

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي



BADJI MOKHTAR - ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR - ANNABA

جامعة باجي مختار - عنابة

Année 2017

Faculté des sciences et de l'ingénierie
Département d'informatique

THESE

Présentée en vue de l'obtention du diplôme de DOCTORAT

**L'intégration d'un système de recommandation au
sein d'un EIAH**

**Option
Intelligence Artificielle
Par
M^{me} BENHAMDI SOULEF**

DIRECTEUR DE THESE : Pr. BABOURI Abdesselam UNIVERSITE DE GUELMA

DEVANT LE JURY

Président :	Pr. BOUHADADA Tahar	Prof. Université d'Annaba
Rapporteur :	Pr. BABOURI Abdesselam	Prof. Université de Guelma
Examineurs :	Pr. LAFFIFI Yassine	Prof. Université de Guelma
	Dr. SARI Toufik	MCA. Université d'Annaba

Remerciements

Cette thèse est le fruit d'un grand effort et d'un travail de plusieurs années. Ce travail qui n'aurait pas pu voir le jour sans la participation, l'aide, les conseils, ou encore la présence de certaines personnes.

Que tous ceux qui, de près ou de loin, ont contribué, par leurs conseils, leurs encouragements ou leur amitié à l'aboutissement de ce travail, trouvent ici l'expression de ma profonde reconnaissance.

Mes vifs remerciements vont tout d'abord à mon encadreur : Pr Babouri Abdesselam, docteur de l'Université de Guelma, pour ses conseils pendant ces années de thèse. Sa disponibilité, son soutien et son esprit pédagogiques m'ont permis d'apprendre beaucoup de choses et de donner le meilleur de moi-même.

Je remercie également Dr Raja Chiky, enseignante chercheuse à l'ISEP, de m'avoir guidé et jugé mes travaux. Je lui suis infiniment reconnaissante de m'avoir encouragée et soutenu dans ce travail.

Je tiens à remercier Professeur BOUHADADA Tahar (Université d'ANNABA) qui me fait l'honneur de présider ce jury. Je remercie également Dr SARI Toufik de l'université d'ANNABA ainsi que Professeur LAFFIFI Yassine de l'université de GUELMA, qui ont aimablement accepté d'examiner cette thèse. J'espère que leurs remarques me permettront de parfaire ce travail.

Mes remerciements vont également à ma famille : mes chers parents, mes frères et sœurs, mon cher fils: Mohamed, qui m'ont poussé jusqu'au bout pour effectuer cette thèse.

Résumé

Les environnements d'apprentissage classiques offrent des informations statiques, c'est-à-dire ils proposent le même contenu pour tous les étudiants. Une fois un système d'apprentissage est développé, il sera difficile, voire impossible de l'adapter selon leurs besoins et préférences. De nouveaux systèmes doivent être mis en place pour assurer la personnalisation du contenu pédagogique. Ce travail de thèse a pour objectif de proposer une nouvelle approche de recommandation des documents aux étudiants selon leurs préférences, niveau et la capacité de leurs mémoires. La nouvelle approche de recommandation est basée sur le filtrage hybride et les taxonomies : NPR-eL (New multi-Personalized Recommender System for e-Learning). NPR-eL a été intégrée dans une autre approche de modélisation et création de cours afin de développer un nouvel système d'apprentissage personnalisé (NPLE : New Personal Learning Environment). Les expérimentations que nous avons réalisées prouvent la performance de l'approche proposée.

Mots clés : e-Learning, Taxonomies, IMSLD, Personnalisation, systèmes de recommandation ; démarrage à froid, EIAH.

Abstract

Traditional e-Learning environments are based on static contents considering that all learners are similar, so they are not able to respond to each learner's needs. These systems are less adaptive and once a system that supports a particular strategy has been designed and implemented, it is less likely to change according to student's interactions and preferences. New educational systems should appear to ensure the personalization of learning contents. This work aims to develop a new personalization approach that provides to students the best learning materials according to their preferences, interests, background knowledge, and their memory capacity to store information. A new recommendation approach based on collaborative and content-based filtering is presented: NPR_eL (New multi-Personalized Recommender for e-Learning). This approach was integrated in a learning environment in order to deliver personalized learning material. We demonstrate the effectiveness of our approach through the design, implementation, analysis and evaluation of a personal learning environment.

Keywords: personalized recommender system; personalization; e-Learning; personal learning environments; IMSLD, Cold start problem.

ملخص

أنظمة التعليم الإلكتروني التقليدية توفر نفس المعلومات لجميع الطلبة علما أن لكل طالب خصائص ومتطلبات مختلفة عن الآخر، وبناء عليه فإن تطوير هذه الأنظمة وجعلها أكثر استعابا لهاته الخصائص أصبح ضرورة ملحة.

في السنوات الأخيرة قام بعض الباحثين بتطوير أنظمة تسمح بتقديم المعلومة المناسبة لكل طالب. هذه الأنظمة تعرف باسم أنظمة الاقتراح.

في هذه المذكرة تم تطوير نظام اقتراح جديد يسمح بتقديم المعلومات إلى الطلبة وذلك استنادا إلى رغبات ومستوى كل طالب بالإضافة إلى قدرته على تخزين هذه المعلومات. النظام المقترح يسمح بحل مشكلة نقص المعلومات الخاصة بالطالب و/أو المعلومات التي سيتم إستخراجها بطريقة تسمح له بتوفير الوقت والجهد. حيث تم دمج هذا النظام مع نظام للتعليم التقليدي بهدف تطويره.

النتائج المتحصل عليها أثبتت قدرة نظام الاقتراح الجديد على تقديم معلومات مناسبة لكل طالب.

الكلمات المفتاحية : أنظمة الاقتراح ، أنظمة التعليم الإلكتروني ، التعليم عن بعد، مشكلة نقص المعلومات.

Table des matières

Introduction générale

Chapitre 1- Contexte du travail : Environnements Informatiques pour l'Apprentissage Humain (EIAH)

1.1 Introduction.....	1
1.2 Contexte du travail.....	1
1.2.1 Enseignement Assisté par Ordinateur (EAO).....	1
1.2.2 EIAO ₁	2
1.2.3 EIAO ₂	3
1.2.4 EIAH.....	3
1.2.5 Formation Ouverte et A Distance (FOAD).....	3
1.2.6 Les plates formes d'enseignement.....	5
1.3 Les langages de modélisation pédagogiques.....	6
1.4 Les normes pour le e-Learning.....	8
1.4.1 Learning Object Model (LOM).....	9
1.4.2 Sharable Content Object Référence Model(SCORM).....	9
1.5 Les normes IMS.....	10
1.5.1 IMS simple sequencing.....	11
1.5.2 IMS Learner information package.....	12
1.5.3 IMS Content Packaging.....	12
1.5.4 IMS Learning Design.....	13
1.6 L'apport des EMLs.....	19
1.7 Conclusion	20

Chapitre 2- Les systèmes de recommandation

2.1 Introduction.....	21
2.2 Les systèmes de recommandation.....	21
2.3 La typologies des systèmes de recommandation.....	22
2.3.1 La recommandation éditoriale.....	22
2.3.2 La recommandation sociale.....	23
2.3.3 La recommandation contextuelle.....	24
2.3.4 La recommandation personnalisée.....	25
2.4 Evaluation des recommandations.....	37
2.4.1 Mésures statistiques de précision.....	37
2.4.2 Mesures permettant l'aide à la décision	38

2.4.3 Couverture.....	40
2.4.4 Temps de calcul.....	41
2.5 Les problèmes des systèmes de recommandation.....	41
2.5.1 Le démarrage à froid.....	41
2.5.2 La robustesse.....	43
2.5.3 Précision des recommandations	44
2.6 La personnalisation.....	44
2.6.1 Définition.....	44
2.6.2 Profils ou modèles d'utilisateurs.....	44
2.6.3 Représentation du profil apprenant.....	45
2.6.4 Utilisation du profil apprenant.....	46
2.6.5 Services de la personnalisation.....	50
2.7 Les types d'apprentissage.....	50
2.8 Personnaliser l'apprentissage.....	53
2.9 Les approches de la personnalisation de l'apprentissage.....	54
2.10 La personnalisation multidimensionnelle.....	59
2.11 Les environnements d'apprentissage personnalisé (PLEs).....	60
2.11.1 Définition.....	60
2.11.1 Les avantages des PLEs.....	62
2.12 Conclusion.....	63

Chapitre 3- Vers un système d'apprentissage personnalisé (NPLE)

3.1 Introduction.....	64
3.2 Les solutions au problème de démarrage à froid.....	64
3.2.1 Les taxonomies.....	65
3.2.2 Relation entre préférences de l'item et préférences taxonomiques.....	66
3.3 travaux existants.....	67
3.4 Un nouveau système d'apprentissage personnalisé (NPLE).....	69
3.4.1 Modélisation d'une situation d'apprentissage	71
3.4.1.1 la particularisation du scénario.....	71
3.4.1.2 la création d'une unité d'apprentissage UoL.....	88
3.4.2. La phase d'apprentissage.....	89
3.5 Conclusion.....	91

Chapitre 4- Expérimentation et évaluation

4.1 Introduction.....	92
-----------------------	----

4.2 Données exploitées.....	93
4.2.1 University Datatset.....	93
4.2.2 Book-Crossing Dataset.....	94
4.3 Métriques d'évaluation.....	95
4.4 Expérimentation en utilisant University Dataset.....	95
4.4.1 Calcul des métriques d'évaluation.....	101
4.4.2 Explication des résultats.....	103
4.4.3 Discussion des résultats.....	104
4.5 Expérimentation en utilisant Book-Crossing Dataset.....	105
4.5.1 Calcul des métriques d'évaluation.....	106
4.5.2 Discussion des résultats.....	107
4.6 Discussion Globale.....	108
4.7 Caractéristiques du NPLe.....	109
4.8 Conclusion.....	110

Conclusion et Perspectives

Bibliographie

ANNEXES

Annexe A XML (eXtensible Markup Langage)

Annexe B UML (Unified Modeling Language)

Annexe C The University's Taxonomy

Liste des figures

Figure 1.1 : Exemple d'architecture de plateforme pour la FAD (George, 2001)...	6
Figure 1.2 : L'environnement d'exécution de SCORM (Mbala, 2003)	10
Figure 1.3 : IMS Content Packaging scope.....	13
Figure 1.4 : Modèle conceptuel d'IMS Learning Design.....	15
Figure 2.1 : Exemple de recommandation éditoriale présente sur le site Alapage	23
Figure 2.2 : Exemple de recommandations sociales proposées par Amazon et Youtube.....	24
Figure 2.3 : Exemple de recommandations contextuelles sur le site de la Fnac.....	25
Figure 2.4 : Exemple de recommandations personnalisées proposées par Allociné	26
Figure 2.5 : Le modèle général du filtrage collaboratif (Berrut et Denos, 2003).....	30
Figure 2.6 : Le concept de compétence (Paquette , 2002).....	49
Figure 3.1 : L'architecture du système (NPLE).....	71
Figure 3.2 : Particularisation du scénario	72
Figure 3.3 : La structure d'un <i>manifest</i>	72
Figure 3.4: Diagramme en arbre dérivant l'élément rôles	73
Figure 3.5 : Diagramme en arbre dérivant l'élément <i>activities</i>	74
Figure 3.6 : Diagramme en arbre dérivant l'élément <i>method</i>	74
Figure 3.7 : Diagramme en arbre dérivant l'élément <i>resources</i>	75
Figure 3.8 : Le système de recommandation NPR-eL	77
Figure 3.9 : Les éléments d'un manifest.....	78
Figure 3.10 : Réduction de l'espace de recherche en utilisant <i>clustering</i> (Wreng, 2008).....	84
Figure 3.11 : La formation des groupes des similaires (NPR-eL).....	84
Figure 3.12 : Diagramme en arbre dérivant l'élément <i>method</i>	89
Figure 4.1 : L'environnement NPLE	92
Figure 4.2 : Espace enseignant	93
Figure 4.3 : Espace apprenant	93
Figure 4.4 : Les évaluations réelles de a1	94
Figure 4.5 : Le questionnaire proposé et un exemple de test (niveau 1).....	96
Figure 4.6 : Le test de RSI	97
Figure 4.7 : Le résultat de la recommandation (University Dataset).....	98
Figure 4.8 : Exécution d'un exemple d'UoL	100
Figure 4.9 : Un exemple du post test	101
Figure 4.10 : L'évaluation de NPR-eL (University Dataset).....	102
Figure 4.11 : Evaluation de Boosted_CSHTR (Benhamdi et Séridi 2011).....	102
Figure 4.12: Les résultats des pré-tests et post test (Boosted_CSHTRL).....	103
Figure 4.13 : Les résultats des pré-tests et post test (NPR-eL).....	103
Figure 4.14 : Le résultat de la recommandation.....	105
Figure 4.15 : Evaluation du CSHTR en utilisant F1 (Weng, 2008).....	106
Figure 4.16 : Les résultats des pré-test et posttest (Boosted_CSHTRL).....	107
Figure 4.17 : Les résultats des pré-tests et post test (NPR-eL).....	107

Liste des Tableaux

Tableau 2.1 : Méthodes d'hybridations (Burke, 2002)	35
Tableau 2.2 : Catégories d'items basées sur l'intersection entre listes de recommandation et 39 préférences réelle (Esslimani, 2010).....	
Tableau 3.1: Les modèles d'unités pédagogiques	75
Tableau 3.2 : Le modèle d'UP 10	76
Tableau 3.3 : CSHTR et NPR-eL : comparaison.....	77
Tableau 3.4 : Les étapes d'exécution d'une UoL.....	90
Tableau 3.5 : Le processus d'apprentissage.....	91
Tableau 4.1 : Les évaluations explicites des apprenants.....	94
Tableau 4.2 : Les résultats des tests.....	98
Tableau 4.3 : Les items recommandés.....	105
Tableau 4.4 : Les évaluations réelles faites par u0.....	108
Tableau 4.5 : Les items recommandés.....	108

Introduction générale

1. Contexte et problématique

L'apparition de l'internet a permis d'offrir aux utilisateurs une grande masse de données, qui ne cesse d'augmenter chaque jour. Ces données, appelées 'ressources' ou 'items' peuvent être hétérogènes et distribuées. Les items sont des documents électroniques sous format textuel ou multimédia regroupant un ensemble de données informatives.

Selon une évaluation de l'Internet World Stats¹ réalisée en 2010, il y aurait plus de 1,9 milliards d'internautes dans le monde pouvant consulter environ 109,5 millions de sites Web opérationnels et 25, 21 milliards de pages (Esslimani, 2010). Or, devant cette surabondance d'items, l'accès aux informations qui permettent de satisfaire au mieux les besoins des utilisateurs devient une tâche fastidieuse.

Dans ce contexte, le développement des outils permettant de filtrer ce volume important de données s'avère crucial. L'objectif principal de la personnalisation de l'accès à l'information est la satisfaction des attentes des utilisateurs. En effet, la personnalisation a pour finalité d'amener à l'utilisateur des items pertinents relativement à ses besoins et préférences. La personnalisation est un axe de recherche qui a suscité l'intérêt et l'engouement de nombreux chercheurs. Les approches de personnalisation qui ont été mises en place sont : les approches basées sur le contenu (Mladenic, 1999), les techniques à base de critiques issue du domaine de raisonnement à partir des cas ("Case Based Reasoning" (CBR)) (Burke, 2000), les approches basées sur la navigation sociale (Svensson et al., 2005), etc.

Les systèmes de recommandation sont considérés comme l'un des outils de la personnalisation de l'accès à l'information. Ils peuvent exploiter les approches citées ci-dessus, en vue de proposer à un utilisateur actif (i.e. un utilisateur courant), des

¹ <http://www.internetworldstats.com>

conseils d'items qu'ils jugent pertinents par rapport à ses attentes. En d'autres termes, les systèmes de recommandation ont pour but d'orienter l'utilisateur vers l'information qui lui convient. Au titre d'exemple, sur un portail Extranet d'entreprise, les informations que peuvent être proposées à l'utilisateur actif par le système de recommandation sont : un article spécialisé, une actualité ou bien un rapport technique. Sur un site d'e-commerce, le système de recommandation peut proposer à cet utilisateur un produit à acheter, un livre à lire ou un film à regarder (Esslimani, 2010).

Ces dernières années, plusieurs systèmes de recommandation ont été mis en place. CSHTR (Cold Start Hybrid Taxonomic Recommender) est l'un de ces systèmes qui a prouvé son efficacité par rapport aux autres techniques en terme de résistance au problème de démarrage à froid, et en terme de prédiction. Ce système exploite les préférences taxonomiques des utilisateurs pour traiter le problème de démarrage à froid. Mais, cette solution n'est pas idéale parce que les utilisateurs doivent lire et évaluer des documents pour pouvoir capturer leurs préférences taxonomiques, avant de recevoir des recommandations. Pour cela, nous avons proposé aux apprenants de répondre à un questionnaire pour pouvoir choisir directement leurs domaines d'intérêt. De plus, CSHTR prend en considération seulement les préférences des utilisateurs, alors que d'autres propriétés importantes sont absolument négligées.

Dans le domaine de l'EIAH, qui est notre domaine de recherche, les préférences des étudiants en termes d'items sont jugées insuffisantes, du moment où la qualité de recommandation et d'apprentissage sont fortement liés au système cognitif de l'apprenant.

2. Approche proposée

L'objectif de cette thèse est d'améliorer la solution proposée par l'approche CSHTR au problème de démarrage à froid et l'utiliser dans le domaine de l'EIAH. Nous avons proposé un questionnaire afin de construire les profils d'apprenants. De cette façon, nous pouvons savoir les préférences des utilisateurs en leur évitant une perte de temps dans la lecture de quelques documents qui peuvent être loin de leurs préférences. Nous avons aussi ajouté d'autres caractéristiques relatives au système cognitif des apprenants qui sont : leur niveau et la capacité de leur mémoire. En effet, notre

hypothèse est que l'intégration de plusieurs caractéristiques, peut améliorer la qualité de recommandation.

La nouvelle technique de recommandation (NPR-eL : New Personalized Recommender System for e-Learning) a été intégrée dans une méthode permettant la création de scénarios pédagogiques pour développer un nouveau environnement d'apprentissage personnalisé (NPLE : New Personalized Learning Environment).

3. Structure du document

Dans ce manuscrit, nous présentons dans le premier chapitre le contexte général en éclaircissant la relation entre les technologies de l'information et de communication, les environnements d'apprentissage humain et le domaine du e-Learning.

Dans le deuxième chapitre, nous décrivons l'origine des systèmes de recommandation ainsi que les concepts liés à ce domaine. De plus, il sera question de décrire les principales techniques de recommandation en discutant leurs avantages et leurs inconvénients.

Le troisième chapitre présente l'approche de recommandation proposée: NPR-eL et une description détaillée de l'approche de modélisation et création de cours au sein de laquelle, nous avons intégré la nouvelle approche de recommandation.

Le dernier chapitre montre les résultats de notre expérimentation. Ces résultats prouvent la performance de l'approche proposée : NPR-eL.

Chapitre 1

Contexte général: Environnements Informatiques pour l'Apprentissage Humain (EIAH)

1.1 Introduction

L'objectif de cette thèse est de mettre en relation le domaine de l'éducation et les technologies de personnalisation : les systèmes de recommandation afin de créer un environnement d'apprentissage personnalisé. Nous allons consacrer le premier chapitre pour l'éducation et plus précisément notre domaine de recherche : l'EIAH. Depuis plusieurs années, le glissement des EAO (Enseignement Assisté par Ordinateur) (Bruillard, 1997) vers des environnements permettant l'interaction entre des machines et des humains distribués dans l'espace, a permis l'apparition de ce qu'on appelle « les Environnements Informatiques pour l'Apprentissage Humains » (EIAH) (Tchounikine, 2002). Ce sigle permet de prendre en compte l'évolution des technologies de télécommunication et d'évoluer vers la formation à distance. L'e-formation est née grâce à ces environnements et l'apparition des technologies de l'information et de la communication appliquées à l'éducation (TICE) ainsi que les grands réseaux d'information (Broisin, 2006).

Dans ce chapitre, nous présentons un bref historique de l'EIAH. Par la suite, nous évoquerons l'évolution de la formation à distance vers la formation ouverte et à distance en utilisant les TICE. De plus, nous allons donner une définition des plates formes d'enseignement. Enfin, nous définissons quelques langages de modélisation pédagogique en mettant l'accent sur le langage utilisé dans ce travail : IMSLD (IMS Learning Design) (Laforcade, 2004).

1.2 Contexte du travail

Nos travaux de recherche se situent dans le domaine des EIAH. Notre objectif est de personnaliser l'apprentissage au sein d'un EIAH. Depuis plusieurs années, le glissement des EAO vers des environnements permettant l'interaction entre des machines et des humains distribués dans l'espace, a permis l'apparition de ce qu'on appelle « les environnements informatiques pour l'apprentissage humain » (EIAH) (Tchounikine, 2002). En effet, le terme EIAH couvre une diversité des systèmes dont le point commun est la mise en relation d'une intention didactique et d'un environnement informatique.

1.2.1 L'enseignement assisté par ordinateur (EAO) :

Depuis plusieurs années, de nombreuses recherches ont eu pour objectif d'utiliser les machines en éducation. Les premières tentatives sont apparues dans les années 50 (George,

2001). La construction et l'utilisation de ces machines étaient très compliquées. Ces machines à enseigner sont utilisées alors pour automatiser les fichiers papiers de l'enseignement programmé. Cet enseignement programmé se fonde sur le courant de la psychologie comportementale de l'époque (Skinner, 1954). L'apparition de L'ordinateur donne la naissance à L'EAO et offre la possibilité de l'individualisation de l'enseignement (Bruillard, 1997).

Les années soixante ont été marquées par l'apparition des premiers systèmes informatiques tels que les systèmes de type « PLATO » avec présentation imagée et variée de l'information, suivie de questionnaire de différents types. Ces systèmes nommées « si..alors..Sinon » étaient très coûteux à produire et rigide à utiliser car les éditeurs étaient moins sophistiqués à cette époque. L'EAO est un domaine pluridisciplinaire qui dépasse largement le domaine de l'informatique, ce qui signifie qu'il dépend des résultats des recherches d'autres disciplines telles que : la didactique, la psychologie.... Ce domaine a donné la naissance de l'« EIAO » avec ses différentes significations (Bruillard, 1997):

- EIAO₁ : Enseignement Intelligemment Assisté par Ordinateur.
- EIAO₂ : Environnement Interactif d'Apprentissage avec Ordinateur.

1.2.2 EIAO₁ : enseignement intelligemment assisté par ordinateur

Les limites constatées des systèmes d'EAO et des recherches menées en intelligence artificielle (IA), en science de l'éducation et en sciences cognitives ont conduit à l'émergence du sigle EIAO₁ aux états –Unis, avec le célèbre système SCHOLAR de (Carbonell, 1970), d'où la connaissance, représentée dans des structures d'information, était utilisée, non seulement pour être présentée, pour déterminer des questions à poser à l'apprenant et pour vérifier ses réponses, mai aussi pour répondre à des questions de ce dernier, non explicitement prévues à la conception du système, et ce en langage naturel (Baron ,1995).

Les Systèmes Tuteurs Intelligent (STI ou en anglais : Intelligent Tutoring System) sont apparus dans les années 80, étaient fortement liés au développement des systèmes à base de connaissance en IA (Prévot, 1992). Ces systèmes cherchent à jouer le rôle d'un enseignant. Le modèle des STI constitué de trois expertises : celle du domaine à enseigner, celle de l'enseignement (module pédagogique), celle des compétences de l'apprenant (modèle d'apprenant). Ce modèle a été conçu dans le but de s'adapter à l'apprenant.

1.2.3 EIAO₂: Environnement Interactif d'Apprentissage avec ordinateur

Ces environnements cherchent à combiner l'apprentissage par résolution de problème et l'apprentissage par la découverte. En effet, l'arrivée du terme « Interactif » signifie que ces environnements s'intéressent beaucoup plus à l'interactivité qui recouvre d'une part les activités qu'ils permettent à un apprenant et d'autre part leurs capacités d'intervention et d'adaptation à leur utilisateur (Baron et al., 1991). Le glissement du terme « enseignement » vers « apprentissage » révèle qu'on s'intéresse plus à la construction des connaissances chez un apprenant. Le terme « avec ordinateur » souligne que la machine accompagne l'apprenant tout au long du processus de l'apprentissage.

1.2.4 les Environnements Informatiques pour l'Apprentissage Humain (EIAH) :

Plus récemment, c'est le terme EIAH « les environnements informatiques pour l'apprentissage humain » qui a pris la place des termes précédents. La définition des EIAH la plus adoptée est celle de Tchounikine (Tchounikine, 2002) :

« un EIAH est un environnement informatique conçu dans le but de favoriser l'apprentissage humain, c'est-à-dire, la construction de connaissances chez un apprenant. Ce type d'environnement mobilise des agents humains (élève, enseignant et tuteur) et artificiels (agents informatiques qui peuvent eux aussi tenir différents rôles) et leur offrent des situations d'interaction, localement ou à travers les réseaux informatiques, ainsi que des conditions d'accès à des ressources formatives (humaines et/ou médiatisées), ici encore locales ou distribuées ».

1.2.5 La Formation Ouverte et A Distance

Avant de passer à la définition de la FOAD (Betbeder, 2003), il faut tout d'abord définir les termes « Enseignement », « Apprentissage » et « Formation ». Le premier a pour objectif de transmettre le savoir. Il est organisé pour préparer, assister, suivre, contrôler et améliorer l'apprentissage. Ce dernier est le processus de transformation des informations acquises en entrée vers la sortie pour construire des connaissances chez un apprenant. La formation est un point de vue qui englobe les deux processus : l'enseignement (point de vue enseignant) et l'apprentissage (point de vue apprenant).

Dans cette partie, nous présentons une définition des sigles : EAD (Enseignement A Distance), FAD (Formation A Distance) (Marot et Darnige, 1996), et nous terminons par la présentation de la Formation ouverte et A Distance (FOAD). La formation est une relation

triangulaire, d'où l'apprenant construit ses connaissances en interagissant avec son environnement, le formateur ou l'enseignant contrôle l'activité de l'apprenant.

1.2.5.1 EAD : Enseignement A Distance

C'est le sigle utilisé, il désigne le transfert des connaissances dans une situation éducative en dehors d'une situation en face à face ou en « présentiel ». Donc, il prend en compte la distance géographique, d'où l'usage des nouveaux outils de communication (messagerie, téléchargement de fichier,...) au lieu d'utiliser l'envoi postal (Lameul, 2000).

1.2.5.2 FAD : Formation A Distance

La formation à distance couvre l'ensemble des dispositifs techniques et des modèles d'organisation qui ont pour but de fournir un enseignement ou un apprentissage à des individus qui sont distants de l'organisme de formation prestataire du services. En effet, il existe pas mal de définitions proposées par plusieurs spécialistes de la FAD (Keegan ,1996), l'un de ces spécialistes, qui a proposé une définition de la FAD indépendamment de l'avènement des TICE et d'Internet :

La formation à distance est une forme d'enseignement caractérisée par (Oubahssi, 2005) :

- la séparation quasi-permanente entre le formateur et l'apprenant tout au long du processus d'apprentissage (ceci différencie la formation à distance de la formation présentielle);
- l'influence d'une organisation administrative aussi bien en ce qui concerne la planification et la préparation des matériaux pédagogiques que la mise à la disposition des apprenants des services d'accompagnement et de support (ceci différencie la FAD de l'autoformation);
- l'utilisation de médias techniques (imprimerie, audio, vidéo, ordinateurs) pour assurer le lien entre le formateur et l'apprenant et médiatiser le contenu de la formation;
- L'existence de mécanismes de communication bidirectionnelles afin que l'apprenant bénéficie mais prenne aussi l'initiative de dialogues avec le formateur (ceci distingue la FAD des autres usages de la technologie dans l'éducation);
- la quasi-absence de la notion de groupe tout au long du processus d'apprentissage, de sorte que les apprenants sont toujours vus comme des individus isolés et non comme faisant partie d'un groupe, avec la possibilité d'organiser occasionnellement des rencontres, soit en présentiel, soit via des moyens électroniques à des fins didactiques ou de socialisation.

1.2.5.3 FOAD : Formation Ouverte et A Distance :

Les formations ouvertes et à distance (FOAD) peut être définie comme suit (Betbeder, 2003):

« *La formation ouverte et à distance...*

- est un dispositif organisé, finalisé, reconnu comme tel par les acteurs ;
- qui prend en compte la singularité des personnes dans leurs dimensions individuelle et collective ;
- et repose sur des situations d'apprentissage complémentaires et plurielles en termes de temps, de lieux, de médiations pédagogiques humaines et technologiques, et de ressources.etc.

La Formation Ouverte et A Distance (FOAD) englobe les concepts de formation ouverte (entre différents techniques d'enseignement) ainsi que la formation à distance. Cette approche, cherche à intégrer tous les éléments et toutes les phases du cycle de formation. Les débuts de ce mouvement d'intégration en Europe étaient l'apparition du concept du moteur d'enseignement (Claës, 1988) : système qui tente de répondre à toutes les phases de la formation, aussi bien dans la préparation (contenu et cursus), dans l'utilisation (apprentissage et soutien) et dans le suivi de la formation (pédagogique et administratif). Le domaine de la FOAD est approprié à l'introduction des technologies (CD-ROM, Internet, Intranet, etc.) dans la formation ainsi que les différents types de situations d'apprentissage (présentiel, à distance, autoformation). L'utilisation de l'Internet et les TICE dans la FOAD constitue ce que l'on désigne couramment par : e-formation (Broisin, 2006). Grâce aux TIC (technologies de l'information et de la communication), la FOAD devrait pouvoir aujourd'hui concilier économie (réduction des frais), réponse aux contraintes (individualisation de la formation) et efficacité (rapidité du déploiement à grande échelle).

1.2.5.4 E-formation ou e-Learning

L'e-formation peut être définie comme toute utilisation de l'Internet dans la gestion et la diffusion de la formation. Les concepteurs de la e-formation cherchent essentiellement un certain nombre de qualités comme : une distribution plus large et plus efficace de la formation, l'individualisation et l'adaptation de parcours de formation aux besoins individuels des apprenants, des possibilités de suivi et de tutorat, la rapidité,... (Broisin, 2006).

Bien que la e-formation permette de satisfaire un nombre important de besoins, elle pose des problèmes très généraux qui apparaissent aussi bien dans la formation en présentiel que pour certains dans la FAD.

1.2.6 Les plates formes d'enseignement

Le domaine de la formation à distance a bénéficié de l'évolution des technologies de l'information et de la communication éducative (TICE), du fait qu'elle apporte une plus-value

organisationnelle et pédagogique. Ces technologies offrent la possibilité d'accéder à des contenus distants et d'échanger des informations entre les différents acteurs (apprenant/enseignant, apprenants entre eux, enseignants entre eux).

Les outils de formation qui adoptent les réseaux informatiques comme support sont appelés : plates formes informatique pour la formation à distance ou encore plates forme d'enseignement (PFE ou en anglais : LMS Learning Management System). Ces plates formes sont développées dans le but de la consultation à distance des contenus pédagogiques, l'individualisation de l'apprentissage et le tutorat à distance. Les PFE intègrent des dispositifs permettent de faciliter les activités réalisées par les différents acteurs de la FAD (George, 2001): enseignant, concepteur informatique, tuteur, apprenant et administrateur. On commence par le rôle de l'enseignant qui est chargé de concevoir le matériel pédagogique. Les médias (texte, image,...) sont alors réalisés par le concepteur informatique, qui peut être l'enseignant lui-même. L'enseignant conçoit ensuite les parcours pédagogiques typés et individualisés de son enseignement. Le tuteur prend en charge le suivi et le contrôle du travail de l'apprenant. Il peut éventuellement réguler leurs parcours pédagogiques. L'apprenant consulte, télécharge les contenus nécessaires pour son travail, effectue les activités qui lui sont demandées, ensuite il envoie son travail au tuteur pour l'évaluation. L'apprenant peut communiquer avec l'enseignant pour, par exemple, discuter sur le thème proposé. L'installation et la maintenance du système, l'inscription administrative des étudiants, la gestion des droits d'accès aux ressources pédagogiques, sont réalisés par l'administrateur.

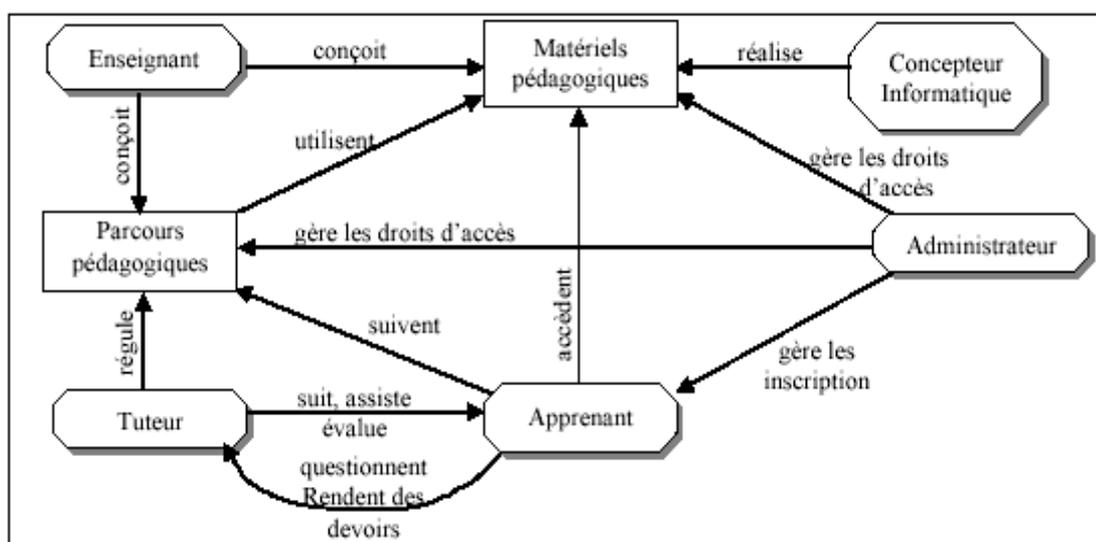


Figure 1.1 Exemple d'architecture de plateforme pour la FAD (George, 2001).

Cette figure présente un modèle de PFE avec cinq acteurs. Il existe des plates formes qui ne prennent pas en considération tous ces acteurs, alors que d'autres sont plus complètes. Dans certains cas, l'enseignant joue le rôle d'un concepteur informatique ou d'un tuteur (George, 2001).

1.3 Le langage de modélisation pédagogique

En ingénierie logicielle, l'activité de modélisation est primordiale comme elle l'est en EIAH. Les travaux de recherches ont montré qu'il n'est pas suffisant de mettre des ressources d'apprentissage à la disposition des plates formes de formation à distance. Il faut également fournir des outils permettant de les gérer. Dans ce sens, des langages de modélisation pédagogique ou éducative (*Educational Modeling Language*) sont apparus (Giacomini, 2005).

La définition d'EML qu'on utilise dans ce document est utilisée comme définition de référence de nos jours :

“An EML is a semantic information model and binding, describing the content and process within a “unit of learning“ from a pedagogical perspective in order to support reuse and interoperability” (Rawlings, 2002).

Les EML proposent un modèle d'information décrivant les concepts, leurs relations, leur sémantique et le processus dans une unité d'apprentissage selon une perspective pédagogique donnée pour supporter la réutilisabilité et l'interopérabilité, ce qui implique que les modèles décrits avec les EML doivent être indépendants des diverses théories d'apprentissage : ils permettent aussi de décrire une correspondance vers une formalisation de ce modèle dans un langage compréhensible par la machine.

Une unité d'apprentissage (unit of Learning appelée aussi : unit of study ou unit of instrument) est la plus petite unité définissant un modèle d'apprentissage, elle peut être un cours, un module de formation,... Nous définissons maintenant quelques concepts utilisés par les EMLs : Scénario, acteurs : enseignant et apprenant.

Scénario :

Un scénario pédagogique peut être décrit comme le déroulement du processus de l'apprentissage, ou encore la manière d'exécution des activités par les différents acteurs. Les activités à réaliser, les acteurs et leurs rôles et les scénarios sont déterminés lors de la phase de conception.

Acteurs :

L'un des concepts qui doit être défini lors de la conception de situation d'apprentissage est l'acteur qui joue un certain rôle durant L'apprentissage. Dans notre travail on s'intéresse à deux types d'acteurs : L'enseignant autant qu'un créateur du contenu pédagogique, et l'apprenant.

L'enseignant :

L'enseignant a le rôle d'un concepteur de situations d'apprentissage et d'un contrôleur d'activité de l'apprenant. Le processus de conception consiste en premier lieu à déterminer l'objectif de l'apprentissage et le type du contenu pédagogique qui doit être utilisé durant l'apprentissage en répondant à un questionnaire. Ensuite, il doit donner une description détaillée de chaque activité pour la rendre plus compréhensible à l'apprenant. Finalement, il doit choisir les services indispensables pour le processus d'apprentissage. L'enseignant peut exercer un contrôle sur le travail des apprenants, il peut ainsi donner des informations ou des conseils, répondre à leurs questions....

L'apprenant :

L'apprenant est la personne intéressée par l'apprentissage, et pour laquelle des situations d'apprentissage sont conçues, dans le but de construire un savoir. De ce fait, L'apprenant doit effectuer l'ensemble d'activités déterminées par l'enseignant en utilisant des ressources et des services disponibles dans l'environnement d'apprentissage. Ces ressources peuvent être des fichiers ou des documents contenant des : cours, exercices, projets, problèmes à résoudre,....

Les services permettent aux apprenants de s'interagir et de communiquer aussi avec l'enseignant.

1.4 Les normes pour le e-Learning

Avant de déterminer l'intérêt des normes, on doit tout d'abord définir le terme « norme » : une norme est un ensemble de règles de conformité, étudié par un organisme de normalisation au niveau national et international (Aubert, 2005).

Le domaine du e-Learning permet de présenter aux apprenants un ensemble de cours et faciliter la création de ces cours pour les enseignants. La normalisation d'après Michel Arnaud : « a pour but d'assurer l'interopérabilité et la réutilisabilité des outils d'enseignement médiatisé et à distance, ainsi que des systèmes de description des contenus organisés en métadonnées ». La mise en œuvre des cours en utilisant une norme permet de les rendre

présentable sur n'importe quelle plateforme, ce qui signifie que l'enseignant n'écrit qu'une seule fois son cours.

1.4.1 Learning Object Model (LOM)

LOM (Aubert, 2005) est un standard qui permet la description des objets pédagogiques selon 9 catégories et comporte 71 attributs. Cette description est exprimée à l'aide d'un fichier XML, contenant les caractéristiques d'un document, ou le contenu d'un cours ou d'un exercice pour l'e-Learning.

Le schéma de métadonnées LOM est très détaillé et fournit un ensemble de vocabulaire de référence. Plusieurs organismes de standardisation (IEEE LTSC, IMS, ADL, ARIADNE) ont adopté le modèle LOM qui a été intégré le plus souvent dans des profils d'application (SCORM, CanCore, etc.).

Le modèle de métadonnée générique Dublin Core qui permet de décrire n'importe quel type de document a été intégré au LOM.

1.4.2 Sharable Content Object Référence Model (SCORM)

SCORM (Giacomini, 2005) est le modèle de référence pour le partage de « contenus » et définit un modèle d'agrégation permettant de créer et exécuter des objets pédagogiques structurés. SCORM regroupe un ensemble de spécifications techniques permettant aux systèmes d'apprentissage de réaliser plusieurs fonctionnalisées telles que : trouver, importer, partager, réutiliser et exporter les contenus de manière normalisée.

Découpage en plusieurs parties

SCORM peut être constitué de plusieurs parties:

1. Modèle d'agrégation de contenu : qui assure la promotion de méthodes cohérentes en matière de stockage, d'identification, de conditionnement d'échange et de repérage du contenu.

Ce modèle peut comporter plusieurs fonctionnalités : la première consiste en la définition des métadonnées LOM (Learning Object Metadata) qui permettent d'identifier les termes décrivant le contenu de l'objet d'apprentissage. La deuxième établit la liaison entre les métadonnées et les fichiers (XML), pour définir la manière de leur codage afin de les rendre compréhensibles par la machine. La dernière fonctionnalité définit comment emballer un ensemble d'objets d'étude, leurs métadonnées et des informations concernant le contenu.

2. Environnement d'exécution:

Décrit les exigences du système de gestion de l'application nécessaire à la gestion de l'environnement d'exécution. Il fournit (Mbala, 2003) :

- Un mécanisme commun de lancement des objets pédagogiques.
- Un dispositif commun de communication entre ces objets et le système de gestion d'apprentissage (LMS : Learning Management System).
- Une base constituée d'un vocabulaire pour cette communication.

L'interface de programme (API) joue le rôle d'un intermédiaire entre un objet pédagogique et le LMS afin d'assurer la communication entre eux. Cette API informe le LMS de l'état de la ressource pédagogique (initialisée, terminée, en état d'erreur...). Il n'y a pas d'interaction entre les ressources pédagogiques «Assets» et le 'LMS'. Pour ceux, l'API n'est pas utile pour ce type de ressources. Le modèle de données fourni par l'environnement d'exécution définit un vocabulaire commun pour la communication entre le LMS et l'objet pédagogique (SCO).

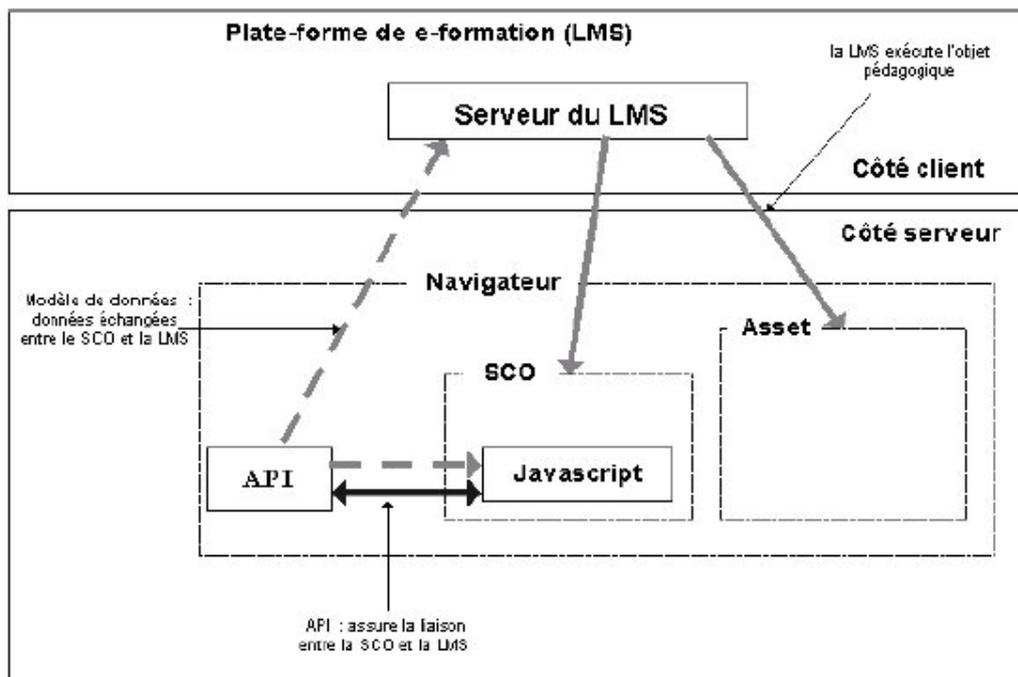


Figure 1.2 L'environnement d'exécution de SCORM (Mbala, 2003).

3. Le modèle de séquençement et de navigation:

Il décrit comment le système interprète les règles de séquençement exprimées par un développeur et les événements de navigation qui peuvent être lancés par l'apprenant ou par le système.

1.5 Les normes IMS

Il existe également plusieurs spécifications du consortium IMS : IMS Content Package (IMS, 2003a), IMS LIP et IMS *Simple Sequencing* (IMS, 2003b) et IMS LD (Laforcade, 2004).

1.5.1 IMS Simple Sequencing

IMS Simple Sequencing (IMS SS) permet de définir l'ordre de présentation des activités du e-Learning à l'apprenant et les conditions selon lesquelles une ressource est sélectionnée, délivrée (*delivered*), ou ignorée lors de la présentation. Cette spécification ne prend en considération qu'un nombre limité des stratégies de séquençement communes. De ce fait, elle est nommée : simple IMS SS et utilise l'instruction de lecture du système d'ordinateur (*computer system*) pour déterminer l'activité suivante qui doit être présentée à l'apprenant. Ces instructions sont appelées : informations de séquençement, qui sont divisées en deux catégories : le modèle de définition d'information et le modèle de trace d'information. Les informations définies sont ajoutées à un fichier XML par un auteur ou un outil. Actuellement, le fichier *imsmanifest.xml* utilisé dans IMS CP est le seul fichier utilisé pour cet objectif. Le résultat est donc un ou plusieurs arbres d'activités intégrés avec un ensemble d'instructions de séquençement. Les informations de trace sont obtenues lors de l'exécution des activités par l'apprenant.

Le processus qui détermine l'activité suivante est appelé : le comportement de séquençement. Ce dernier est défini réellement par six processus : navigation, terminaison, *rollup*, sélection, séquençement et exécution. Le comportement de séquençement se déroule par la répétition de ces processus, formant ainsi une boucle répétitive. Cette boucle sera déclenchée quand l'apprenant clique sur un bouton ou un lien, pour passer à l'activité suivante. Cet événement fait référence au processus de navigation, qui génère deux événements : Le premier événement déclenche le processus de terminaison en assurant l'achèvement du traitement des informations concernant l'activité courante avant de passer à l'activité suivante. Le deuxième événement déclenche le processus du séquençement dans lequel une combinaison des règles de séquençement et des contraintes limitées permettant d'identifier l'activité suivante.

Le processus de séquençement donne la possibilité de contrôle au processus d'exécution (*delivery process*) qui initialise l'état du séquençement avant de présenter l'activité à l'apprenant. Si l'apprenant veut continuer, toute la boucle sera répétée.

Le processus du *rollup* est appelé dans le cas où l'activité est terminée et possède des activités précédentes et suivantes dans l'arbre des activités. Afin de déterminer le modèle de

trace d'information du parent, l'état de l'activité terminée est combiné avec l'état des activités adjacentes.

Le processus de sélection est appelé quand l'activité possède des activités dites : filles (sous activité). Ce processus consiste donc à choisir un sous ensemble d'activités à partir de l'ensemble total d'activités.

1.5.2 IMS Learner Information Package

La spécification IMS Learner Information Package (IMS LIP) (Laforcade, 2004) identifie une structure permettant l'échange d'informations concernant l'apprenant entre différents systèmes coopérants tels que : Les systèmes de gestion d'apprentissage, les systèmes des ressources humaines, les systèmes de gestion des données des apprenants, etc.

Les éléments contenus dans le package LIP sont généralement optionnels, l'utilisateur a le choix de les implémenter. On peut citer quelques exemples de ces éléments : l'identification pour identifier un individu en indiquant quelques éléments tels que : nom, prénom, adresse, etc. Les autres éléments sont : le but que l'apprenant vise à atteindre, les activités, les compétences, les intérêts, etc. Dans le but de satisfaire les différents besoins, LIP offre des possibilités d'exécution. En arrivant à un certain niveau de granularité, soit plus général ou plus détaillé, n'importe quel élément de la spécification peut être étendu. Des éléments extérieurs à la spécification, peuvent être aussi rajoutés au package.

1.5.3 IMS Content Packaging

IMS Content Packaging information model (Lorteau et Peralta, 2004) permet la définition des structures de données utilisées dans le but de fournir l'interopérabilité à l'Internet basé contenu avec : des outils de création de contenu, systèmes de gestion d'apprentissage (LMS), et environnement d'exécution. Cette spécification a pour objectif de définir un ensemble de structures standard pouvant être utilisés pour l'échange du contenu. Ces structures offrent la base pour les liaisons de données standards qui permettent aux développeurs et programmeurs de développer des matériaux d'apprentissage (*instructional*) interopérant via les outils auteurs et le LMS environnement d'exécution qui ont été créés indépendamment par différents développeurs.

➤ Le modèle conceptuel d'IMS CP

Le diagramme conceptuel illustrant les différents composants d'IMSCP information model, est présenté par la figure suivante :

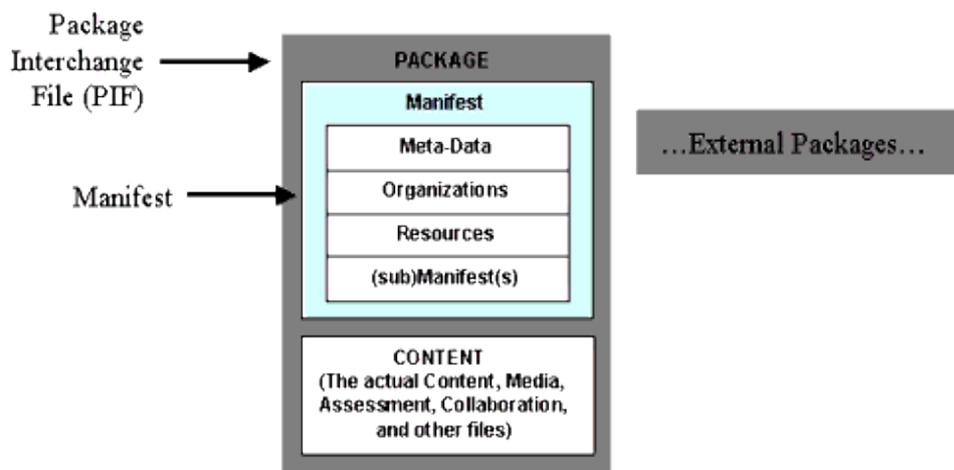


Figure 1.3 IMS Content Packaging scope (Lorteau et Peralta, 2004).

Cette figure montre que l'IMS package comporte deux éléments : un fichier XML (eXtensible Markup Language) décrivant l'organisation du contenu et les ressources dans un package, et les fichiers physiques écrits en XML (Lorteau et Peralta, 2004). Le fichier XML est appelé *imsmanifest.xml*, car le contenu et l'organisation du cours sont définis dans le contexte des « *manifests* ». Une fois que le package soit incorporé dans un fichier pour le transporter, il est nommé : fichier d'échange de package (package interchange file). Ce dernier est un fichier unique compressé contenant un fichier *manifest* de haut niveau (*top-level*) appelé : *imsmanifest.xml*, ainsi que tous les autres fichiers physiques décrits par le *manifest*. Ce fichier est considéré comme un moyen de transport des informations structurées. Le package est un répertoire logique qui comporte un fichier XML et n'importe quel document XML de contrôle dont il fait référence, et les ressources physiques actuelles. Ces ressources peuvent être organisées en sous-répertoire.

1.5.4 IMS Learning Design

IMS LD a été publié en 2003 par IMS/GLC (Instruction Management System Global Learning Consortium) : Consortium pour l'apprentissage global avec les systèmes de gestion de normalisation. Il permet de décrire un processus pédagogique de façon formelle pour être déployé sur une plateforme de formation. Il supporte tout type de pédagogie et il n'est pas attaché à une classe particulière de plateforme. Nous donnons les principales caractéristiques de cette spécification en présentant son méta-modèle (pour plus de détail, sur cette spécification se reporter à (Lafocade, 2004)).

Le méta-modèle d'information d'IMS-LD se décompose en trois composants principaux:

- Le modèle conceptuel : il représente le vocabulaire, les relations entre les concepts et les relations avec la spécification IMS Content Packaging.
- Le modèle d'information : il décrit les éléments d'IMS-LD sur trois niveaux :
 - le niveau A décrit le vocabulaire au cœur de la spécification afin de concevoir des unités d'apprentissage supportant diverses pédagogies ;
 - le niveau B ajoute au niveau A de nouveaux éléments permettant une personnalisation ainsi que des scénarios et des interactions plus élaborées ;
 - le niveau C ajoute un nouveau mécanisme de notification permettant d'étendre la dynamique des scénarios construits.

La description de chacun des trois modèles d'informations s'appuie sur une version restrictive du modèle conceptuel précédent selon le niveau concerné.

- Le modèle de comportement : il décrit un ensemble de comportement d'exécution (*runtime*) que des systèmes délivrant les unités d'apprentissage devront implémenter.

Le concept au cœur de cette spécification pour la conception d'apprentissage est qu'une personne (*person*) joue des rôles (*role*) dans un processus d'apprentissage/enseignement (typiquement, un rôle d'apprenant *learner* ou de personnel staff). En jouant ce rôle la personne cherche à aboutir à des résultats (*outcomes*) en réalisant des activités d'apprentissage (*Learning activity*) ou de soutien (*support activity*) plus ou moins structurées (par le biais des structures d'activités *activity-structure*). Ces activités se réalisent dans le contexte d'un environnement (*environment*). Cet environnement est constitué d'objets d'apprentissage (*Learning Object*) et de services (*services*) qui seront utilisés pendant la réalisation des activités associées à cet environnement. La méthode (*method*) permet de déterminer quel rôle assurera quelles activités à tel moment du processus. Le mécanisme de notification (notification) permet également de préciser certaines structures du processus de l'apprentissage ; ce mécanisme appartient au niveau C. La méthode est conçue pour atteindre des objectifs d'apprentissage (*Learning objectives*) qui correspondent à la spécification de résultats pour l'apprenant.

La méthode est décrite sur l'hypothèse de certains pré requis (*prerequisites*) correspondants alors à la spécification des pré requis d'entrée pour les apprenants.

La méthode est constituée d'une ou plusieurs pièces de théâtre (*play*) concurrentes (leur réalisation est en parallèle). Un *play* est constituée d'un ou de plusieurs actes (*act*) séquentiels. Chaque acte fait référence à un ou plusieurs rôle-part concurrents également; ces rôle-part permettent d'associer un rôle avec une activité ou à une structure d'activité.

Au niveau B, une méthode peut contenir des conditions (condition) ou règles (par exemple du type If-Then-Else) qui permettent alors de raffiner la visibilité des activités et des environnements pour les personnes ou rôles. Ces conditions définissent alors des expressions booléennes sur les propriétés de ces personnes ou rôles. Une propriété peut être regroupée dans des property-groups qui sont un exemple de propriétés globales (global éléments) pouvant être définies. Les propriétés peuvent être de différents types, locales comme globales pour les personnes comme pour les rôles.

Une notification peut être générée par un résultat et peut ainsi créer une nouvelle activité disponible pour un rôle donné (la personne recevant la notification n'est pas forcément celle qui l'a émise). Par exemple, si un apprenant termine une activité (ce qui correspond au fait d'avoir atteint un résultat), alors un autre apprenant ou bien un enseignant peut se voir attribuer une nouvelle activité en conséquence. Ce mécanisme peut aussi être utilisé pour des conceptions d'apprentissage dans lesquelles la ressource d'une activité est dépendante du type de résultats obtenus à des activités précédentes (utiles pour des apprentissages coopératifs).

Deux rôles explicites sont spécifiés dans IMS-LD : *learner* et *staff*. Ces rôles peuvent être spécialisés en sous-rôles d'après la spécification d'IMS-LD mais cela reste à la charge du concepteur d'apprentissage.

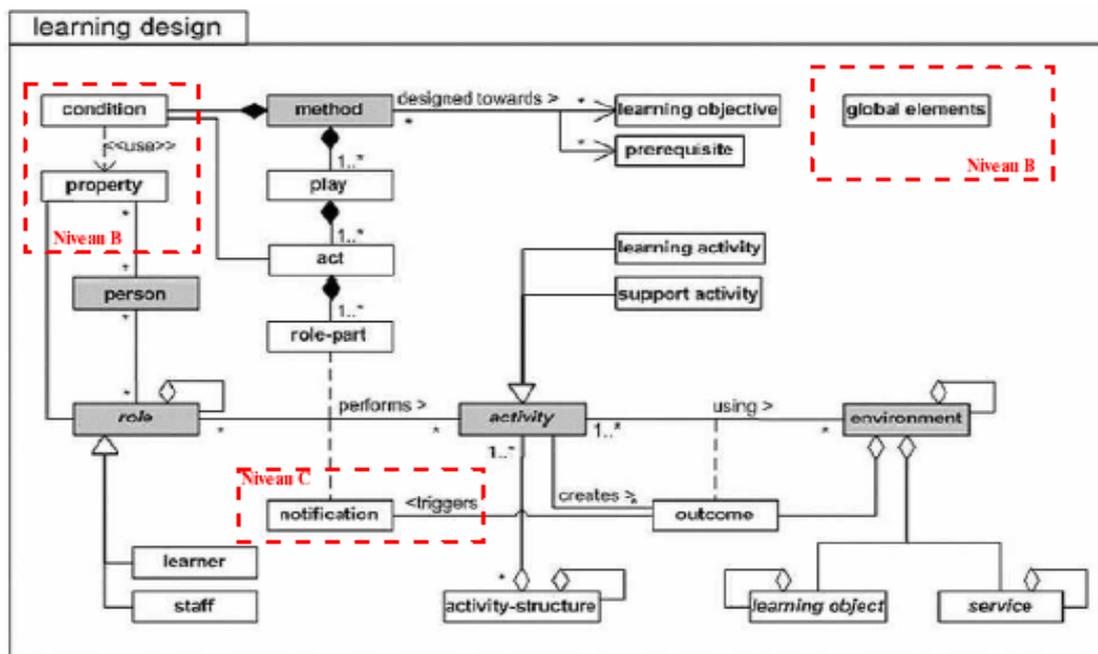


Figure 1.4 Modèle conceptuel d'IMS Learning Design (Lafocade, 2004).

Nous avons vu que les activités pouvaient être assemblées dans des structures. Celles-ci sont alors une agrégation d'un ensemble d'activités dans une unique structure; les activités

sont alors associables à un rôle grâce au rôle-part. Une structure peut modéliser une séquence d'activités, qui devront tous être complétés dans l'ordre fourni, ou bien une sélection d'activités pour laquelle le rôle associé décidera du nombre d'activités qu'il voudra réaliser. Les structures d'activités peuvent également faire référence à d'autres structures comme à d'autres unités d'apprentissage externes.

Les environnements peuvent contenir deux types de base :

- Des objets d'apprentissage localisés (typiquement spécifiés avec une URL accompagnée parfois de métadonnées)
- Des services génériques relatifs à des services concrets disponibles à l'exécution. Ces services n'ont pas d'URL associée à la conception. L'URL sera donnée quand la situation d'apprentissage conçue (Learning Design) sera instanciée à l'exécution.

Dans le but d'offrir aux développeurs des possibilités de traitement automatique par des systèmes informatisés, le consortium IMS fournit une implémentation du modèle d'information sous forme de schéma XML.

Ce langage permet de modéliser des unités d'apprentissage grâce à des concepts pédagogiques. La modélisation d'un plan de cours en IMSLD se fait par la définition des différents rôles, Activités, environnement, méthodes, propriétés, conditions et notifications. Il est utilisé pour transformer les plans de cours en unités d'apprentissage (en anglais : unit of learning (UoL) ou unit of study) décrite de façon formelle et pouvant être exécutés avec l'éditeur IMS LD basé sur un moteur. Ces unités exécutables pourront être conçues dès le début en utilisant un éditeur tel que *Copper Author* ou *Reload*.

Comme les enseignants ne peuvent pas se familiariser avec le formalisme (XML) utilisé par le langage IMSLD, ces éditeurs sont donc nécessaires pour guider les enseignants dans le processus de conception des scénarios pédagogique. Dans ce qui suit nous définissons des exemples d'outils basés sur IMSLD.

Il existe trois catégories principales de ces outils : éditeurs du Learning design, outils d'exécution et les gestionnaires du learning design (*players*) (Giacomini, 2006).

1.5.4.1 Editeurs de 'Learning design'

L'éditeur Reload LD

Reload (Reusable e_learning Object Authoring and Delivery) est un environnement développé par Phillippe Beauvoir et Paul Charples (Beauvoir et Sharples, 2005), dans le but de concevoir et importer des packages conformes avec IMSLD. Un ensemble de boutons

correspondants aux concepts d'IMSLD : rôles, activités, propriétés, environnement, ainsi qu'un ensemble *d'acts* et *role_part* constituant des *plays* sont associés à chaque unité d'apprentissage. Les trois niveaux d'IMSLD : A, B et C, sont pris en compte par la deuxième version de *Reload*

CopperAuthor:

Cet éditeur est développé par OUNL (Giacomini, 2005). Il offre une interface graphique permettant aux utilisateurs de développer et de valider des UOLs, en affichant le code XML résultant et en unifiant les unités d'apprentissages incomplètes. Il ne permet pas encore d'importer des UOLs.

1.5.4.2 Les Players

En effet, Il existe plusieurs *players*, nous définissons parmi eux (Giacomini, 2005) :

CopperCore

C'est un environnement d'exécution développé par OUNL dans le cadre du projet AFANET. Il sert à charger et valider des package IMSLD. Il permet ensuite de définir des exécutables, créer des utilisateurs et leur associer des rôles, cela se fait dans une fenêtre de commandes. L'unité d'apprentissage peut être ensuite exécutée par l'enseignant dans le player. Lorsque l'apprenant termine la réalisation d'une activité, elle est cochée et il peut passer à l'activité suivante. L'enseignant peut suivre le travail de l'apprenant.

EduBox

Ce *player* est développé par *Peros systems* et *OUNL*. Il est utilisé par ce dernier pour mettre ses cours en ligne. Des travaux en cours cherchent à intégrer *EduBox* dans le système *Blackboard* et développer une version respectant la spécification IMSLD. Ce *player* n'est pas téléchargeable.

L'unité d'apprentissage

L'utilisation d'une approche conceptuelle d'apprentissage nous permet d'organiser les unités d'apprentissage. Cette dernière assemble les ressources liées, les liens Web et plusieurs matériaux et services d'apprentissage dans un dossier compressé.

L'IMS LD est le plus approprié d'après Giacomini (Giacomini, 2005) pour la conception d'un outil d'aide à la conception de sites Web éducatifs car il prend en considération plusieurs aspects tel que : les théories et modèles d'enseignement et d'apprentissage, les domaines

d'enseignement, les rôles des acteurs, les interactions entre ces différents acteurs dans l'apprentissage et l'évolution dynamique d'un cours de formation.

Rasseneur (Rasseneur, 2006) a vu que pour la description d'une formation à distance et pour supporter l'appropriation de la formation. Le niveau A de la spécification IMSLD est la partie la plus intéressante, car à ce niveau, les modules d'enseignement sont structurés sous forme de modules, activités et sous-activités et les différents prérequis et les objectifs pédagogiques du module sont explicités. Cependant, d'après le même auteur, certains aspects de cette modélisation restent peu satisfaisants :

- L'IMS LD ne permet pas la catégorisation des objectifs et des prérequis.
- L'IMS LD ne différencie pas entre un objectif et une compétence.
- Il ne permet pas de donner les dates de début et de fin des activités.

Dans le contexte des travaux de Caron (Caron, 2006) concernant l'implémentation de dispositif pédagogique, il considère que l'utilisation d'IMSLD peut être utile pour formaliser des scénarios pédagogique beaucoup plus flexibles et réutilisables, indépendamment de toute plateforme de formation. Cependant, il peut ne pas être capable de conduire correctement à une implémentation sur une plateforme de formation. Il est difficile de décrire les spécificités d'une plateforme donnée en IMSLD et de faire correspondre entre les fonctionnalités du scénario définies en IMSLD et les fonctionnalités réellement disponibles. Pour cela, l'implantation automatique du scénario pédagogique à partir du langage IMSLD pourrait être inutile. Caron a proposé donc à adopter l'approche orientée modèles appliquée aux EIAH. Dans cette approche, l'IMS LD est considéré comme une étape intermédiaire entre la description du scénario en langage naturel et la traduction de son modèle par un langage qui prend en compte les spécifications de la plateforme de formation.

1.5.4.3 Impact d'IMSLD

Il existe aujourd'hui, plus de 30 groupes et projets internationaux qui adoptent la spécification IMSLD comme : LORNET, L'ENT 'cartable électronique'..., ou des éléments qui sont directement attachés à elle. Les trois facteurs clés démontrant la pertinence et l'impact d'IMSLD définis par plusieurs auteurs sont (Laforcade, 2004) :

Le premier facteur :

L'existence de plusieurs applications qui sont basées sur IMS LD ou qu'il intègre dans leur modèle conceptuel (LAMS, ASKLDT, MOT Plus,...).

Deuxième facteur :

L'augmentation des communautés virtuelles autour d'IMS LD, l'intérêt de ces communautés réside dans le fait qu'elles permettent aux spécifications d'améliorer l'apprentissage en ligne en adaptant l'enseignement aux plates formes virtuelles.

Troisième facteur :

L'étendue géographique d'IMS LD, partant de son pays d'origine : les Etats-Unis, vers l'Europe, le Canada, l'Australie, etc.

1.5.4.4 Les défis d'IMSLD

L'IMS LD comme les autres spécifications peut avoir quelques défis (Laforcade, 2004), parmi lesquels, on cite :

- La difficulté d'importation de lots d'information 100% conforme IMSLD bien que l'échange et l'interopérabilité de ces lots d'information pour être exécutables par plusieurs systèmes d'exécution, soient les buts les plus importants de toutes les spécifications de e-formation, ainsi que l'obtention de la plus grande flexibilité possible.
- Les outils auteur et les éditeurs sont plus centrés sur les performances techniques au lieu d'être plus orientés vers les utilisateurs finaux qui sont : l'enseignant ou le concepteur pédagogique pour garantir une meilleure utilisation de ceux-ci.
- la mise en place d'un environnement virtuel compatible à 100% avec des UOL de IMS LD, tout en facilitant l'échange et l'interopérabilité avec les éditeurs d'UOL et les services interconnectés.
- l'élaboration d'une liste la plus complète possible des modèles pédagogique pour être réutilisés directement par l'enseignant, contenant des scénarios pédagogique modélisés auparavant et stockés dans une bibliothèque. L'enseignant peut donc économiser du temps et des efforts en utilisant ces scénarios avec une personnalisation et une adaptation à ses besoins.

1.6 L'apport des EMLs

Apparus au début des années 2000 et situés dans la mouvance de *l'Instructional Design*, les langages de modélisation pédagogique apparaissent de plus en plus nécessaires aux acteurs de la formation ouverte et à distance. Un EML (*Educational Modelling Language*) est défini par (CEN ISS) comme « un modèle d'information et d'agrégation sémantique, décrivant les contenus et les processus engagés dans une "unité d'apprentissage" selon une perspective pédagogique et dans le but d'assurer la réutilisabilité et l'interopérabilité ». S'inscrivant dans ce

cadre, les travaux du consortium IMS ont abouti à la spécification "Learning Design V.1.0" (IMS LD 2003). Cette proposition, issue du langage EML développé par (Koper, 2001), fournit un cadre méthodologique de modélisation d'une Unité d'Apprentissage (UA) et vise à proposer un compromis entre d'une part la neutralité permettant la mise en œuvre d'approches pédagogiques variées et d'autre part la puissance d'expression permettant la conception précise d'une situation d'apprentissage.

1.7 Conclusion

Nous avons présenté dans ce chapitre un bref historique de l'EIAH. Ce qui nous intéresse plus particulièrement dans l'EIAH est la modélisation des situations d'apprentissage au sein de ces environnements. Pour cela, nous avons défini quelques langages dédiés à la modélisation des situations d'apprentissage : Les langages de modélisation éducatifs (EML : Educational Modeling Languages), en mettant l'accent sur le langage que nous avons utilisé dans notre travail : IMS Learning Design (IMSLD). Ces dernières années, plusieurs travaux de recherche ont été mis en place sur la personnalisation des contenus pédagogiques. Ces travaux prouvent que la personnalisation est nécessaire pour motiver les apprenants, et par conséquent, améliorer la qualité d'apprentissage. La personnalisation de l'apprentissage au sein des EIAHs constitue l'objectif principal de cette thèse. De ce fait, le chapitre suivant sera consacré pour la présentation de l'un des outils de personnalisation: les systèmes de recommandation, et la personnalisation de l'apprentissage.

Chapitre 2

Les systèmes de recommandation

2.1 Introduction

Ce chapitre a pour objectif de faire un tour d'horizon, non exhaustif, sur les systèmes de recommandation, en évoquant leur origines et leurs applications, et en décrivant les différentes approches de recommandation, les métriques d'évaluation et les problèmes les plus importants dans ce domaine. Ainsi nous définissons quelques concepts relatifs à ce domaine.

2.2 Les systèmes de recommandations

Généralement, on peut accéder à l'information sur Internet en utilisant deux méthodes. La première méthode consiste à accomplir une recherche active de l'information par l'intermédiaire d'outils de recherche d'informations. Ces outils ont pour objectif de fournir à l'utilisateur les documents qui permettent de satisfaire ses besoins en informations (selon la requête de recherche formulée) (Belkin et Croft, 1992). La deuxième méthode repose sur l'usage des systèmes de filtrage d'information ou systèmes de recommandation. Les systèmes de recommandation (ou systèmes de filtrage d'informations) visent à proposer à un utilisateur actif des items susceptibles de l'intéresser. Ces items peuvent être des articles, des livres, des films, des restaurants.... À l'opposé des outils de recherche d'informations, le filtrage d'information ne requiert pas une formulation systématique du besoin informationnel de l'utilisateur. Ainsi, ces systèmes permettent notamment de faciliter l'accès à l'information. Les systèmes de filtrage d'informations ont pour but de faire parvenir aux utilisateurs les informations jugées pertinentes.

En 1992, le centre de recherche *Xeros* aux états unis a développé, l'un des premiers systèmes de recommandation basé sur le filtrage collaboratif: *Tapestry*. Ce système a été intégré à une application de mail électronique pour la recommandation des listes de diffusions aux utilisateurs. Dans ce système, les utilisateurs annotent les listes de diffusion par des tags pour être exploité dans la proposition des listes de diffusion qui sont pertinentes à chaque utilisateur. En 1994 et en 1995, d'autres systèmes ont vu le jour tels que le système de recommandation d'articles d'actualité et de films développé par "GroupLens" (Resnick et al., 1994) et le système de recommandation de musique "Ringo" proposé par (Shardanand et Maes, 1995). Ces deux systèmes sont également basés sur le FC.

Par la suite, des systèmes de recommandation ont émergés dans différents domaines d'applications. Nous pouvons en citer (Esslimani, 2010):

– les systèmes de recommandation de films, tels que : MovieLens (Herlocker et al., 1999) et Eachmovie (Breese et al., 1998), (<http://www.grouplens.org>)

- les systèmes de recommandation de livres (Bookcrossing (Ziegler et al., 2005)), (http://www.informatik.uni-freiburg.de/_cziegler/BX).
- les systèmes de recommandation de musique (LastFM (Jäschke et al., 2007)), (<http://www.lastfm.fr>)
- les systèmes de recommandation d'articles d'actualités (Billsus et al., 2002),
- les systèmes de recommandation de blagues (Jester (Goldberg et al., 2001)), (<http://eigentaste.berkeley.edu>).
- les systèmes de recommandations introduits sur des sites e-commerce (Amazon (Linden et al., 2003)), (<http://www.amazon.com>)
- les systèmes de recommandation de restaurants (Burke, 2002),
- les systèmes de recommandation intégrés aux Extranets documentaires (l'Extranet documentaire du Crédit Agricole (Bertrand-Pierron, 2006)),
- les systèmes de recommandations intégrés aux moteurs de recherche (le moteur de recherche d'AOL (Pass et al., 2006)), (<http://www.gregsadetsky.com/aol-data>)
- les systèmes de recommandations implémentés sur des sites de recrutement (Job-Finder (Rafter et al., 2000)),
- les systèmes de recommandations de citations bibliographiques (McNee et al., 2002) (Cosley et al., 2002).

2.3 Typologie des systèmes de recommandation

Dans la littérature, nous trouvons différentes formes de recommandation selon les données à recommander, les informations disponibles et bien évidemment selon l'objectif visé. Dans cette thèse, nous avons adopté la typologie définie par (Poirier, 2011).

2.3.1 La recommandation éditoriale

La recommandation éditoriale est généralement adoptée lorsque le système ne dispose d'aucune information sur le visiteur du site. Cette forme de recommandation a pour principal objectif d'attirer rapidement l'œil de l'utilisateur novice afin de lui procurer l'envie de parcourir une partie du catalogue, comme le ferait la une d'un journal. Pour cela, on peut présenter en premier niveau les produits les plus populaires, les nouveautés, les articles les mieux notés, les promotions, etc.

alapage.com

NOUVEAUTÉS LIVRES

 Katiba Jean-Christophe Rufin 24h Livr. Gratuite 19,00 € ! Remise : - 5% > Voir l'article	 Orages ordinaires William Boyd 24h Livr. Gratuite 20,71 € ! Remise : - 5% > Voir l'article	 Au-delà des pyramides Douglas Kennedy 24h Livr. Gratuite 18,53 € ! Remise : - 5% > Voir l'article	 La Beauté et l'Enfer. Ecrits 2004.. Roberto Saviano 24h Livr. Gratuite 19,95 € ! Remise : - 5% > Voir l'article
---	---	--	--

[> Voir plus de nouveautés livres](#)

MEILLEURES VENTES

 1 Les écureuils de Central Park sont .. Katherine Pancol 24h Livr. Gratuite 22,71 € ! Remise : - 5% > Voir l'article	 2 La Fille de papier Guillaume Musso 24h Livr. Gratuite 18,91 € ! Remise : - 5% > Voir l'article	 3 Je ne sais pas maigrir. Edition 2.. Pierre Dukan 24h Livr. Gratuite 6,18 € ! Remise : - 5% > Voir l'article	 4 Le crépuscule d'une idole. L'affa.. Michel Onfray 24h Livr. Gratuite 20,90 € ! Remise : - 5% > Voir l'article
---	---	--	--

Figure 2.1 Exemple de recommandation éditoriale présente sur le site Alapage (Poirier, 2011)

Cette approche est adoptée par tous les sites de vente par correspondance comme par exemple, le site : Alapage (<http://www.alapage.com>) (voir figure 2.1).

2.3.2 La recommandation sociale

Avec cette méthode, les internautes proposent les informations jugées pertinentes pour d'autres internautes. Il peut s'agir de simples utilisateurs, comme sur Youtube (<http://www.youtube.com>) ou *Flixster* (<http://www.flixter.com>), ou de consommateurs comme sur Amazon, *Priceminister* (<http://www.priceminister.com>) ou encore le site de la *Fnac* (<http://www.fnac.com>). Ce type de recommandations repose sur le principe du bouche à oreille. Sur *Flixster* par exemple, les utilisateurs peuvent proposer des recommandations à l'intérieur de leur réseau social en transmettant leurs évaluations à leurs amis.

La création des listes de «coups de cœur» par les utilisateurs est l'une des solutions utilisée par certains sites. Ces listes sont ensuite utilisées pour construire les profils des produits, comme sur *Amazon*, ou les profils des utilisateurs, comme sur *Youtube*. Ces deux exemples sont illustrés par la figure 2.2.



Figure 2.2 Exemple de recommandations sociales proposées par *Amazon* et *Youtube* (Poirier, 2011).

Sur *Myspace*, (<http://www.myspace.com>) le principe est un petit peu différent. Ce sont les artistes, soit les producteurs de contenu, qui recommandent d'autres artistes. Un artiste conseille d'autres artistes dont il apprécie les œuvres et réciproquement, ce qui offre à chacun d'entre eux l'opportunité d'être vu par plus de personnes. Les auditeurs qui le désirent peuvent ainsi découvrir des nouveautés en se promenant de liens en liens (Poirier, 2011).

2.3.3 La recommandation contextuelle

La proposition des items similaires à l'item consulté est le principe de la recommandation contextuelle. Cette approche peut utiliser des techniques de rapprochement des items simples. On peut par exemple sélectionner des items du même univers, du même auteur, du même réalisateur, du même compositeur, de même couleur, etc. Elle peut également utiliser des techniques de rapprochement plus complexes, comme avec les méthodes basées sur les usages. *Youtube* ou *Amazon* sont des exemples parfaits de cette technique. Lorsqu'un item est consulté par un internaute, par exemple un disque de *Georges Brassens*, le système suggère une liste d'items qui ont été appréciés par les utilisateurs ayant également apprécié ce disque de *Brassens* (Poirier, 2011). Un exemple de recommandation contextuelle basée sur les usages est montré par la figure 2.3.



Figure 2.3 Exemple de recommandations contextuelles sur le site de la *Fnac* (Poirier, 2011).

On trouve également des méthodes basées sur le contenu, c'est-à-dire que les recommandations sont générées en fonction des concepts (ou mots clés) décrivant le contenu des items. Les descripteurs des items sont créés en appliquant une analyse humaine ou automatisée, *Flickr* (<http://www.flickr.com>) ou *IMD* (<http://www.imdb.com>) par exemple utilisent les tags (ou étiquettes en français) pour rapprocher les items entre eux. *Pandora* (<http://www.pandora.fm>) compare les morceaux musicaux de son catalogue en analysant plus de 400 caractéristiques sonores (Poirier, 2011).

2.3.4 La recommandation personnalisée

Ce type de recommandation vise à identifier, pour un utilisateur particulier, les contenus jugés pertinents. Les systèmes basés sur la recommandation personnalisée génèrent des recommandations en fonction de l'utilisateur qui en est bénéficiaire et non pour un groupe d'utilisateurs comme c'est le cas avec les moteurs de recherche par exemple ou encore avec les autres types de recommandation décrit au dessus.

Le système de recommandation remplacera un commerçant en chair et en os qui écoute les préférences de ses clients, au cas par cas, afin de les guider à retrouver les bons produits. Le client y gagne en temps car il n'a pas besoin de fouiller tout le catalogue et le système peut également lui permettre de découvrir de nouvelles choses. Quant au fournisseur, lorsque les recommandations générées sont de bonne qualité, cela peut entraîner un gain de confiance du client et ainsi favoriser la fidélité, ce à quoi aspire toute stratégie marketing. La recommandation permet également au fournisseur de valoriser son catalogue en guidant l'utilisateur vers des contenus autres que les plus populaires (Poirier, 2011). La figure 2.4

présente l'exemple type de recommandations personnalisées que le site *Allociné* propose à ses utilisateurs : le système génère des prédictions pour toutes les notes manquantes au niveau de la matrice de notes "Utilisateur x Item", c'est-à-dire pour les films ou séries que l'utilisateur n'a pas encore notés. Les items ordonnancés permettent alors à l'utilisateur de faire ses choix sur une partie du catalogue et non sur le catalogue tout entier. Ce système en l'occurrence est un service offert à l'utilisateur ayant pour unique objectif de fidéliser l'utilisateur, le site *Allociné* (<http://www.allocine.fr>) ne distribuant pas directement les items présents dans le catalogue (Poirier, 2011).

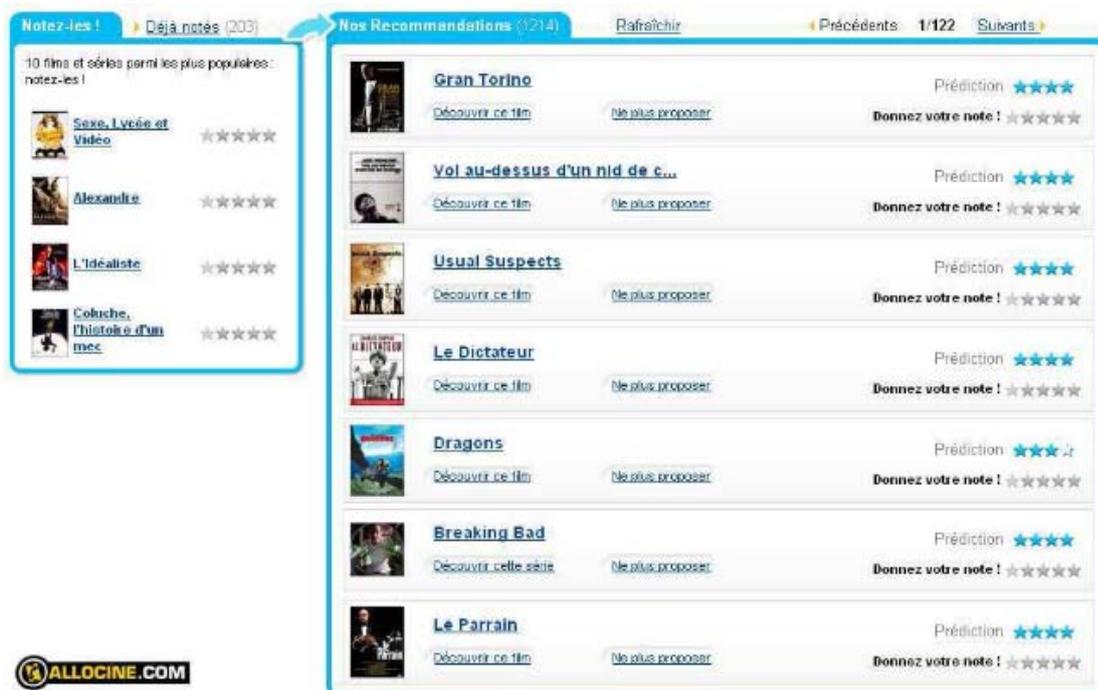


Figure 2.4 Exemple de recommandations personnalisées proposées par *Allociné* (Poirier, 2011).

La recommandation personnalisée est certainement l'approche la plus célèbre du moment. Il intéresse autant les laboratoires de recherche pour toutes les problématiques qu'il entraîne que le monde industriel pour les possibilités marketing qu'offre ce type de systèmes (Poirier, 2011).

2.3.4.1 L'approche basée contenu :

Le filtrage basé contenu ou le filtrage cognitif (Malone et al, 1987) prend en compte les caractéristiques du contenu d'un item (document) et les besoins des utilisateurs en termes d'information. En d'autres termes, cette technique est basée sur la recommandation des items similaires à ceux préférés par l'utilisateur dans le passé (Pazzani and Billsus, 2007; Malone et al., 1987). Les systèmes basés contenu sont souvent utilisés pour recommander des documents textuels (Adomavicius et Tuzhilin, 2005).

Les techniques du filtrage basé contenu proposent aux utilisateurs des items via les approches de classification de données ou les méthodes de plus proches voisins. Les approches de classification associe à chaque utilisateur un classifieur représentant son profil. Le *classifier* compare les nouveaux documents avec le profil de l'utilisateur et recommande ceux qui sont les plus proches (Adomavicius et Tuzhilin, 2005). Plusieurs techniques de classification ont été adoptées par les systèmes de recommandation basé-contenu dont les plus utilisées sont : les réseaux de neurones, les arbres de décision, les réseaux bayésiens. Par exemple : Re : Agent et personal news recommender proposé par Jennings (Boone, 1998) sont basées sur les réseaux de neurone. RIPPER (Cohen 1995, Cohen 1996), MovieLens (Good et al.,1999), Recommender (Basu et al., 1998), and WebSIFT(Cooly et al. 1999) sont basés sur les règles d'induction. News Dude (Billsus et Pazzani, 1998), Personal WebWatcher (Mladenic, 1999), la technique de filtrage basée catégories de Sollenborn (Sollenborn et Funk, 2002) sont basés sur les réseaux Bayésien.

A l'inverse, les méthodes de plus proche voisin compare le contenu d'un nouvel item avec ceux évalués par l'utilisateur. La pertinence de l'item sera déterminée selon les items les plus proches (Montaner et al., 2003). Les systèmes de recommandation les plus connus sont : WEBSSELL (Cunningham et al., 2001), Daily learner (Billsus et al., 2000), LaboUr (Schwab et al.,2000).

Nous pouvons utiliser la technique de recommandation basée sur le contenu pour la recommandation de pages Web, de films, d'articles actualités, de restaurants, etc. Si nous prenons l'exemple d'un système de recommandation d'articles scientifiques basé sur le contenu, lorsqu'un utilisateur a tendance à consulter souvent des articles portant sur le domaine de la génétique, le système lui proposera des recommandations liées à la génétique. En effet, ces articles disposent de mots-clés communs tels que : "ADN", "gène" ou "protéine". Il est à signaler que ces mots-clés sont généralement soit extraits sur la base d'une indexation automatique, soit attribués manuellement (Esslimani, 2010).

Pour ce qui est des systèmes de recommandation de films ou de restaurants, le contenu est plutôt structuré et représenté par des métadonnées définies au préalable et valables pour tous les items (Pazzani et Billsus, 2007).

Dans le cadre de la technique basée sur le contenu, la mesure TF-IDF (“Term Frequency-Inverse Document Frequency”) (Salton, 1989) est l’une des mesures les plus connues pour l’analyse du contenu. Cette mesure statistique permet de décrire l’importance d’un mot dans un document ou dans un item faisant partie d’une collection ou d’un corpus.

Le principe de cette mesure est que les mots-clés paraissant dans beaucoup d’items ne permettent pas de distinguer un item pertinent d’un autre qui ne l’est pas. Or, les mots clés qui sont rares et communs à quelques items ne définissent plus la similarité de contenu ainsi que la pertinence d’un item par rapport à un autre (Esslimani, 2010). En général, les avantages de l’approche basée contenu sont:

- L’utilisateur peut avoir une idée sur le contenu de chaque document (item) et connaître pourquoi tel item est important.
- Cette approche permet de pallier au problème de démarrage à froid. Le démarrage à froid se traduit notamment par l’introduction d’un nouvel item au système de recommandation. Dans le cas d’un système de recommandation basé sur le filtrage collaboratif, ce nouvel item ne sera pas incorporé aux recommandations, puisque le système ne dispose d’aucune information sur cet item. Ainsi, grâce à l’analyse de contenu, cet item peut être intégré aux recommandations proposées à un utilisateur actif.

Néanmoins, la technique basée sur le contenu présente quelques limites, notamment:

- la sur spécialisation des recommandations est un problème majeur de cette technique. En effet, les items recommandés sont toujours similaires et identiques (en termes de contenu) aux items précédemment consultés par l’utilisateur. Un item proche avec une description thématique différente sera automatiquement exclu de la recommandation, alors qu’il pourrait intéresser l’utilisateur.
- La représentation des items est toujours limitée aux descriptions ou aux attributs qui leur sont associés. Par conséquent, afin d’avoir un ensemble suffisant d’attributs, il est nécessaire soit de prétraiter le contenu pour permettre une extraction automatique d’attributs, soit d’attribuer les descriptions manuellement (Shardanand et Maes, 1995). Dans les deux cas, l’extraction d’attributs demeure une opération fastidieuse surtout lorsqu’il s’agit d’items multimédia tels que : les images, les documents audio et vidéo,

etc. De ce fait, certains aspects pertinents du contenu peuvent être négligés, ce qui peut avoir un impact sur la qualité des recommandations (Esslimani, 2010).

2.3.4.2 L'approche collaborative

Le filtrage collaboratif est basé sur L'hypothèse : le bouche à oreille. Plus précisément, Le filtrage collaboratif vient renforcer l'idée que les personnes à la recherche d'informations devraient pouvoir se servir de ce que d'autres ont déjà trouvé et évalué. Par exemple, les personnes, qui veulent regarder un film ou lire un livre, demandent à leurs amis leurs opinions. Donc, dans le filtrage collaboratif, la sélection des documents à proposer à un utilisateur ne dépend plus du contenu du document (filtrage basé sur le contenu), mais des appréciations faites par des utilisateurs ayant des préférences similaires. Ainsi, si deux utilisateurs u_1 et u_2 ont évalué un certain nombre de documents de façon similaire, il y a de fortes chances que u_1 aime ce que u_2 aime, et inversement. Donc les documents que u_1 a aimés peuvent être recommandés à u_2 et inversement. De la sorte, cette approche résout un problème important rencontrée par l'approche de filtrage basé sur le contenu, à savoir le traitement de documents multimédias (Adomavicius et Tuzhilin, 2005 ; Sarwar et al., 2000a).

L'algorithme général d'un système de filtrage collaboratif est illustré par la figure 2.5. Typiquement, les étapes de cet algorithme sont les suivantes (Adomavicius et Tuzhilin, 2005 ; Berrut et Denos, 2003), (Zaier, 2010).

- Proposer à l'utilisateur un ensemble de documents pour les évaluer.
- Construire le profil de l'utilisateur en se basant sur ses appréciations;
- utiliser ce profil pour aider l'utilisateur dans ses prochaines recherches d'information.

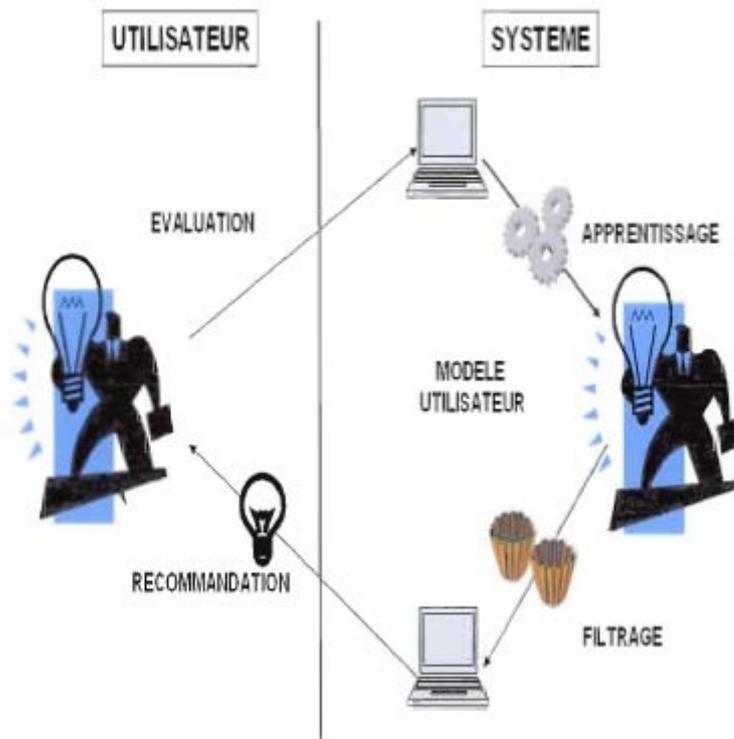


Figure 2.5 Le modèle général du filtrage collaboratif (Berrut et Denos, 2003).

a. Méthodes basées sur la mémoire

L'approche basée sur la mémoire calcule les recommandations en exploitant les évaluations explicites des utilisateurs sur les items (sous forme de notes par exemple) (Sarwar et al., 2001). Des techniques statistiques sont utilisées par cette approche dans le but d'identifier des utilisateurs similaires ayant, sur un même ensemble d'items, des appréciations similaires à celles de l'utilisateur actif. Une fois le groupe des utilisateurs similaires identifiés, l'approche basée sur la mémoire adopte différents algorithmes afin de combiner les évaluations des voisins et générer des recommandations à l'utilisateur actif (McLaughlin et Herlocker, 2004). Le FC basé sur la mémoire recherche les "k plus proches voisins" (k Nearest Neighbors "kNN") (Resnick et al., 1994), i.e. les k voisins les plus similaires à l'utilisateur actif, dans le but de proposer des recommandations fiables. Ces utilisateurs similaires sont identifiés à partir d'une évaluation de la similarité des appréciations sur les items communs à l'utilisateur actif et les autres utilisateurs.

Dans un système de FC basé sur la mémoire, les données sont représentées sous forme d'une matrice "Utilisateur x Item", où les lignes représentent les utilisateurs $U = \{u_1; \dots, u_m\}$ et les colonnes constituent les items $I = \{i_1, \dots, i_j\}$. Les utilisateurs fournissent leurs opinions concernant les items sous forme de notes v . Pour un utilisateur actif u_a (par exemple Jean) n'ayant pas exprimé son avis concernant un item i_k (le film "Les visiteurs"), le système

recherche les utilisateurs voisins les plus proches notés U_a (parmi Rose, Ryan et Hélène ayant noté le film “Les visiteurs” et qui ont déjà co-noté le film “Pulp Fiction” avec Jean) et utilisent leurs opinions pour prédire la note manquante $v(u_a; i_k)$ ($v(\text{Jean}, \text{Les visiteurs})$).

Ainsi, deux phases essentielles en FC basé sur la mémoire sont distinguées: la phase d’identification du voisinage et la phase de calcul des prédictions. Ces deux phases sont décrites dans les sous-sections suivantes (Esslimani, 2010).

➤ **Calcul du voisinage**

Le FC basé sur la mémoire utilise plusieurs mesures dans le but d’évaluer les similarités d’appréciations entre utilisateurs et identifier les utilisateurs voisins (les plus proches). Parmi ces mesures nous pouvons citer : le coefficient de corrélation de Pearson (Herlocker et al., 1999), la mesure basée sur le cosinus (Sarwar et al., 2000b), la corrélation de Spearman (Resnick et al., 1994), “*Mean squared difference*” (qui représente une mesure de dissimilarité) (Shardanand et Maes, 1995), etc. Les mesures les plus connues sont le coefficient de corrélation de *Pearson* et la mesure basée sur le cosinus. Ces deux mesures sont populaires grâce à leur contribution à la performance des systèmes de recommandation (Anand et Mobasher, 2005).

Ces deux mesures sont décrites (Anand et Mobasher, 2005) ci-dessous. $CorrP(u_a; u_b)$ et $Cos(u_a; u_b)$ représentent les similarités calculées respectivement avec la mesure de corrélation de *Pearson* et la mesure basée sur le cosinus, entre deux utilisateurs u_a et u_b . I_a et I_b représentent respectivement l’ensemble des items notés par u_a et u_b . $\overline{v(u_a)}$ dénote la moyenne de notes de u_a et $v(u_a; i)$ indique la note de u_a sur l’item i . I_c désigne les items co-notés (notés en commun) entre l’utilisateur actif u_a et l’utilisateur u_b (Esslimani, 2010).

– Le coefficient de corrélation de *Pearson* : cette mesure est calculée par l’équation (1) (Anand et Mobasher, 2005). Lorsque $CorrP(u_a; u_b)$ vaut 1, cela signifie que les utilisateurs u_a et u_b sont fortement corrélés. Or, si $CorrP(u_a; u_b)$ vaut -1, cela signifie que u_a et u_b ont des évaluations totalement opposées. Quand cette corrélation vaut 0, aucune relation n’existe entre les deux utilisateurs (Esslimani, 2010).

$$CorrP(u_a, u_b) = \frac{\sum_{i \in I_c} (v(u_a, i) - \overline{v(u_a)})(v(u_b, i) - \overline{v(u_b)})}{\sqrt{\sum_{i \in I_c} (v(u_a, i) - \overline{v(u_a)})^2 \sum_{i \in I_c} (v(u_b, i) - \overline{v(u_b)})^2}} \quad 1$$

– La mesure basée sur le cosinus : cette mesure est très fréquemment utilisée dans le domaine de la recherche d’information. Dans ce contexte, elle consiste à évaluer la similarité entre deux documents représentés par des vecteurs de poids (fréquence d’apparition des mots), en

calculant le cosinus de l'angle formé par ces deux vecteurs (Salton et McGill, 1983) (Esslimani, 2010).

En FC, cette mesure peut être utilisée pour mesurer la similarité entre deux utilisateurs u_a et u_b en calculant le cosinus de l'angle entre les vecteurs correspondant à ces deux utilisateurs sur la base de l'équation (2) (Breese et al., 1998), en prenant en compte les items co-notés Ic. La valeur calculée par la mesure cosinus est comprise entre 0 et 1.

$$\text{COS}(u_a, u_b) = \frac{\sum_{i \in I_c} v(u_a, i) * (u_b, i)}{\sqrt{\sum_{i \in I_a} v(u_a, i)^2 \sum_{i \in I_b} v(u_b, i)^2}} \quad 2$$

L'inconvénient majeur des deux mesures *Pearson* et *Cosinus*, est que le calcul des similarités devient non fiable voire impossible, lorsque le nombre d'items co-notés entre utilisateurs est limité. Certaines extensions ont été proposées notamment par (Breese et al., 1998), afin de remédier à ce problème telle que "La note par défaut "consistant à attribuer une valeur par défaut à une note manquante. Mais l'enjeu à ce niveau est de savoir quelle valeur par défaut choisir (appréciation positive, négative ou bien neutre) et d'évaluer son impact sur le calcul des similarités.

Par ailleurs, (Breese et al., 1998) ont proposé, en vue d'améliorer la performance des systèmes de recommandation exploitant le FC basé sur la mémoire, d'utiliser (Esslimani, 2010) :

- "L'amplification de cas" permettant de transformer les similarités en amplifiant les valeurs proches de 1 et en pénalisant celles qui sont proches de 0, afin d'attribuer un poids important aux utilisateurs fortement similaires à l'utilisateur central.
- "La fréquence inverse utilisateur" inspirée de la méthode IDF ("Inverse Document Frequency"). Cette méthode repose sur l'hypothèse suivante : les items appréciés par un grand nombre d'utilisateurs sont jugés not pertinents pour le calcul des similarités comparés à ceux qui sont évalués par un nombre limité d'utilisateurs.

Ainsi, chaque note est convertie en la multipliant par la fréquence inverse utilisateur qui est équivalente à $\log \frac{n}{n_{i_k}}$, n désigne le nombre total des utilisateurs et n_{i_k} le nombre d'utilisateurs ayant noté i_k (Esslimani, 2010).

➤ Calcul de prédiction

Cette phase est importante dans la mesure où l'objectif de tout système de FC est le calcul des prédictions pour générer des recommandations pertinentes à un utilisateur actif. La

méthode la plus utilisée pour le calcul de ces prédictions est la “somme pondérée” (Herlocker et al., 1999). Selon l’équation (3) (Herlocker et al., 1999), cette méthode prend en considération les plus proches voisins U_a (corrélés avec l’utilisateur actif) ayant déjà noté l’item i_k , pour calculer la prédiction de la note de u_a sur i_k notée $Pred(u_a; i_k)$. $Sim(u_a; u_b)$ est la valeur de similarité entre u_a et un voisin u_b ($u_b \in U_a$) et peut être instanciée par les similarités calculées en utilisant coefficient de *Pearson* ($CorrP(u_a; u_b)$) ou bien la mesure basée sur le cosinus ($Cos(u_a; u_b)$) (Esslimani, 2010).

$$Pred(u_a, i_k) = \overline{v(u_a)} + \frac{\sum_{u_b \in U_a} Sim(u_a, u_b) * (v(u_b, i_k) - \overline{v(u_b)})}{\sum_{u_b \in U_a} Sim(u_a, u_b)} \quad 3$$

b. Les méthodes basées modèle

Les méthodes basées modèle ont été mise en place afin de pallier aux problèmes des méthodes basées sur la mémoire, dont notamment : la non robustesse au manque de données ainsi que le non passage à l’échelle (Sarwar et al., 2000b). Les méthodes basées sur un modèle utilisant notamment les techniques de réduction de dimensionnalité ou le *clustering* ont été proposées pour faire face à ces deux problèmes. Ces méthodes visent à écarter les utilisateurs ou les items non représentatifs. Ainsi l’espace de représentation utilisateur-item est plus réduit et le taux de données manquantes est moins important comparé à l’espace de représentation original. Les utilisateurs similaires peuvent ainsi être identifiés dans cet espace réduit, ce qui permet de garantir le passage à l’échelle.

Dans le cadre des méthodes basées sur un modèle, le processus de FC consiste à construire des modèles (généralement en hors ligne “*off-line*”) en exploitant les données collectées sur l’utilisateur et/ou sur l’item. Les prédictions sont alors générées sur la base des modèles construits. Ces prédictions seront proposées à l’utilisateur actif lors de son interaction avec le système. Le processus de construction du modèle est basé sur les techniques d’apprentissage automatique, telles que : le *clustering*, les réseaux bayésiens, les arbres de décision, etc.

L’approche collaborative présente quelques avantages parmi lesquels on cite (Esslimani, 2010):

- La possibilité d’indexer tout type de document (vidéo, son, multimédia,...), du fait que le contenu n’est pas pris en compte et le document est seulement connu par son identifiant.

- Cette approche génère des recommandations de qualité par rapport à l'approche basée contenu car elle permet de distinguer entre un document bien écrit et un autre mal écrit si les deux utilisent les mêmes termes.

Tapestry et *GroupLens* sont un exemple des systèmes de recommandations basé sur l'approche collaborative. *Tapestry* (Goldberg et al. 1992, Resnick et Varian 1997) génère des recommandations en se basant sur les opinions explicites des gens appartenant à des communautés. *GroupLens* exploite les évaluations des utilisateurs du *Usenet newsgroups* pour prédire l'importance d'une nouvelle histoire pour l'utilisateur actif (Weng, 2008).

Malgré sa popularité, l'approche collaborative souffre aussi de quelques défauts (Weng, 2008):

- Démarrage à froid (Middleton et al., 2002) : lors de l'ajout d'un nouvel item le système n'est pas capable de recommander un item avant qu'il soit suffisamment apprécié par les utilisateurs. Autrement, les nouveaux documents ne peuvent être diffusés que si un minimum d'informations les concernant est collecté à partir de l'avis de l'un des utilisateurs.

Un autre cas du démarrage à froid est l'ajout d'un nouvel utilisateur. Les utilisateurs commencent avec un profil vide et doivent le constituer à partir de zéro. Même avec un profil de démarrage, une période d'apprentissage est toujours nécessaires avant que le profile ne reflète concrètement les préférences de l'utilisateur. Pendant cette période le système ne peut pas filtrer efficacement pour le compte de l'utilisateur.

- L'absence de préférences communes entre les utilisateurs conduit nécessairement à des recommandations pauvres (Mantaner et al., 2003)
- le problème de *scalability*: les systèmes de recommandations doit être utiliser par un grand nombre d'utilisateurs pour qu'il soit efficace comme il requièrent une grande masse de données fournies par chacun de ces utilisateurs, alors que les recommandations sont limitées aux items évalués par ces utilisateurs (sarwar et al., 2000).

2.3.4.3 Filtrage hybride

L'approche de filtrage hybride vise à tirer profit des avantages des deux approches précédentes, en résolvant les problèmes qui leur sont liés. En fait, ces deux approches paraissent complémentaires. Les chercheurs du domaine supposent que la combinaison des deux méthodes pourrait limiter les problèmes qui leur sont liées. D'où l'émergence de plusieurs techniques d'hybridation dont l'objectif consiste à combiner les deux approches

(filtrage collaboratif et filtrage basé sur le contenu) de manière efficace (Adomavicius et Tuzhilin, 2005 ; Burke, 2002).

Les récents travaux dans ce domaine visent à développer des algorithmes hybrides de plus en plus efficaces. Burke (2002) a identifié sept types de méthodes d'hybridation (présentées par le tableau 2.1). Toutes ces méthodes reposent sur 2 approches principales (Adomavicius et Tuzhilin, 2005). La première approche est basée sur des méthodes basées sur la mémoire « *Memory-based* » pour calculer les recommandations (Billsus et Pazzani, 2000 ; Good et al., 1999).

Les méthodes basées sur les modèles « *Model-based* » sont utilisées par la deuxième approche pour déterminer les recommandations (Basu et al., 1998), (Zaier, 2010).

Méthodes d'hybridation	Description
Pondéré	Les résultats pondérés de plusieurs techniques de recommandation sont combinées pour produire une nouvelle recommandation.
Permutation	Le système permute entre les différentes techniques de recommandation selon le résultat de la recommandation.
Mixte	Les recommandations de plusieurs techniques sont présentées en même temps.
Combinaison	Différentes techniques de recommandation sont combinées en un unique algorithme de recommandation.
En cascade	. système de recommandation raffine les résultats fournis par un autre système
Augmentation	Le résultat « output » d'une technique de recommandation est utilisé comme données en entrée « input » pour l'autre technique.
Méta-niveau	Le modèle appris par une technique de recommandation est utilisé comme données en entrée « input » pour l'autre technique.

Tableau 2.1 méthodes d'hybridations (Burke, 2002).

2.3.4.4 L'approche démographique

Cette approche exploite les informations démographiques des utilisateurs (éducation, âge, occupation,...) pour générer des recommandations (Krulwich et Burkey, 1996). Ces informations sont utilisées pour relier un item aux utilisateurs qui peuvent l'apprécier. Dans le cas où un utilisateur intéressé par la culture britannique par exemple, le système lui proposera

des articles prenant en considération les informations démographiques suivantes (Weng, 2008) :

_ l'âge et le niveau culturel de l'utilisateur. L'utilisateur Est-il un étudiant et il a seulement besoin d'informations générales sur la culture australienne pour faire son devoir de maison, ou un enseignant universitaire et son travail de recherche requière des informations plus riches.

_ la nationalité de l'utilisateur. Est-il capable de lire des documents en anglais ?

La recommandation des mêmes documents aux utilisateurs ayant des profils démographiques similaires alors que ces utilisateurs peuvent avoir des préférences différents, est l'inconvénient majeur du filtrage démographique. Outre, l'évolution des profils des utilisateurs est un autre problème de cette approche car ils ne s'adaptent pas aux éventuels changements dans les préférences des utilisateurs (Montaner et al., 2003) comme ils ne sont pas prises en compte.

2.3.4.5 L'approche à base de connaissance

Des modèles de décision sont employés par ce type de système pour faire des inférences des besoins et préférences d'un utilisateur. Ces systèmes sont composés de trois parties : une base de règles, une base de faits et un moteur d'inférence. Les règles de prédiction exploitées par ces systèmes représentent les connaissances définies par un expert de domaine (Schafer et al., 1999). Le moteur d'inférence sert à contrôler l'activité du système. Ces systèmes peuvent aussi utiliser les interactions passées d'un utilisateur avec le système et exploiter les cas similaires (système à base de cas) pour générer une recommandation (Lops et al., 2010).

Les avantages de ces systèmes sont (Lops et al., 2010):

- leur flexibilité, i.e leur capacité à prendre en compte tout type de règles dans le processus de recommandation.
- Un autre point fort de ces systèmes est qu'ils ne sont pas affectés par le problème de démarrage à froid comme aucune information n'est requise au démarrage du système.
- Sensitivité aux changements des préférences.
- Inclusion des attributs divers (non pas seulement de l'objet recommandé).

Les inconvénients présentés par les systèmes à base de connaissance sont (Lops et al., 2010) :

- Le cout de développement est élevé car les données nécessaires pour ce type de systèmes doivent être définies par un expert de domaine.
- Capacité de suggestion statique.
- Une gestion de connaissance est requise.

2.4 Evaluation des recommandations

2.4.1 Mesures statistiques de précision

MAE

Les mesures statistiques de précision consistent à évaluer la différence existant entre les notes prédites et les notes réellement attribuées par les utilisateurs. La MAE (*Mean Absolute Error*) (McLaughlin et Herlocker, 2004) est la mesure de précision la plus célèbre pour l'évaluation des systèmes de recommandation. La MAE mesure la moyenne d'erreur absolue (l'équation (4)) entre les notes prédites $Pred(u_a; i)$ et les notes réelles des utilisateurs $v(u_a; i)$ (McLaughlin et Herlocker, 2004). n représente le nombre total d'items présents dans le corpus test. Plus la valeur de MAE est faible, plus les prédictions sont précises et le système de recommandation est performant (Esslimani, 2010).

$$MAE = \frac{\sum_{i=1}^n |v(u_a, i) - Pred(u_a, i)|}{n} \quad 4$$

La MAE a été fréquemment utilisée pour l'évaluation des systèmes de recommandation et du FC (Shardanand et Maes, 1995) (Herlocker et al., 1999). L'avantage de la MAE est qu'elle est simple à utiliser, facile à interpréter et qu'elle est largement utilisée par la communauté scientifique, ce qui permet de positionner les approches de recommandation proposées par rapport aux travaux de recherche existants (Esslimani, 2010).

Cependant, la mesure MAE peut ne pas être appropriée pour l'évaluation de systèmes de recommandation qui génèrent des listes ordonnées de recommandation (listes TopN), (McLaughlin et Herlocker, 2004).

“*Root Mean Squared Error*” et “*Mean Squared Error*”, sont des mesures statistiques de précision qui permettent d'évaluer les prédictions numériques. Ces mesures attribuent un poids plus important aux prédictions dont l'erreur est élevée, par rapport aux prédictions précises (i.e. ces deux mesures pénalisent plus que la MAE les systèmes de recommandation générant des prédictions dont le taux de précision est faible) (Esslimani, 2010).

HMAE

Les systèmes de recommandations visent à calculer les prédictions des notes manquantes concernant le maximum de paires <utilisateur-item>. Une fois ces prédictions calculées, le système de recommandation propose seulement les items ayant les valeurs de prédiction les plus élevées. Dans ce cas, l'erreur concernant les items ayant de faibles valeurs de prédiction

n'est pas utile quant à l'évaluation de la performance des systèmes de recommandation, tandis que l'erreur relative aux items ayant des notes prédites élevées est d'une grande importance en termes d'évaluation.

La HMAE permet en effet d'évaluer les items jugés pertinents par le système, alors qu'ils ne le sont pas réellement (en comparaison avec le corpus test par exemple). Ces items sont appelés : les "faux positifs". Avec la détection des faux positifs, le système ne risque pas d'être pénalisé suite à une recommandation d'item non pertinent susceptible d'engendrer une insatisfaction chez l'utilisateur (Esslimani, 2010).

la HMAE (High MAE) (Baltrunas et Ricci, 2007) peut être utilisée pour évaluer la capacité d'un système de recommandation à proposer des items pertinents aux utilisateurs actifs. Selon l'équation (5) (Baltrunas et Ricci, 2007), la différence entre HMAE et MAE, est que HMAE prend en considération uniquement les prédictions élevées. m désigne le nombre d'items prédits avec des valeurs élevées. Plus la valeur de HMAE est faible, plus le système de recommandation est performant.

$$HMAE = \frac{\sum_{i=1}^m |v(u_a, i) - \text{Pred}^f(u_a, i)|}{m} \quad 5$$

En effet, seuls les items ayant des valeurs réelles élevées dans le corpus test, sont exploités par la HMAE. Son avantage est sa capacité à évaluer la précision des recommandations, jugées pertinentes, qui sont effectivement suggérées aux utilisateurs (Esslimani, 2010).

2.4.2 Mesures permettant l'aide à la décision

Les mesures permettant l'aide à la décision consistent à évaluer jusqu'à quel point le système de recommandation peut recommander des items susceptibles d'intéresser l'utilisateur (Adomavicius et Tuzhilin, 2005).

En d'autres termes, ces mesures calculent, dans une liste de recommandation, la proportion d'items qui sont effectivement utiles et pertinents pour l'utilisateur actif afin d'évaluer la pertinence des recommandations.

Pour les besoins d'évaluation en termes d'aide à la décision, les appréciations ou les notes des utilisateurs doivent être transformées dans le cadre d'une échelle binaire ("Aime" ou "Aime pas") afin de distinguer les items pertinents de ceux qui ne le sont pas, pour un utilisateur donné (Esslimani, 2010). Ainsi, dans le cadre de nos expérimentations, un item est considéré comme pertinent lorsqu'il dispose des valeurs les plus élevées, c'est-à-dire des valeurs entre

0.5 et 1 sur l'échelle choisie [0-1]. Nous considérons que les notes de 0 (non apprécié) à 0.4 correspondent à des items non pertinents pour l'utilisateur.

Les mesures permettant l'aide à la décision sont principalement issues du domaine de la recherche d'information. Ces mesures sont : la précision, le rappel et la mesure F1 (Herlocker et al., 2004).

Précision

La précision représente la possibilité qu'un item sélectionné soit réellement perçu comme étant pertinent par l'utilisateur actif (Anand et Mobasher, 2005). Un item sélectionné représente un item qui est proposé par le système de recommandation à l'utilisateur actif et qui est contenu en même temps dans le corpus test. Les catégories d'items classifiés selon l'intersection entre les listes de recommandation et les évaluations réelles des utilisateurs sont présentées par le tableau 2.2 (Esslimani, 2010).

A partir de ce tableau, la précision est définie par le ratio entre le nombre d'items pertinents sélectionnés N_{ps} et le nombre d'items sélectionnés par un utilisateur actif N_s . Cette métrique est calculée sur la base de l'équation (6) (Herlocker et al., 2004).

$$P = \frac{N_{ps}}{N_s} \tag{6}$$

La précision générale du système de recommandation correspond ainsi à la moyenne des précisions calculées pour chaque utilisateur actif. Plus cette précision est élevée, plus le système de recommandation est performant (Esslimani, 2010).

	Sélectionné (s)	Non Sélectionné (ns)	Total
Pertinent (p)	N_{ps}	N_{pns}	N_p
Non Pertinents (np)	N_{nps}	N_{npns}	N_{np}
Total	N_s	N_{ns}	N

Tableau 2.2 Catégories d'items basés sur l'intersection entre listes de recommandation et préférences réelles (Esslimani, 2010).

Rappel (*Recall*)

Le rappel représente la probabilité qu'un item pertinent soit sélectionné par l'utilisateur actif. Selon l'équation (7) (Herlocker et al., 2004), le rappel peut être défini comme étant le rapport entre le nombre d'items pertinents sélectionnés par l'utilisateur "Nps" et le nombre total d'items pertinents disponibles "Np".

$$R = \frac{N_{ps}}{N_p}$$

7

Comme pour la précision, le rappel relatif à la totalité du système est évalué comme étant la moyenne des rappels calculés individuellement. Il existe une mesure combinant la précision et le rappel (Sarwar et al., 2000a). Il s'agit de la mesure "F1". Elle représente la moyenne harmonique entre la précision et le rappel, suivant l'équation (8) (Sarwar et al., 2000a). La valeur de F1 varie de 0 à 1. Lorsque les scores de précision et de rappel sont équivalents, la qualité des recommandations est considérée comme parfaite (Esslimani, 2010).

$$F1 = \frac{2PR}{P+R}$$

8

Nous avons utilisé ces métriques parce qu'elles ont été utilisées pour l'évaluation de l'approche sur laquelle notre travail est basé (CSHTR : *Cold Start Hybrid Taxonomic Recommender*) (Weng, 2008), ce qui nous permet de comparer entre les résultats de notre approche et ceux du CSHTR.

2.4.3 Couverture

La capacité du système à générer des recommandations est connue sous le nom de « couverture ». En FC basé sur la mémoire, la couverture peut être évaluée par rapport à la capacité du système de recommandation à générer des prédictions pour toutes les notes manquantes au niveau de la matrice de notes "Utilisateur x Item". Elle peut être également évaluée en prenant en considération uniquement les prédictions contenues dans le corpus test (Esslimani, 2010).

En effet, L'incapacité d'un système de recommandation exploitant le FC à calculer des recommandations peut notamment être engendrée par le manque de données sur les items et/ou les utilisateurs. En effet, faute de notes provenant des voisins, le système aura des difficultés à calculer certaines prédictions.

Ainsi, un système de recommandation ne peut être performant que lorsqu'il est susceptible de calculer un nombre suffisant de prédictions concernant un maximum d'utilisateurs. Autrement dit, le système doit être capable de satisfaire les besoins des différents utilisateurs actifs présents dans le système (Esslimani, 2010).

2.4.4 Temps de calcul

Le temps de calcul peut être un facteur important dans l'évaluation de la performance d'un système de recommandation. Il s'agit de calculer le temps réel pour évaluer le temps requis pour l'exécution des algorithmes et l'obtention des résultats escomptés.

Les spécifications matérielles de la machine utilisée pour l'exécution de ces algorithmes ainsi que les applications lancés simultanément sur cette machine au moment des calculs ayant un effet important sur la mesure du temps de calcul est dépendante (Esslimani, 2010).

2.5 Les problèmes des systèmes de recommandation

Les systèmes de recommandation souffrent d'un certains nombres de problèmes qui varient selon les approches et les techniques qu'ils adoptent, nous citons au titre d'exemple: le démarrage à froid, la sélection de voisins fiables, la robustesse et la précision des recommandations. Dans cette section, nous présentons ces problématiques avec les solutions proposées (Esslimani, 2010).

2.5.1 Démarrage à froid

Le démarrage à froid résulte généralement de l'insuffisance des données requise par le système pour produire des recommandations de qualité. Le démarrage à froid peut apparaitre selon trois modalités: le cas d'un nouveau item, un nouveau utilisateur et le dernier cas où le système est nouveau et ne dispose d'aucune information ni sur les items ni sur les utilisateurs.

➤ Nouveaux utilisateurs

Dans le cas d'un nouveau utilisateur, le système ne pourra pas recommander efficacement pour lui parce qu'il connaît peu sur ses préférences parce qu'il n'a pas fait suffisamment d'évaluations. Dans le cadre du filtrage collaboratif, le système ne peut pas définir ses voisins, comme le système ne peut pas extraire les attributs des items qui lui intéressent dans le cas du filtrage basé sur le contenu.

Pour remédier à ce problème, l'élicitations est l'une des solutions qui ont été proposées (à travers la sollicitation de notes explicites, de critiques ou d'informations démographiques). Or, cette sollicitation directe peut entraîner l'abandon de l'utilisateur. La proposition des recommandations arbitraires à l'utilisateur au démarrage, est un autre moyen pour pallier à ce problème. Cependant, cette stratégie risque d'occasionner une insatisfaction chez l'utilisateur, au vu de la faible qualité des recommandations.

(Rachid et al., 2008) suggèrent d'autres techniques pour faire face au problème de nouveauté de l'utilisateur. Ces techniques utilisent la popularité des items et l'entropie consistant à évaluer la dispersion des avis des utilisateurs sur un item.

L'utilisation des informations démographiques a été proposée comme une solution. Ainsi, deux utilisateurs sont considérés comme similaires, s'ils appartiennent à un même segment démographique, (Pazzani, 1999). Or, même si des utilisateurs appartiennent à un même segment démographique, ils ne partagent pas nécessairement les mêmes goûts (Esslimani, 2010).

➤ *Nouveaux items*

Recommander de nouveaux items constitue également un enjeu de taille pour les systèmes de recommandation. Ce problème (connu sous le nom de « latence ») peut être défini par l'insuffisance de données sur le nouvel item, c à d cet item n'a pas eu suffisamment d'évaluations. Donc, il ne pourra pas être facilement recommandé. Les systèmes les plus affecté par ce problème sont les systèmes de recommandation qui recommandent des articles d'actualité (Sollenborn et Funk, 2002).

Le problème de latence a été pallié par la proposition des nouveaux items à l'utilisateur actif pour y attribuer des appréciations. Cependant, cette stratégie pourrait occasionner une lassitude chez l'utilisateur qui risque d'abandonner le système.

L'approche de filtrage basée sur le contenu peut se présenter comme une solution alternative (Krulwich et Burkey, 1996). La technique basée sur le contenu compare le contenu d'un nouvel item avec les items qui ont été évalués par les utilisateurs afin de l'impliquer dans la recommandation. Néanmoins, les recommandations générées par cette approche sont de faible qualité à cause du manque de diversité (le problème de sur spécialisation).

Un nouveau type de filtrage basé sur le contenu exploitant les ontologies a été proposé également comme solution au problème de latence. Ce type de filtrage a été notamment adopté par le système *Entree* (qui recommande des restaurants) (Burke, 2002) et le système *Quickstep-Foxtrot* (Middleton et al., 2004). Ces deux systèmes sont spécialisés dans la recommandation des articles scientifiques. Dans ces systèmes, les ontologies sont utilisées pour classifier et de catégoriser les items et générer les modèles utilisateurs. La nécessité d'une construction préalable d'une ontologie relative au domaine de connaissance est l'inconvénient majeur de cette technique (Esslimani, 2010).

Dans les systèmes de filtrage collaboratif, L'approche KNN est utilisée pour retrouver les k plus proche voisins d'un utilisateur actif dans le but de lui recommander les items préférés par

ces voisins. La détermination d'un seuil de similarité ou la détermination d'un seuil d'items co-notés est nécessaire pour l'approche kNN. Or, la détermination de ce type de seuil reste problématique. En effet, l'ensemble des k voisins les plus proches varie avec les changements de l'état du système (intégration de nouveaux utilisateurs et d'items). De ce fait, ces seuils doivent être adaptés au fur et à mesure de la réinitialisation du système, tout en évitant de fixer des valeurs extrêmes pour que le pouvoir prédictif du système ne soit pas faible et pour que le bruit ne soit pas engendré à cause de voisins peu pertinents. En l'absence de ces items co-notés, le système ne peut pas sélectionner un voisinage fiable parce que cette approche repose sur les items notés en commun.

Dans cette optique, les associations transitives ont été utilisées par plusieurs travaux pour relier les utilisateurs aux items ou entre eux. Dans le but d'améliorer la qualité des recommandations (Papagelis et al., 2005) exploitent le principe d'inférence pour définir des voisins potentiellement fiables. Néanmoins, considérant que les systèmes de recommandation sont dynamiques et que la phase de calcul du voisinage requiert un temps de calcul important, l'application de ce type d'association devrait se baser sur des stratégies permettant de limiter par exemple le nombre d'utilisateurs concernés, afin de permettre le passage à l'échelle (Esslimani, 2010).

➤ *Nouveau système :*

Ce problème se traduit par le manque d'informations sur les items et les utilisateurs. La solution traditionnelle proposée pour résoudre ce cas de démarrage à froid est la spécification d'une phase d'apprentissage après le développement du système, afin d'accumuler les données nécessaires pour générer des recommandations aux utilisateurs.

2.5.2 La robustesse :

La stabilité des systèmes de recommandation constitue un vrai challenge pour leurs développeurs. Personne ne peut faire confiance aux appréciations fournies par les utilisateurs de ces systèmes. En effet, il n'y a pas de garantie que les données exploitées par les systèmes de recommandation représentent les réelles appréciations des utilisateurs.

(O'Mahony et al., 2006) identifient deux catégories de données bruitées :

- Le bruit naturel : ce bruit relève du fait que l'expression des appréciations est souvent perçue par les utilisateurs comme un processus fastidieux, ce qui peut influencer la qualité des opinions attribuées par ces utilisateurs.
- Le bruit malicieux : ce bruit provient de l'insertion d'information biaisée de la part de certains utilisateurs malveillants. Une de leurs motivations par exemple est de promouvoir

leur produit ou leur article en forçant le système de recommandation à générer des notes élevées pour ceux-ci et à en faire un “push”, au détriment d’autres items (concurrents) présents dans le système. De plus, le bruit malicieux peut aussi bien consister à endommager la totalité du système.

(Lam et Riedl, 2004) ont effectué une évaluation, en termes de vulnérabilité d’algorithmes utilisés et de capacité prédictive du système de recommandation, de l’impact des données bruitées sur l’efficacité du système, et proposé des méthodes de détection d’attaques. Ces méthodes permettent d’étudier les propriétés des items attaqués.

2.5.3 Précision des recommandations

L’évaluation des systèmes de recommandation est une étape primordiale dans un processus de recommandation. Cette étape montre l’efficacité du système ou à quel point les informations proposées à l’utilisateur sont pertinentes.

La plupart des travaux de recherches s’appuient sur le critère de précision dans l’évaluation des systèmes de recommandations. La précision permet en effet d’évaluer la capacité du système à recommander des items pertinents en comparant les notes prédictions avec les notes réelles d’un utilisateur.

Il est à signaler que la qualité et la précision des recommandations est étroitement liée à la disponibilité des données sur les appréciations. Dans le cas où ces données ne sont pas disponibles, le système ne peut générer des prédictions précises. En outre, cette qualité de recommandation dépend également de la fiabilité de l’algorithme utilisé pour l’apprentissage des modèles utilisateurs (Esslimani, 2010).

2.6 La personnalisation

2.6.1 Définition:

La notion de personnalisation de l’information est souvent vague dans la littérature et peut se rapporter à des concepts différents. La personnalisation peut être définie comme la sélection et la proposition de l’information qui peut intéresser un utilisateur ou groupe d’utilisateurs prenant en compte leurs caractéristiques. Généralement, les caractéristiques d’un utilisateur sont représentées par ce qu’on appelle : profil.

2.6.2 Profils ou modèles utilisateurs

Différents types d’information peuvent être utilisées pour représenter le profil d’un utilisateur ou un groupe d’utilisateurs (Arapakis et al., 2010). Ces types d’informations sont définis selon

le contexte dans lequel le profil est utilisé. Ces informations peuvent décrire par exemple ses centres d'intérêt, ses préférences, ses compétences, etc.

Il existe trois types de profil utilisateur : le profil cognitif (Lieberman, 1995), le profil qualitatif (Harrathi et Calabretto, 2006) et le profil multidimensionnel (Kostadinov, 2003).

Le profil cognitif exploité dans plusieurs systèmes personnalisés (Lieberman, 1995 ; Leung et al., 2006) est analogue au contexte cognitif de l'utilisateur traité dans la taxonomie du contexte multidimensionnel dans (Tamine et al., 2009), (Daoud, 2009).

Le profil qualitatif dans (Harrathi et Calabretto, 2006) est lié aux préférences de recherche de l'utilisateur quant à la qualité de l'information restituée par le système (fraicheur, crédibilité des sources d'information, cohérence, etc.). Ces critères concernent le contexte du document qualitatif traité dans la taxonomie du contexte multidimensionnel dans (Tamine et al., 2009), (Daoud, 2009).

2.6.3 Représentation du profil apprenant

1) *Représentation du profil* :

Il existe principalement trois types de représentation

- ensembliste : le profil y est généralement composé d'un ensemble de vecteurs de termes pondérés (Budzik et Hummond, 2000) ou classes de vecteurs (MC Gowan, 2003).
- sémantique : la représentation du profil prend en compte, dans ce cas, les relations sémantiques entre les concepts le contenant. Les ontologies (Liu et Yu, 2004) et les réseaux sémantiques probabilistes (Wen et al., 2004) sont les formalismes les plus utilisés pour ce type de représentation.
- multidimensionnelle : le profil est structuré selon un ensemble de dimensions, représentées selon divers formalismes (Bouzeghoub et Kostadinov, 2005).

2) *Construction du profil* :

La construction du profil utilisateur vise à accumuler et exploiter les données et sources d'information pertinentes pour les représenter (Gauch et al., 2003). La collecte de ces sources d'information requiert la spécification du type des données pertinentes à collecter et le mode d'acquisition des données (explicite ou implicite). Le mode d'acquisition explicite des données est le plus facile à mettre en œuvre et permet à l'utilisateur de fournir explicitement des informations requises pour la construction de son profil. En revanche, le mode d'acquisition implicite repose sur des techniques d'extraction des informations en se basant sur l'observation des comportements et interactions de l'utilisateur avec le système (fréquence de clics, temps de lecture, etc.) durant les activités des recherches.

3) *Evolution du profil*. L'évolution des profils implique leur adaptation à la variation des préférences des utilisateurs qu'ils décrivent, et par conséquent, de leurs besoins en information au cours du temps. L'évolution peut être considérée comme un problème de représentation de la diversité des domaines d'intérêts de l'utilisateur en utilisant des techniques de classification (MC Gowan, 2003) ou heuristiques liées à la notion de cycle de vie artificielle d'un centre d'intérêt (Chen et Sycara, 1998) .

2.6.4 Utilisation du profil apprenant

Dans la littérature, différents termes sont utilisés pour désigner la modélisation de l'utilisateur; en ce qui concerne apprenant nous trouvons : utilisateur, apprenant ou étudiant et pour modélisation nous trouvons profil ou modèle. Les différentes utilisations de modèle de l'apprenant seront présentées dans cette section (Moulet, 2006).

Premièrement, nous allons définir un modèle de l'apprenant orienté domaine utilisé dans les tuteurs intelligents. Ensuite, un modèle de l'apprenant orienté interface utilisé dans les hypermédias adaptatifs sera présenté. Troisièmement, nous allons détailler le modèle ePortfolios (un modèle d'apprenant orienté productions). Nous explorerons aussi le modèle de l'apprenant orienté compétences. Et pour finir, nous regarderons un modèle plutôt utilisé dans l'industrie : la carte personnelle électronique.

➤ *Tuteurs intelligents*

Les systèmes tutoriels intelligents (ITS) sont apparus dans les années 80. Leur singularité est qu'ils sont fortement liés aux recherches sur les systèmes à base de connaissance développés en Intelligence Artificielle (IA). L'architecture générale proposée par la plupart des auteurs pour ce type de systèmes, est composée de quatre modules : le modèle du domaine ou de l'expert (représentant les connaissances du domaine), le modèle de l'apprenant (représentant les connaissances de l'apprenant), le modèle pédagogique (représentant le ou les scénarios pédagogiques) et l'interface (Moulet, 2006).

Les connaissances de l'apprenant sur le domaine concerné par l'ITS (par exemple l'arithmétique ou alors les faits historiques de la seconde guerre mondiale) est représenté par un modèle orienté domaine (modèle de l'apprenant). Le modèle de l'apprenant a pour objectif en quelque sorte de construire un portrait des connaissances de l'élève qui s'enrichit à chaque étape de l'apprentissage. Cette partie du système a besoin de recueillir des informations qui peuvent être de nature différente (Nicaud et Vivet, 1988). Des informations implicites issues de l'analyse du comportement de l'élève pendant les sessions, des informations explicites issues du "dialogue" direct entre le système et l'apprenant, permettent, entre autres,

l'élaboration de ce modèle de l'apprenant. Le modèle de l'apprenant est constitué, selon John Self (Self, 1988), de quatre types de connaissances: des connaissances procédurales, des connaissances conceptuelles, des traits individuels et l'historique de l'apprenant. Il explique aussi que les modèles de l'apprenant sont utilisés en ITS pour répondre à des questions que le système se pose sur l'apprenant. En ITS, le modèle de l'apprenant est de nature cognitive, en d'autres termes, il représente les connaissances cognitives de l'apprenant. Ce modèle est utilisé pour adapter le système aux besoins particuliers des apprenants. Historiquement, les modèles de l'apprenant sont donc des modèles cognitifs. Actuellement, des recherches se font sur l'enrichissement du modèle de connaissances de l'apprenant par un modèle psychologique de celui-ci, contenant des informations sur l'état affectif, les émotions, les intentions de l'apprenant (Moulet, 2006).

➤ ***Hypermédias adaptatifs***

Le modèle de l'apprenant est adopté par les hypermédias adaptatifs (HA) pour personnaliser les apprentissages (Garlatti et Prié, 2004). Les hypermédias adaptatifs (HA) peuvent être définies, selon Brusilovsky (Brusilovsky, 2003), comme suit:

"Les hypermédias adaptatifs construisent un modèle des buts, des préférences et des connaissances de chaque utilisateur et ils utilisent ce modèle à travers les interactions avec l'utilisateur pour s'adapter aux besoins de cet utilisateur."

Dans la littérature sur les HA, Nous retrouvons deux types d'adaptation: l'adaptation de présentation et l'adaptation de navigation (Brusilovsky, 2001). L'adaptation de présentation s'effectue sur la présentation d'une ressource. L'adaptation de navigation est l'adaptation du parcours qu'un apprenant doit suivre pour trouver la ressource qu'il recherche. Selon Brusilovsky (2003), dans le domaine de l'éducation, l'adaptation de type navigation est devenue particulièrement populaire et est le type d'adaptation le plus utilisé. Nous avons constaté dans la littérature que les recherches tendent vers des fonctionnalités de personnalisation plus avancées (plus sémantiques et cognitives), souvent fondées sur les possibilités qu'offre la technologie du web sémantique. Par exemple, Dagger et al. (2004) effectuent une recherche dont le challenge est d'étendre les axes d'adaptation pas uniquement au niveau du contenu (fondés alors sur les connaissances antérieures, les buts, les styles d'apprentissage...) mais aussi au niveau des modèles pédagogiques, de la communication et des activités d'apprentissage. L'idée émergente en personnalisation des apprentissages est le référencement sémantique des objets d'apprentissage et des acteurs (Moulet, 2006).

➤ *ePortfolio*

Une nouvelle façon de collecter, d'organiser et de partager les productions des apprenants est l'utilisation d'une Portfolio. Dix définitions d'ePortfolio sont présentées dans (Moulet, 2006), selon deux critères : le contenu et l'utilisation de l'ePortfolio. Le contenu est composé de deux sous-éléments : la forme et le fond, la structure et la sémantique. Le contenu peut être défini comme un ensemble d'informations numériques, ce qui est la structure, la forme de ce contenu. Et ces informations peuvent représenter les résultats scolaires d'une personne, ce qui est la sémantique du contenu. L'utilisation est composée aussi de deux sous-éléments : les services offerts et les buts d'un ePortfolio. Par exemple, un service peut être de permettre la création de présentations de son ePortfolio pour différentes audiences alors qu'un but peut être de faire réfléchir l'apprenant sur ses apprentissages. Les buts ou utilités choisis pour un ePortfolio vont avoir une incidence sur le contenu et sur les services proposés (Moulet, 2006).

Trois types d'ePortfolio ont été identifiés dans la littérature (IMS Global Learning Consortium, 2004 et ePortConsortium, 2003). L'ePortfolio personnel est utilisé pour l'autoréflexion, il peut servir de journal de bord (concernant les expériences), il permet d'organiser des matériaux présentés en classe et il aide les étudiants à reconnaître leurs habiletés et à prendre des décisions. L'ePortfolio d'apprentissage est utilisé pour montrer l'apprentissage des étudiants, il fournit un cadre pour évaluer les progrès académiques et il permet de démontrer comment les habiletés se développent dans le temps. Pour finir, l'ePortfolio professionnel aide à faire des décisions de carrière, il peut démontrer que quelqu'un rencontre les exigences d'un programme, il peut aussi être utilisé pour présenter les habiletés et les accomplissements pour l'employabilité ou alors pour revoir le développement professionnel pour les avancées de carrière (Moulet, 2006).

➤ *Modèle orienté compétences*

L'approche par compétences est une autre méthode utilisée pour la modélisation de l'apprenant. Nous trouvons dans la littérature sur le domaine beaucoup de définitions du concept de compétence. Mais il existe un certain consensus : « une compétence est un savoir-faire qui intègre habileté et connaissance, qui est complexe, qui réfère à des habiletés cognitives, affectives, sociales ou psychomotrices et qui est spécifique à un ensemble de situations (donc à un contexte) » (Lasnier, 2000). Selon Paquette (Paquette, 2002), une compétence désigne la relation entre trois concepts : acteur, habileté et connaissance. Cette relation est montrée par la figure 2.6.

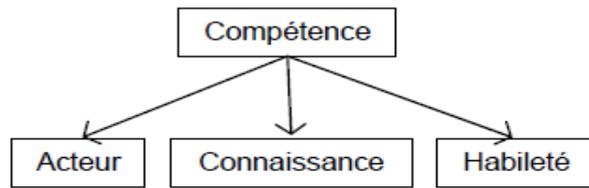


Figure 2.6 Le concept de compétence (Moulet, 2006).

La définition des connaissances inclut : des concepts, des procédures, des principes ou des faits. Les habiletés décrivent les processus qui peuvent être appliqués aux connaissances d'un domaine d'application pour les percevoir, les mémoriser, les assimiler, les analyser, les synthétiser, les évaluer, etc. (Paquette, 2002). Les habiletés se situent donc au niveau métacognitif car ce sont des connaissances agissant sur d'autres connaissances. Une taxonomie permettant de classer les habiletés selon leur complexité cognitive a été proposée par (Paquette, 2002). La création de cette taxonomie est née de l'intégration de quelques travaux, qui ont créé des taxonomies pour différents domaines d'apprentissage (cognitif, affectif, méta-cognitif et psychomoteur). Par exemple, planifier est une habileté de haut niveau : nous pouvons planifier l'exécution d'un projet (Moulet, 2006).

Une Autre définition du concept de compétence a été proposée par IMS Global Learning Consortium (2002). Le terme compétence englobe plusieurs autres termes comme par exemple: habileté, connaissance, tâche et produit de l'apprentissage. La définition contient quatre éléments (Moulet, 2006): (1) un identifiant (étiquette unique qui identifie la compétence), (2) un titre (texte), (3) une description (optionnel, texte interprétable uniquement par des humains) et (4) une définition. Le dernier élément, définition, est le plus intéressant. Ce terme peut être vu comme un ensemble d'attributs constituant ses éléments. Pour cela, habileté et connaissance de Paquette seront des éléments de la définition d'une compétence selon IMS (Moulet, 2006).

➤ *Carte personnelle électronique*

Une autre approche permettant de représenter une personne est la carte personnelle électronique. C'est une sorte de carte de visite. Les informations personnelles et professionnelles d'une personne représentent les types d'informations prises en compte par cette approche. Cette approche est plus utilisée en industrie que dans le domaine de l'éducation. Il existe plusieurs projets de carte de visite électronique. VCard représente un exemple de ces projets. vCard est proposée par l'IMC (Internet Mail Consortium) (IMC, 1998). VCard est une carte de visite virtuelle, elle contient des informations concernant une

personne comme le nom, le prénom, l'adresse, le téléphone, mais aussi des photos, des logos d'entreprise... L'échange d'informations personnelles qui sont typiquement retrouvées sur les cartes professionnelles traditionnelles, a été automatisé grâce à cette carte. Plusieurs types d'applications peuvent utiliser cette carte, nous citons au titre d'exemple : le courriel, les navigateurs web, des applications de téléphonie, des vidéos conférences, des assistants personnels, des fax..., (Moulet, 2006).

2.6.5 Services de la personnalisation

Les services que peuvent être fournies par les systèmes de personnalisation sont :

- Filtrage des résultats : Ce service permet d'affiner les résultats de requêtes en écartant l'information non pertinente. Deux méthodes sont alors utilisées: affiner la requête avant son exécution afin de réduire le résultat (Koutrika et Ioannidis, 2004, 2005), ou appliquer un post traitement sur le résultat après l'exécution de la requête, afin d'écarter les résultats inutiles (Bradley et al., 2000), (Daoud, 2009).
- Tri des résultats : Le principe du tri est de présenter en premier niveau les informations les plus pertinentes afin de rendre les données sélectionnées intelligibles à l'utilisateur (Sun et al., 2008).
- Recommandation : C'est un service de personnalisation visant à proposer les informations qui peuvent intéresser l'utilisateur.

2.7 Les types d'apprentissage

Selon (Werquin, 2010), toutes les données sur les apprentissages tout au long de la vie montrent que, pour la quasi-totalité des individus, la plus haute certification obtenue l'a été dans le système formel d'éducation et de formation initiale, soit de nombreuses années auparavant pour beaucoup d'adultes. Ceci est confirmé par d'autres sources qui montrent que près de 90 % des actions de formation pour adultes ne débouchent pas sur une certification alors même que, en fonction des pays, entre 20 et 60 % des individus s'engagent dans des apprentissages, avec comme motivation principale l'obtention d'une certification. C'est vrai en particulier des individus qui ont un faible niveau ou qui n'ont pas de certification du tout (OCDE, 2007). Il y a donc un manque de visibilité patent quant aux savoirs, savoir-faire et/ou compétences réels des individus puisque ce qui a été acquis au cours de la vie active ou d'autres activités est invisible, et ce d'autant plus qu'ils sont sortis du système initial d'éducation et de formation il y a longtemps. Cette absence de visibilité est d'autant plus préjudiciable que les individus ont un faible niveau de certification, la certification fournissant

souvent une forme de protection dans la mesure où elle est utilisée comme reflet ou indicateur des savoirs, savoir-faire et/ou compétences (Werquin, 2010).

Cette situation peut nuire à l'organisation du travail dans l'entreprise, à la promotion sociale et professionnelle des individus et à la fluidité du marché du travail en général. Remédier, même très partiellement, à cette opacité quant à ce que les individus savent et savent faire devrait aussi permettre d'améliorer le fonctionnement de la société dans son ensemble. C'est en tout cas ce que pensent les nombreux promoteurs de la reconnaissance des acquis d'apprentissages non formels et informels dans les pays de l'étude. Ce serait aussi un élément important des politiques visant à promouvoir l'équité et la seconde chance (Werquin, 2010).

➤ *Apprentissage formel*

L'apprentissage formel se déroule dans un contexte organisé et structuré et est explicitement désigné comme apprentissage (en termes d'objectifs, de temps ou de ressources (Cedefop, 2008). Il répond à un objectif clair et une intention de la part de l'apprenant, et/ou de celui qui a décidé de l'apprentissage : acquérir des savoirs, savoir-faire et/ou compétences. D'où l'idée d'associer acquis d'apprentissages d'un côté et savoirs, savoir-faire et/ou compétences de l'autre dans cet ouvrage, même s'il est clair que la notion d'acquis est plus vaste que celle de savoirs, savoir-faire et/ou compétences ; mais il n'est sans doute pas possible, ni souhaitable, de tout codifier ni de tout reconnaître (Werquin, 2010).

Les exemples typiques sont les apprentissages qui ont lieu dans le système d'éducation et de formation initiale ou bien lors de formations organisées par l'employeur sur le lieu de travail. On peut aussi parler d'éducation et/ou de formation formelle ou, plus rigoureusement, d'éducation et de formation dans un contexte formel (Werquin, 2010). Il y a bien eu quelques hésitations par le passé, notamment lorsque seules l'éducation et la formation initiale pour les jeunes étaient considérées comme relevant d'apprentissages formels (Werquin, 2007), mais la définition plus large proposée ci-dessus est maintenant assez consensuelle.

➤ *Apprentissage informel*

L'apprentissage informel est l'apprentissage découlant des activités de la vie quotidienne liées au travail, à la famille ou aux loisirs. Le caractère de ce type d'apprentissage est généralement non intentionnel de la part de l'apprenant (Cedefop, 2008). Il s'agit d'un apprentissage expérientiel ou d'expérience tout court. L'idée est que le fait même d'exister expose en permanence les individus à des situations d'apprentissage (Werquin, 2010).

À l'autre extrême par rapport à l'apprentissage formel, cette définition est aussi plutôt consensuelle, à quelques exceptions près (Werquin, 2007). Une première difficulté, on le perçoit déjà, dans une démarche de reconnaissance d'acquis d'apprentissages informels est qu'il est souvent très difficile, voire impossible, de faire en sorte que les candidats à la reconnaissance réalisent véritablement la nature et la portée de leurs propres apprentissages informels. Une deuxième difficulté vient du fait que ces apprentissages peuvent ne pas mener à une quelconque forme de reconnaissance si les acquis sont insuffisants par rapport à la norme (ou, comme vu plus haut, au référentiel en cas de certification) fixée par l'évaluateur ou l'instance d'évaluation (Werquin, 2010).

➤ *Apprentissage non formel*

Les activités d'apprentissage non formel sont des activités organisées qui ne sont pas explicitement désignées comme activités d'apprentissage (en termes d'objectifs, de temps ou de ressources). L'apprentissage non formel est intentionnel de la part de l'apprenant (Cedefop, 2008). Une grande variété des approches peuvent être employées par cet apprentissage, ce qui le rend moins consensuel. Les activités d'apprentissage non formel ne sont pas explicitement définies comme des activités d'apprentissages par exemple sans pour autant être des apprentissages informels. L'intérêt de ce concept, par son positionnement même, est qu'il donne toute latitude pour satisfaire un éventuel besoin d'un concept intermédiaire entre apprentissages formels et apprentissages informels et les utilisateurs ne s'en sont pas privés (Werquin, 2010).

Par exemple, des apprentissages non formels peuvent survenir en marge d'autres activités qui peuvent avoir ou ne pas avoir d'autres objectifs d'apprentissage. Ce peut être, par exemple, un cours de mécanique automobile sur le lieu de travail (apprentissage formel) où les participants apprennent aussi sur eux-mêmes (ponctualité, sens de l'initiative), sur le travail en équipe ou sur la résolution de problèmes (apprentissage non formel). Dans ce cas, les apprentissages non formels surviennent donc en marge d'autres activités à visée plus ou moins pédagogique. Si la participation à l'activité principale est intentionnelle, les apprentissages non formels qui en découlent peuvent ne pas l'être. En tout état de cause, ils peuvent ne pas être perçus directement et c'est ce qui rend souvent très difficiles les démarches de reconnaissance pour les publics qui n'ont pas conscience de ces apprentissages non formels annexes, ni des acquis potentiels correspondants. Un autre exemple est fourni par l'Allemagne où la totalité de la formation pour adultes relève d'apprentissages non formels. Un dernier exemple est donné par la multitude de situations où les individus se mettent en

position d'autodidacte, de manière intentionnelle et avec des objectifs clairs (maîtriser un nouveau logiciel dans l'entreprise ou à la maison) sans que pour autant il y ait un financement ou un créneau horaire défini (Werquin, 2010).

2.8 Personnaliser l'apprentissage

Un système d'apprentissage a besoin des informations pertinentes sur l'apprenant pour personnaliser, adapter l'apprentissage à ses caractéristiques ou préférences. Ces informations sont fournies par ce qu'on appelle : modèle d'apprenant. Historiquement, les modèles de l'apprenant sont des modèles cognitifs, c'est-à-dire représentant les connaissances de l'apprenant sur un domaine particulier (celui sur lequel porte l'environnement d'apprentissage). De plus en plus, apparaissent des modèles affectifs portant sur les émotions, les buts des apprenants. (Hibou et Py, 2006) précisent que les modèles de l'apprenant contiennent trois types d'informations : informations de nature cognitive, comportementale ou psychologique. Ces informations peuvent être des connaissances, des savoir-faire, des intentions ou des émotions (Moulet, 2006).

Outre les besoins exprimés par des apprenants de plus en plus familiers avec les technologies, le concept d'EAP (Environnement d'Apprentissage Personnalisé) émerge également des besoins liés à l'exploration et à la réutilisation de matériel existant, au travail collaboratif et aux communications pédagogiques qui ne surviennent pas toujours dans un cadre institutionnel.

L'EAP peut être considéré comme une approche novatrice encourageant l'utilisation de différentes technologies et leur appropriation en soutien à l'apprentissage. Les étudiantes et étudiants y apprennent à se responsabiliser face à leurs apprentissages, à développer leur jugement, leur pensée critique, etc, (Kirshenman et al., 2010).

Dans un contexte où l'autonomie est encouragée chez l'étudiant, les enseignants voient leur rôle évoluer vers l'accompagnement de ces derniers dans leurs apprentissages. Ce peut être l'occasion de découvrir les possibilités offertes par des outils autres que ceux déjà connus. Enfin, il s'agit possiblement d'un des germes d'un important mouvement de personnalisation de l'apprentissage et des transformations de l'enseignement correspondantes.

Dans le domaine des EIAH (Environnements Informatiques pour l'Apprentissage Humain), la personnalisation de l'apprentissage englobe deux facettes (Marty et Mille, 2009). La première consiste à adapter les ressources pédagogiques au contexte d'enseignement. Ce contexte rassemble le niveau et le contenu de la formation pour laquelle on souhaite adapter les ressources, mais également les besoins et les habitudes pédagogiques des enseignants.

Dans ce cas, on parle de configuration des ressources (des logiciels pédagogiques). La seconde facette concerne l'adaptation à un apprenant ou un groupe d'apprenants présentant une même caractéristique. Dans ce cas, on parle de personnalisation des ressources pédagogiques.

L'adaptation à l'apprenant peut se faire dans le cadre d'un apprentissage individualisé ou d'un apprentissage personnalisé (Verpoorten et al., 2009). Dans un apprentissage individualisé, les ressources pédagogiques sont adaptées aux buts et besoins de chaque apprenant, en fonction de ses caractéristiques. Dans un apprentissage personnalisé, c'est l'apprenant qui choisit les ressources qui lui semblent pertinentes après avoir mené une activité réflexive sur lui-même et sur son apprentissage.

2.9 Les approches de la personnalisation de l'apprentissage

La personnalisation au sein des Environnement Informatique d'apprentissage Humain (EIAH), selon (Lefevre et al., 2011b), portent toujours sur:

- le contenu des activités proposées à l'apprenant, qui peut différer de par les données elles-mêmes (valeurs ou sujet traité) mais aussi de par le niveau de difficulté,
- l'interface du logiciel ou la présentation de certaines données, et/ou ses fonctionnalités (dont l'aide et le feedback),
- la séquence d'activités (ou le parcours).

Quatre approches de personnalisation liées au type de l'objet à personnaliser (contenu, interface ou séquence de travail), ont été défini par (Lefevre et al., 2011b) . Les deux premières sont des systèmes de recommandation qui suggèrent les ressources numériques (ressources pédagogiques). Les deux dernières proposent d'adapter les séquences de travail ou le parcours que devrait suivre un apprenant dans un cursus pédagogique (Lefevre et al., 2011b).

-La première approche concerne un système d'apprentissage actif et sémantique (Szilagyi et al., 2011). Adossé à une modélisation du profil des apprenants, ce système utilise les technologies liées au Web sémantique (Szilagyi et al., 2010), pour reposer des parcours personnalisés sur mesure et en temps voulu. L'opérabilité du système s'appuie sur la définition au préalable d'objectifs d'apprentissage balisés dans le temps. Prenons par exemple un groupe d'étudiants inscrits dans un module à l'université, pour lesquels les objectifs sont présentés via un parcours pédagogique. La référence est ce parcours, car c'est principalement à partir de la réalisation des activités pédagogiques balisées de ce parcours que le système détermine les acquis et les lacunes de chacun en rapport avec les objectifs fixés. En effet, le

point d'articulation pour sélectionner les ressources pédagogiques sur mesure repose en priorité sur le rapprochement des lacunes et des acquis. C'est pourquoi une ontologie est intégrée dans le système pour les questionnaires de test des connaissances. Cette ontologie modélise le concept de *Quiz* (Balog-Crisan et al., 2009). Un quiz est considéré comme composé de plusieurs « question/réponse ». L'ontologie Quiz modélise le concept de « question/réponse » en plusieurs catégories selon le type de réponse associé : réponse booléenne, numérique, à choix multiples, etc. Chaque « question/réponse » fait référence à une ou plusieurs compétences identifiées dans une autre ontologie du système (*competency ontology*) (Lefevre et al., 2011b).

La personnalisation de l'apprentissage, adossée ainsi sur l'évaluation des connaissances, consiste à fournir à l'apprenant des compléments pédagogiques sélectionnés sur les réseaux numériques, au moment où l'apprenant en a besoin (« en temps voulu ») pour combler une lacune ou pour avancer plus vite. La sélection des ressources pédagogiques est active : elle tient compte des ressources pédagogiques enregistrées dans une base de données locale ou publiées tous les jours sur le Web. C'est pourquoi on peut dire que le système est ouvert.

Comment le système réalise-t-il un apprentissage dit sur mesure et pertinent ? Le profil de l'apprenant comprend ses préférences (style d'apprentissage), son niveau, ses acquis et ses lacunes. Ce profil est modélisé par des ontologies (ex : ontologie des préférences). De cette manière, le profil de chaque apprenant est décrit par des annotations conformes aux ontologies. C'est par ces profils que le système « connaît », à un temps donné, les besoins de l'apprenant à satisfaire pour favoriser la réussite de son apprentissage. Le système est ensuite capable de rapprocher les annotations sémantiques décrivant le profil de l'apprenant avec les métadonnées des ressources pédagogiques publiées sur les réseaux, pour sélectionner les ressources qui répondent le mieux aux besoins de l'apprenant. Les métadonnées prises en compte dans le système, pour décrire les ressources pédagogiques, sont celles du schéma de description LOM (Learning Object Meta data) (IEEE, 2002). On dispose par exemple d'une description des pré requis et des acquis de la ressource pédagogique, que l'on peut rapprocher des lacunes et acquis de l'apprenant (Lefevre et al., 2011b).

L'ontologie des préférences s'appuie sur le modèle proposé par (Felder et Silverman, 1988) qui décline les façons d'apprendre de l'apprenant en quatre axes : la perception des apprenants (sensoriel/intuitif), le format préféré de l'apprenant (visuel/verbal), la participation de l'apprenant (actif/réflexif) et la présentation du contenu (séquentiel/global).

Les informations relatives au profil de l'apprenant évoluent au fur et à mesure du déroulement de l'apprentissage, par l'exploitation de traces laissées par l'apprenant sur les

réseaux. On comprend bien que les acquis et les lacunes de l'apprenant vont évoluer au fur et à mesure du déroulement de la formation. Le système permet aussi d'affiner les préférences de l'apprenant. Les ressources pédagogiques utilisées par les apprenants peuvent être annotées par ces derniers suivant une ontologie interne au système appelée « ontologie d'appréciation ». On a alors la possibilité de retrouver les ressources pédagogiques bien « notées » par un apprenant afin de voir quelles sont leurs caractéristiques communes. On pourra aussi rapprocher les préférences des apprenants qui apprécient les mêmes ressources pédagogiques. On peut dire que ce système est évolutif car les informations liées au profil de l'apprenant évoluent au fur et à mesure des résultats et des interactions par l'apprenant avec le système.

Les systèmes de collecte et de réutilisation de données, également appelés Systèmes à Base de Traces (SBT) (Settouti et al., 2006), représentent une alternative aux techniques et outils du Web sémantique. Les SBT sont le plus souvent intégrés au sein des systèmes d'apprentissage, et restreignent ainsi le nombre et la nature des traces recueillies puisqu'elles sont limitées à celles issues des interactions des utilisateurs avec l'EIAH hôte. En conséquence, les outils de personnalisation doivent également être renfermés dans l'environnement d'apprentissage, et les données dont ils disposent sont restreintes aux traces produites par cet EIAH. Or, le nombre et la nature des outils utilisés dans le cadre de l'apprentissage étant en constante augmentation, les SBT doivent être capables de fédérer ces différents outils et proposer des solutions unificatrices fondées sur les interactions des utilisateurs avec des systèmes hétérogènes (Lefevre et al., 2011b).

- La seconde approche (Butoianu et al., 2011) propose un SBT indépendant des plateformes (Butoianu et al., 2010), c'est-à-dire un système externe capable d'intégrer des données d'observation issues de tout type d'activités réalisées par n'importe quel utilisateur (aussi bien les apprenants, enseignants et tuteurs que les personnels administratifs ou les administrateurs informatiques) sur des systèmes de natures hétérogènes (plateformes d'apprentissage, viviers de connaissances, systèmes de gestion de contenus, etc.). Ce cadre de travail repose sur trois entités : un modèle extensible décrivant les traces, une base de données implémentant ce modèle, et un ensemble de services Web capables d'interagir avec la base de données.

Deux de ces services sont fondamentaux pour le partage et la réutilisation des traces : le premier, fondé sur la spécification *Simple Publishing Interface* (SPI) (Ternier et al., 2008), est responsable de l'indexation d'une nouvelle trace dans la base de données, alors que le second respecte le standard *Simple Query Interface* (SQI) (Simon et al., 2005) pour retrouver les données d'observation. Le service SQI offre ainsi l'opportunité à diverses applications de proposer des fonctions de personnalisation à partir de la formulation de requêtes pertinentes.

À titre d'exemple, un service de recommandation personnalisé de ressources d'apprentissage a été conçu en exploitant à la fois les préférences de l'utilisateur en termes de langue et format de fichiers, mais également la ressource pédagogique la plus récemment consultée par l'utilisateur. Une application cliente de ce service a été intégrée dans l'outil de recherche d'objets pédagogiques ARIADNE Finder (Broisin et al., 2010) : lorsqu'un utilisateur consulte un objet pédagogique, le Finder invoque le service de recommandation qui recherche, dans la base de traces, les ressources dont le titre ou la description contient au moins un mot clé présent dans le titre de l'objet pédagogique consulté par l'utilisateur, et dont la langue et le format de fichier correspondent aux préférences de l'utilisateur ; les résultats sont finalement retournés au Finder qui les expose à l'utilisateur à travers l'IHM. Dans la mesure où les ressources recommandées sont stockées dans la base de traces et qu'elles ont été générées suite à l'interaction des utilisateurs avec les EIAH intégrés dans ce cadre de travail (par exemple ARIADNE Finder ou la plateforme d'apprentissage MOODLE), elles comprennent aussi bien des cours entiers ou des objets pédagogiques caractérisés par une plus faible granularité, que des messages postés dans des forums de discussion (Lefevre et al., 2011b).

-Dans la troisième approche, on considère que l'enseignant conçoit ou réutilise des activités (leçons, exercices en relation avec les leçons, etc.) qu'il va organiser dans le but de faire progresser ses élèves. L'approche présentée ici ne s'intéresse qu'à la prise de décision concernant l'enchaînement des activités à proposer au moment le plus opportun afin que les connaissances de l'élève « augmentent ». La séquence choisie, c'est-à-dire l'ensemble des activités proposées dans un ordre fixé par l'enseignant, pourra éventuellement être adaptée à chaque élève mais cela peut devenir très coûteux en temps pour l'enseignant. L'approche de (Daubigny et al., 2011), se propose d'automatiser la prise de décision sur la séquence d'activités à proposer (décisions séquentielles) pour faire progresser les élèves. Cette approche permet de trouver pour chaque élève la séquence qui pourra statistiquement le faire progresser le mieux. En effet, celle-ci n'est pas forcément la même pour tous (en fonction de la fatigue, du niveau initial, des affinités de chacun avec les tâches proposées, etc.). L'enseignement est envisagé ici comme un problème de prise de décision séquentielle dans le but de maximiser un critère : la progression de l'élève. Ce problème peut être résolu par une méthode d'apprentissage automatique basée sur les interactions entre l'élève et l'enseignant : l'apprentissage par renforcement (Sutton et Barto, 1998). Chaque élève est placé devant un ordinateur dont le rôle est de trouver la stratégie qui, en utilisant les activités les unes à la suite des autres, va faire progresser au mieux l'élève. Pour apprendre la meilleure stratégie, l'ordinateur dispose d'exemples. Un exemple est une activité proposée par l'ordinateur dont

l'impact sur l'élève a été quantifié grâce à une « récompense ». Cette récompense est définie de façon à quantifier la qualité du choix de l'activité à un moment donné, ce qui permet à l'ordinateur d'avoir un retour sur l'utilité de l'activité à ce moment précis. Choisir la récompense est une étape importante de cette approche : elle ne doit pas être trop explicite de façon à ne pas trop guider l'apprentissage de l'ordinateur mais doit être néanmoins suffisamment informative (Lefevre et al., 2011b).

Les exemples utilisés pour l'apprentissage de l'ordinateur peuvent provenir, entre autres, d'une séquence définie manuellement pour un élève de niveau moyen. Le rôle de l'ordinateur va être de trouver la séquence qui permette d'obtenir la plus grande somme des récompenses, somme constituée par les récompenses obtenues après chaque activité. En fonction de la récompense associée à chaque exemple, le comportement de l'ordinateur va être renforcé ou bien atténué. Une fois la stratégie apprise par l'ordinateur, elle peut être utilisée avec l'élève. Ainsi, cette approche, fondée sur l'apprentissage par renforcement, permettra de proposer une séquence d'activités, propre à chaque élève, déterminée par un critère (la récompense). Les activités sont choisies parmi un ensemble d'activités mises à disposition au préalable par l'enseignant. La séquence est apprise par la machine à partir d'exemples (Lefevre et al., 2011b).

-Dans la dernière approche le logiciel *Adapte* fournit à chaque apprenant des activités pédagogiques adaptées à son profil et aux intentions pédagogiques de son enseignant (Lefevre et al., 2011a). Les activités proposées sont de deux types : soit des activités sur support papier ayant différentes formes (QCM, exercices sur des textes, sur des illustrations...) et portant sur diverses disciplines (Lefevre et al., 2009a) ; soit des activités incluses dans des logiciels pédagogiques. Dans ce dernier cas, l'activité englobe son contenu, mais aussi l'environnement logiciel support de l'activité (l'interface, les fonctionnalités, les rétroactions et l'ordre des activités) (Lefevre et al., 2009b). Pour pouvoir proposer ces activités personnalisées, le logiciel *Adapte* s'appuie sur le modèle *PERSUA2* (Lefevre et al., 2011b). La mise en œuvre de ce modèle permet à chaque enseignant de définir son propre modèle de personnalisation (Lefevre et al., 2011b).

En utilisant le logiciel *Adapte*, un enseignant va dans un premier temps définir sa stratégie pédagogique en explicitant la façon dont il souhaite que les activités soient affectées aux apprenants. Pour cela, l'enseignant restreint le profil des apprenants en choisissant un ou plusieurs éléments et définit des contraintes sur les valeurs de ces éléments afin de préciser les caractéristiques recherchées. Il va, par exemple, indiquer qu'il veut faire travailler les apprenants qui ne maîtrisent pas la conjugaison des verbes du premier groupe au présent de l'indicatif. Ensuite, il définit les contraintes permettant de générer une activité ou de la choisir

dans les bases d'activités des logiciels pédagogiques dont il dispose. Ces contraintes sur le profil des apprenants et ces contraintes sur la création d'activités sont associés et hiérarchisés pour refléter les priorités pédagogiques de l'enseignant. L'enseignant peut par exemple décider de faire travailler en priorité les mathématiques, puis la conjugaison et enfin l'orthographe. Une fois que l'enseignant a défini sa stratégie pédagogique, il précise le contexte dans lequel la séance de travail va se dérouler. Pour cela, il indique au système les apprenants pour lesquels il souhaite des séquences de travail, puis il peut préciser des contraintes sur le temps, le nombre d'exercices, les supports à utiliser (papier et/ou logiciel), etc. La souplesse du modèle PERSUA2 permet aux enseignants d'associer à une stratégie pédagogique autant de contextes d'utilisation qu'ils le souhaitent, et inversement (Lefevre et al., 2011b).

À partir de la stratégie pédagogique et du contexte d'utilisation de l'enseignant d'une part, et des profils des apprenants d'autre part, Adapte propose pour chaque apprenant une séquence de travail. L'enseignant peut valider, modifier ou supprimer les séquences proposées. La modification d'une séquence peut se faire en ajoutant ou supprimant une activité donnée, en demandant un autre énoncé pour une activité, ainsi qu'en changeant l'ordre des activités, soit manuellement, soit à l'aide de fonctions de tri. Une fois les séquences validées, Adapte fournit, pour les activités sur support papier, une feuille d'exercices pour chaque apprenant et la correction des activités pour l'enseignant. Pour les logiciels pédagogiques, quand cela est possible, les fichiers de configuration sont modifiés, ou à défaut une feuille d'instructions indiquant la configuration à effectuer sur l'interface du logiciel est produite.

Avec cette approche, chaque enseignant peut obtenir des activités personnalisées, sur différents supports, respectant à la fois ses propres choix pédagogiques et les profils de chacun de ses élèves (Lefevre et al., 2011b).

2.10 La personnalisation multidimensionnelle

La personnalisation de l'apprentissage est au cœur des recherches actuelles en EIAH (Environnements Informatiques pour l'Apprentissage Humain). Les approches pour développer des EIAH permettant une personnalisation de l'apprentissage prennent en compte de différentes caractéristiques de l'apprenant et/ou du contenu proposés. Certaines de ces approches reposent sur les préférences ou domaines d'intérêts des apprenants, tandis que d'autres préfèrent mettre l'accent sur le niveau des apprenants ou leurs styles d'apprentissage. Très peu sont les approches qui regroupent toutes ces facteurs dans la personnalisation des ressources pédagogiques, malgré que quelques travaux de recherches menés ces dernières

années, aient montré l'impact positif de la personnalisation multidimensionnelle sur la qualité de l'apprentissage. Les facteurs qui ont été introduit dans la personnalisation d'apprentissage varient selon le point de vue de/des auteurs. Par exemple, Bobadilla (Bobadilla, 2009) a proposé un nouvel système de recommandation collaboratif pour personnaliser l'apprentissage prenant en compte les préférences et le niveau des apprenants. une nouvelle approche de personnalisation multidimensionnelle a été mise au point par Tzu-Chi Yang (Yang et al., 2013), en introduisant le style d'apprentissage d'un apprenant ainsi que son style cognitif . Dans notre travail, nous avons ajouté aux facteurs mentionnés précédemment, un autre facteur important lié au système cognitif de l'apprenant, on parle ainsi de la capacité mémoire de l'apprenant, en d'autres termes, sa capacité de mémorisation des informations. Les expérimentations que nous avons effectuées ont montrées l'impact positif de ce facteur sur la qualité de recommandation et d'apprentissage.

2.11 Les environnements d'apprentissage personnels (EAP)

Les plates-formes d'apprentissage en ligne ou Learning Management Systems (LMS), sont devenues un support incontournable pour les organismes de formations (universités ou entreprises). Elles permettent l'organisation des enseignements, l'affectation des apprenants aux enseignements, l'accès aux ressources pédagogiques et leur structuration. La conception et l'usage de ce type de plate-forme sont essentiellement guidés par les besoins de l'organisation et imposée aux apprenants. La plate-forme de formation offre dès lors un monde clos, rigide et par nature déconnecté des expériences d'apprentissage quotidiennes (Taraghi et al., 2009). Cela est particulièrement vrai pour l'apprentissage tout au long de la vie et l'apprentissage professionnel.

En parallèle, les services Web 2.0 tels que les réseaux sociaux, les sites de partage de ressources ou les blogs rencontrent un fort engouement et sont utilisés quotidiennement dans le cadre des loisirs ou du travail. La possibilité de combiner ces services et les informations qu'ils hébergent a mené à l'émergence d'un nouveau type d'environnement d'apprentissage : les Environnements Personnels d'Apprentissage (EAP) ou Personal Learning Environments (PLE) (Kirshenman et al., 2010).

2.11.1 Définition PLE

Les environnements d'apprentissage personnel (EAP) sont devenus de plus en plus populaires dans le domaine du *Technology-Enhanced Learning* (TEL). Ces environnements sont apparus pour pallier aux problèmes des modèles traditionnels du TEL (*Learning Management System*

(LMS), Learning Content Management System (LCMS),...). Un environnement personnel d'apprentissage (EPA, ou PLE pour *Personal Learning Environment*) peut être défini comme le «réseau de personnes entourant un individu, ces personnes utilisant des artefacts et outils pendant qu'ils sont impliqués dans des activités isolées ou collaboratives de construction de connaissances et d'informations plus ou moins intentionnelles » (Wild et al., 2011). L'EPA a tout d'abord été un concept technique dans lequel l'apprenant occupe une place centrale, auparavant occupée par le serveur dans les plates-formes d'apprentissage classiques (Olivier et Liber, 2001).

En effet, les PLEs ont été définis de plusieurs façons : un système contrôlé par l'apprenant ; un ensemble de ressources numériques regroupées dans un seul espace; un système qui supporte un apprentissage tout au long de la vie ; un système qui s'adapte si nécessaire (Lubensky, 2006) ; et un système qui permet aux apprenants d'interagir entre eux et avec d'autres personnes (enseignants, tuteurs, ...).

Dans un PLE, l'apprenant est le contrôleur de : le contenu à présenter, the *look and feel* de l'environnement, et l'interaction avec des individus et communautés (Severance et al., 2008).

Un autre point fort des PLEs est la diversité des ressources qu'ils offrent aux utilisateurs. Des chercheurs de l'université de Bolton ont défini 77 différents *patterns* dans le projet : The *personal learning environment model*, développé en 2006. Ces *patterns* sont regroupés dans 8 catégories: outils de chat et messagerie, outils des *groupware* et communautés, outils de calendrier, *schedeling* et gestion du temps (time management), outils de news *aggregation*, outils de *weblogging* et publication personnelles, outils de social software, outils de *authoring and collaboration*, outils d'intégration. Les ressources d'un PLE doivent être créées selon des standards communes permettant leurs réutilisation dans divers systèmes.

L'agrégation des ressources est l'une des caractéristiques les plus importantes des PLEs ; ces ressources peuvent être : personnes, documents, ou autres services d'apprentissage. Les PLEs sont responsables de la coordination de différentes technologies (CETIS, 2006); un système comme *NetVibes* peut être transformé en un PLE offrant ainsi à un apprenant la possibilité d'organiser des ressources comme : bookmarks, photos, vidéos, etc.

Un élément clé des PLEs est : les réseaux sociaux. Les activités qui peuvent être fournies par un environnement d'apprentissage interactif, sont : création et livraison du contenu, organisation et autorisation, chat, ... (Downes, 2006). Selon Downes, les éléments nécessaires pour développer un environnement d'apprentissage robuste, sont : la diversité, c-à-d différentes perspectives doivent être présentes, l'autonomie (chaque entité est indépendante de

l'autre, et la connectivité telle que chaque entité peut communiquer avec l'autre **sans** contraintes.

2.11.2 Les avantages des PLEs

Les PLEs ont beaucoup de point fort par rapport aux LMSs. les PLEs offrent des environnements centré-apprenant, en d'autres termes ces environnements sont totalement contrôlés par l'apprenant. L'apprenant a plus de responsabilités ; il peut prendre des décisions concernant l'interface et les services du PLEs. Avec les PLEs, L'apprenant peut aussi penser à comment construire, organiser, et contrôler son apprentissage (CETIS, 2006). Les PLEs supportent un apprentissage tout au long de la vie (lifelong learning) et non pas seulement une situation d'apprentissage présentée dans les cours traditionnels (Corlett et al., 2005).

Contrairement aux LMSs, Les PLEs sont plus robustes car ils s'adaptent rapidement aux changements des centres d'intérêt des apprenants (Utecht, 2006).

Les apprenants ne sont pas par ce qu'ils apprennent à l'école, ils peuvent apprendre : à la maison, au travail, et à n'importe quel endroit (Downes, 2006 ; CETIS, 2006), ce qui permet de créer des liens entre les activités de chacun de ces espaces.

L'apprenant peut utiliser les PLEs pour interagir avec d'autres apprenants dans leurs classes et à travers le monde (Waters, 2008).

Les PLEs répondent aux besoins des apprenants en proposant diverses informations à partir de différentes sources. Ces systèmes permettent aux apprenants de bien contrôler leurs propres ressources et interactions.

Les systèmes créés par les apprenant sont plus robustes que ceux contrôlés par les instituts (Johnson & Liber, 2008). Ces systèmes sont plus flexibles et ne requièrent pas l'installation des logiciels ou bases de données spécifiques. La majorité des fonctionnalisées demandées par les apprenants ne sont pas offertes par les systèmes développés par les instituts. Avec les PLEs, les apprenants peuvent créer les services qui ne sont pas disponibles ; ceci est très difficile avec les LMSs (CETIS, 2006).

2.12 Conclusion

La théorie des SRs est assez vaste, elle fait intervenir plusieurs domaines connexe comme la psychologie, la sociologie, la statistique, etc. En fait, tout domaine qui peut être utilisé pour étudier et modéliser le comportement des utilisateurs dans les systèmes, prédire leurs préférences ou évaluer la qualité du système, peut être appliqué. Il n'est pas loin de voir la médecine, par exemple, appliquée pour décrire la fonction de prédiction ; le jour où on découvrira que les goûts des êtres humains ont un certain lien avec leurs physiologies, on va probablement exploiter cette donnée pour générer la recommandation ! Mais pour le moment, les techniques de recommandation existantes n'utilisent que les statistiques et les fondements mathématiques, mais malgré cela les résultats de prédiction peuvent être considérés bons, et ils s'améliorent de plus en plus.

Chapitre 3

Un système d'apprentissage personnalisé

3.1 Introduction

Dans le contexte du e-Learning, les environnements d'apprentissage offrent des contenus de même type aux apprenants, malgré que leurs besoins en termes d'informations soient différents. Dans la perspective de proposer de l'information personnalisée aux apprenants, nous avons proposé une nouvelle approche de recommandation qui repose sur un filtrage hybride permettant ainsi une personnalisation multi critères du contenu pédagogique. Cette approche est appelée « New multi-Personalized Recommender for e-Learning (NPR-eL)» (Benhamdi et al. 2016). Nous avons intégré l'approche : NPR-eL dans une autre approche qui permet la modélisation du processus d'apprentissage en utilisant le langage : IMSLD (Benhamdi et Seridi, 2007), afin de créer un environnement personnalisé (NPLE : New Personal Learning Environment).

Ce chapitre est organisé comme suit : la section 2 présente les solutions que nous avons utilisées pour pallier au problème de démarrage à froid. La section 3 présente quelques systèmes de recommandation développés pour le domaine du e-Learning. La section 4 décrit les deux parties qui composent notre approche. La dernière section conclut ce chapitre.

3.2 Solutions au problème de démarrage à froid

Le problème de démarrage à froid (DàF) est l'un des problèmes majeurs des systèmes de recommandation. Pour le filtrage collaboratif (FC) (Salehi et al., 2014), ce problème émerge le plus souvent dans le cas où les évaluations explicites communes entre les utilisateurs sont très peu ou nulles. Dans ce cas, le système ne peut pas trouver des utilisateurs similaires à l'utilisateur cible. Les systèmes de recommandation basés contenu ne sont pas capables de recommander efficacement à un nouvel utilisateur avant d'obtenir un certain nombre d'appréciations de sa part. Outre, la difficulté d'indexer des documents multimédia (audio, vidéo,...) est un goulot d'étranglement pour cette approche.

Plusieurs chercheurs ont favorisé l'hybridation du FC (Filtrage Collaboratif) et le filtrage basée sur le contenu pour résoudre ce problème. Mais cette solution n'est pas réellement efficace car elle pose d'autres problèmes.

D'autres solutions ont été proposées basées sur de nouvelles approches pour la représentation du contenu, nous citons par exemple: les ontologies (Middleton, 2002), les folksonomies (Duraio, 2010), et les taxonomies (Weng, 2008). Les ontologies sont difficiles à construire et exploiter car elles sont liées à d'autres technologies. Les folksonomies sont un

ensemble de mots clés (tags) proposés arbitrairement par des utilisateurs pour décrire le contenu d'un certain nombre de documents. En plus, les folksonomies ont une structure non hiérarchique ce qui implique que ce type d'information n'est pas appropriées pour représenter le contenu pédagogique. Les taxonomies offrent une représentation hiérarchique et simple de n'importe quel type de contenu ; ceci rend ce type d'information plus populaire. Weng (Weng, 2008) a proposé une approche puissante pour résoudre le problème de démarrage à froid relatif à un nouvel utilisateur et/ou un nouvel item en se basant sur une approche hybride qui utilise le concept des taxonomies (CSHTR : Cold Start Hybrid Taxonomic Recommender). On est parti de ses principes de fonctionnement globaux pour proposer notre approche : NPR_eL. CSHTR propose aux utilisateurs d'évaluer un ensemble de documents (principe adopté par le filtrage collaboratif et basé sur le contenu pour construire les profils des utilisateurs) et transformer ces évaluations en préférences taxonomiques (profil). Dans notre travail, nous avons proposé aux apprenants de choisir directement leurs domaines d'intérêt en répondant à un questionnaire (une taxonomie représentant un ensemble de domaines d'apprentissage) pour éviter la perte du temps dans la lecture des documents qui peuvent être loin de leurs préférences.

3.2.1 Les taxonomies

Dans cette section nous présentons une nouvelle source d'information, nous parlons ainsi des taxonomies. Une taxonomie est un ensemble contrôlé de termes ou de thèmes (en anglais: topics) avec une structure hiérarchique, permettent de décrire et classifier des documents (items) (Levy, 2004). Vu la croissance du volume d'information, les sites du e-commerce et les applications *Business to Business* (B2B), les taxonomies sont devenues très répandues. Par exemple, L'UNSPSC (the *United Nations Standard Products and Services Classification*) a identifié plus de 11000 codes taxonomiques pour classifier des services et des produits pour être utilisés dans '*global market place*' (Levy, 2004 ; Leo et al., 2003). Les sites du e-commerce comme : Amazon.com (<http://www.amazon.com>), BARNES & NOBLE (<http://www.book.com>), art.com (<http://www.art.com>), eBay (<http://www.ebay.com>),..., ont développé leurs propre taxonomies pour décrire leurs produits.

Vu l'absence d'une taxonomie standard pour le domaine du e-Learning permettant de décrire les matières enseignées dans les universités, nous avons développé notre propre taxonomie orientée pour représenter les matières programmées par notre département (département d'informatique, université 8 mai 1945, Algérie). Nous avons utilisé aussi la

taxonomie des livres utilisée par Weng (<http://www.amazon.com>), pour pouvoir comparer les résultats de notre approche (NPR-eL) et ceux du CSHTR (Weng , 2008).

3.2.2 relation entre préférences de l'item et préférences taxonomique

La plupart des systèmes de recommandation génèrent des recommandions en se basant sur leurs préférences des items. L'idée la plus communément connue est que les utilisateurs partagent les mêmes préférences des items ayant certainement les mêmes goûts. En partant de cette idée, le filtrage collaboratif compare les utilisateurs entre eux sur la base de leurs jugements passés pour créer des communautés, et chaque utilisateur reçoit les documents jugés pertinents par sa communauté (Breese et al., 1998). Le principe est de filtrer le flot de documents entrant en fonction de l'opinion que d'autres utilisateurs de la communauté ont déjà porté sur les documents. Si un document a été jugé intéressant par un utilisateur, il sera diffusé automatiquement aux utilisateurs qui ont eu des opinions similaires par le passé. Des travaux récents ont exploité les préférences taxonomiques pour générer des recommandations ; ceux sont basés sur l'hypothèse : « les utilisateurs qui partagent les mêmes préférences taxonomiques, partagent aussi les mêmes centres d'intérêt » (Ziegler et al., 2004 ; Middleton et al., 2002). Li-Tung Weng (Weng, 2008) a proposé une hypothèse, qui a été prouvée par des expérimentations. En se basant sur cette hypothèse, il a développé ses systèmes de recommandation (HTR « Hybrid Taxonomic Recommender», CSHTR).

CSHTR identifie le groupe des similaires d'un utilisateur en exploitant ses préférences taxonomique. Autrement dit, au lieu d'affecter un utilisateur donné à un groupe à la base de ses préférences standards, le système compare ses préférences taxonomiques personnelles aux préférences taxonomiques de chaque groupe, et l'affecte à celui qui se trouve à être le plus proche. Les préférences taxonomiques de chaque groupe sont obtenues en combinant tous les vecteurs personnels des préférences taxonomiques des utilisateurs dans ce groupe.

$$\left\{ \begin{array}{l} \vec{v}_{uc} = (v_{uc1}, v_{uc2}, \dots, v_{uc|c|}) \\ v_{uck} = \frac{\sum_{u \in uc} v_{uk}}{|uc|} \end{array} \right.$$

uc est le groupe des plus proche voisin, et v_{uck} est le $k^{ième}$ composant du vecteur taxonomique du groupe uc : \vec{v}_{uc} . Puis, le vecteur taxonomique de l'utilisateur u sera comparé au vecteur de chacun des groupes, afin de trouver le plus similaire. La similarité entre deux vecteurs est calculée par la mesure suivante :

$$t_sim(\vec{v}_i, \vec{v}_j) = \frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|}$$

Après les calculs de similarité, le groupe à choisir sera celui qui vérifiera :

$$t_cluster(u_i) = \operatorname{argmax}_{uc \in UC} new_sim(\vec{v}_{u_i}, \vec{v}_{uc})$$

\vec{v}_{u_i} est le vecteur taxonomique de l'utilisateur u_i . \vec{v}_{uc} est le vecteur taxonomique du groupe uc . Donc, un utilisateur qui n'a aucune évaluation commune avec les autres aura toujours la possibilité d'appartenir à l'un des groupes grâce à ses préférences taxonomiques.

Dans la présente thèse, nous avons calculé non seulement la similarité taxonomique, mais aussi la similarité cognitive entre l'apprenant cible et tous les apprenants du système. Ces calculs vont déterminer pour l'utilisateur cible son groupe des similaires. Un nombre n des utilisateurs les plus proches 'taxonomiquement' va être choisi tel que la somme des similarités de ces utilisateurs soit égale au moins à 1.

3.3 Travaux existants

Ces dernières années, nous trouvons beaucoup de travaux qui ont suggéré les ontologies comme un outil de personnalisation (Cheung, 2010). Cuiciu et Tang (Cuiciu et Tang, 2010) ont proposé des stratégies de *Data matching* pour recommander des objets d'apprentissage (*learning material*) personnalisés. TagSES (Kim, 2011) est un environnement d'apprentissage personnalisé qui offre des services de recommandation. Au sein de cet environnement, les apprenants peuvent tagger (le mécanisme de *tagging* signifie la possibilité de décrire des objets d'apprentissage par un ensemble de mots clés appelés: *tags*) des objets d'apprentissage pour les partager avec d'autres apprenants. Le modèle d'ontologie établit une relation entre les apprenants, les objets d'apprentissage et les *tags*. Cheung and Kaseiman (Cheung and Kaseiman, 2011) ont développé un environnement d'apprentissage personnalisé dans le cadre des *Adaptative Educational Hypermedia* (AEH) pour proposer aux apprenants des cours de java en se basant sur leurs pré requis. Les ontologies ont été utilisées pour représenter le modèle d'apprenant. Galip and Altun (Galip and Altun, 2011) ont élaboré un modèle d'apprenant basé ontologie pour les environnements d'apprentissage. En effet, les ontologies sont difficiles à créer car celles-ci sont dépendantes d'autres technologies. Pour construire des ontologies, il faut avoir des prés connaissances non seulement sur les outils développés pour les créer, mais aussi sur les langages qui permettent de les interroger. Malgré que les systèmes de recommandation sont basés sur des méthodes statistiques, leurs implémentation est plus

facile. Donc, nous avons décidé d'utiliser les systèmes de recommandation entant qu'un outil de personnalisation. Dans le contexte du e-Learning, nous citons quelques travaux qui adoptent cet outil pour la personnalisation du contenu pédagogiques. ReMashed (Drashler et al., 2009) est un système de recommandation hybride qui a bénéficié des sources du Web 2.0. Les utilisateurs de *ReMashed* ont la possibilité d'évaluer les données (tagging) de tous les utilisateurs du système. (Salehi et al, 2014) ont proposé un système de recommandation personnalisé basé sur le filtrage collaboratif et *sequential pattern mining*, pour la recommandation des objets d'apprentissage. Bobadilla (Bobadilla, 2009) a prouvé l'importance du niveau de l'apprenant et son impact positive sur la qualité des recommandations. Durao and Dolog (Durao and Dolog, 2010) exploitent la similarité entre les tags définis par les utilisateurs dans la production des recommandations. Anjorin et son équipe (Anjorin et al., 2011) ont créé un système de recommandation et l'intégré au sein de l'environnement d'apprentissage CROCODIL. Les recommandations sont générées en se basant sur le *tagging* collaborative des ressources. Un système de recommandation hybride à base de cas est développé par (Vladoiu et al., 2013) pour la recommandation des ressources pédagogiques. Un autre système de recommandation des objets d'apprentissage prenant en compte le profil des apprenants, les propriétés de ces objets, et le contexte, a été proposé par (Lichtnaw et al., 2011). Les folksonomies sont un ensemble de tags (mot clés) non hiérarchique, ce qui signifie que ce type d'information n'est pas approprié pour la représentation des domaines d'apprentissage. Les taxonomies ont une structure hiérarchique et leur construction est simple. Vu leur capacité de représenter n'importe quel type de données, nous l'avons utilisé dans ce travail.

Tous les travaux cités au-dessus introduisent seulement le niveau des apprenants et/ou leurs préférences. Dans la littérature, nous trouvons pas mal de travaux en psycho pédagogie qui ont mis l'accent sur l'importance de certaines stratégies qui aident la mémoire humaine à mémoriser des informations (Oxford, 2001) et mettre en évidence la relation entre la capacité de la mémoire de l'être humain et sa performance cognitive et plus précisément sa compréhension. Malgré son importance dans le processus d'apprentissage, d'après nos connaissances, il y'a aucun travail jusqu'à nos jour qui implique la capacité mémoire dans la recommandation des informations aux apprenants. Outre, aucun de ces travaux ne traite le problème de démarrage à froid. Ces dernières années, une technique de recommandation très puissante, a été développée: CSHTR, qui exploite les taxonomies pour faire face à ce problème. Pour cela, nous avons développé une extension de cette approche, appelée : NPR-eL, qui propose une solution améliorée au démarrage à froid.

3.4 Un nouveau système d'apprentissage personnalisé (NPLE)

Aujourd'hui, la personnalisation est apparue au sein de chaque système qu'un individu interagit avec. La personnalisation est devenue nécessaire dans chaque système ayant pour but de fournir des informations concernant nos préférences ou centres d'intérêt. Dans le contexte du e-Learning, la personnalisation consiste à présenter à un apprenant ou groupe d'apprenants des contenus pertinents, et ceci en fonction de ses caractéristiques. Les systèmes de recommandation sont l'une des technologies de personnalisation les plus flexibles ; ce qui signifie qu'on peut l'utiliser dans n'importe quel domaine. Pour cela nous avons décidé de l'adopter dans notre domaine de recherche : EIAH. Beaucoup de chercheurs ont montré par expérimentation l'utilité de prendre en compte plusieurs facteurs, soit liés à l'apprenant ou aux items recommandés, dans la personnalisation d'apprentissage. Quels sont les facteurs qui doivent être présents dans le calcul des recommandations ? La réponse varie d'un chercheur à un autre ; certains ont mis l'accent sur les centres d'intérêt et le niveau de l'apprenant, d'autres préfèrent d'ajouter son style d'apprentissage, style cognitif,.... En effet, nous avons décidé d'ajouter la capacité de mémoire de l'apprenant parce que nous avons convaincu que :

Hypothèse 1: « Un apprenant qui possède une bonne mémoire, il va certainement bien bénéficier de l'apprentissage ».

Partant de ce principe, nous avons proposé une nouvelle approche hybride (basée sur le filtrage collaboratif et basé contenu) de recommandation personnalisée : NPR_eL (New multi-Personalized Recommender system for e-Learning), qui permet de prédire l'utilité d'un item pour un apprenant donné prenant en considération trois facteurs : ses centres d'intérêt, son niveau et la capacité de sa mémoire ou son aptitude de retenir les informations acquises. Les expérimentations que nous avons faites ont montré que la qualité de l'apprentissage a été affectée par la capacité mémoire de l'apprenant. En effet, nous avons prouvé aussi par expérimentation l'hypothèse suivante :

Hypothèse 2 : « la qualité de recommandation est fortement liée au nombre de propriétés de l'apprenant (ou utilisateur) prises en compte dans le processus de recommandation, en d'autres termes, plus le nombre de propriétés est élevé, plus le système de recommandation est performant ».

Cette approche est constituée de trois étapes : construction du profil, regroupement et prédiction d'évaluation. L'apprenant cible commence tout d'abord par la sélection de leurs domaines d'intérêt en répondant à un questionnaire. La première partie de ce questionnaire

présente la nouvelle taxonomie que nous avons créée pour le e-Learning (pour plus de détail voir section 3.4.1.1). Puis, il doit passer : cinq pré-tests de différents niveaux (très faible, faible, moyen, fort et très fort), pour évaluer ses prés-connaissances, et un autre test basé sur la méthode RSI (Recall Serial Information) afin de mesurer la capacité de sa mémoire. La méthode RSI a été développée par Daneman et carpenter (Daneman et Carpenter, 1980) dans le cadre du domaine 'psycho-cognitif'. En psychologie expérimentale et en neuropsychologie, les mesures classiques pour évaluer la mémoire à court terme sont des taches de rappel sériel immédiat (RSI) (empan de chiffres ou de mots). Le principe de cette méthode est de présenter à une personne un ensemble de listes de : lettres, chiffres, lettres et chiffres, mots, la capacité de la mémoire est le nombre des éléments de la plus longue liste (span) que la personne puisse rappeler après une seule présentation.

Le profil de l'apprenant est construit en se basant sur ses réponses. La deuxième étape se traduit par la définition de son groupe des similaires ou les apprenants ayant les mêmes préférences d'items et taxonomiques, et partagent les mêmes caractéristiques. Dans la troisième étape, le système prédit la préférence d'un item par un apprenant en fonction de la préférence de cet item dans son groupe et la similarité taxonomique entre l'apprenant et cet item.

Pour le développement d'un environnement d'apprentissage personnalisé, nous avons combiné la nouvelle approche de recommandation « NPR-eL » (Benhamdi et al., 2016) avec une autre permettant la modélisation d'une situation d'apprentissage ou la création d'un cours en utilisant le langage de modélisation le plus utilisé : IMS Learning Design (IMSLD). L'approche globale (figure 3.1) est constituée de deux grandes phases : la première phase est la modélisation d'une situation d'apprentissage individuel, qui se traduit par la définition de quelques éléments nécessaires pour l'apprentissage (objectif, activités, ...). La deuxième phase est celle durant laquelle le modèle de scénario créé dans la phase précédente, sera instancié. Dans cette phase, l'apprenant doit réaliser les activités définies dans le scénario. Nous détaillons ces étapes dans la section suivante.

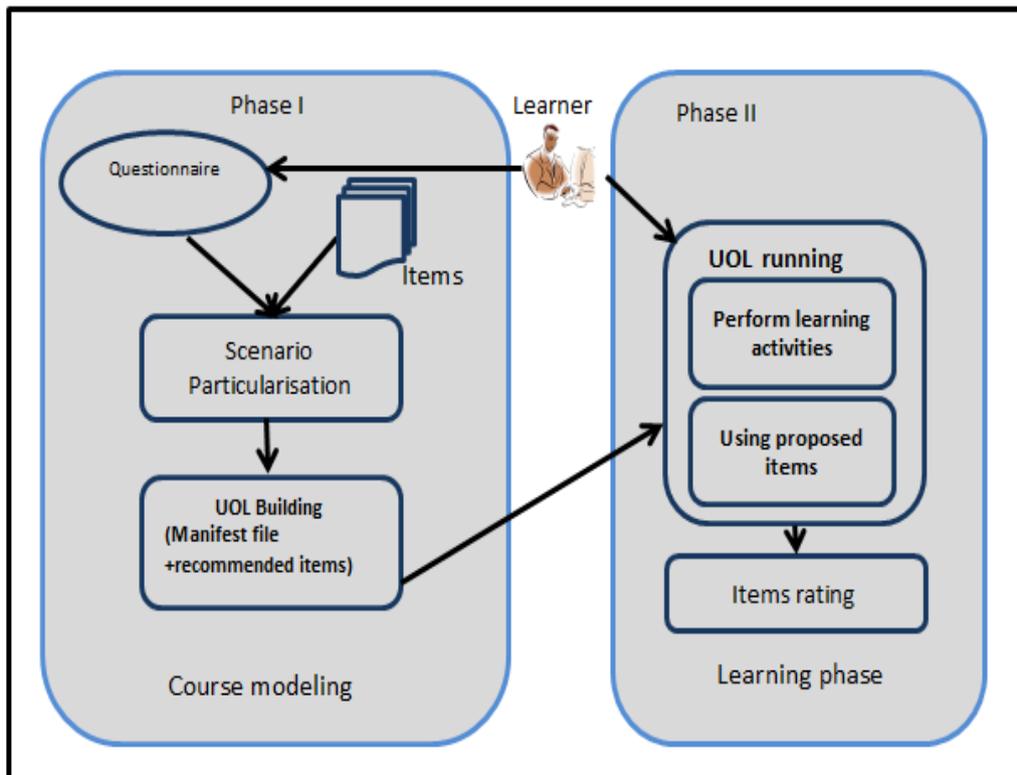


Figure 3.1 L'architecture du système (NPLE).

3.4.1 Modélisation d'une situation d'apprentissage

Un scénario d'apprentissage représente la description du déroulement d'une situation d'apprentissage ou unité d'apprentissage visant l'appropriation d'un ensemble précis de connaissances, en précisant les rôles, les activités ainsi que les ressources de manipulation de connaissances, outils, services et résultats associés à la mise en œuvre des activités. Cette phase est composée de son tour de deux étapes : La particularisation du scénario et la création d'une unité d'apprentissage (UoL).

3.4.1.1 La particularisation du scénario

Au début l'apprenant doit répondre à un questionnaire pour choisir son domaine d'intérêt et le modèle du contenu qu'il préfère, et passer les cinq pré-tests et le test de la capacité mémoire pour construire son profil.

La première étape consiste en l'extraction du *manifest* décrivant le modèle de contenu (ou modèles d'unités pédagogique (UP)) choisi par l'apprenant, à partir de la base des modèles (*courses' templates*). En effet, nous avons pris en compte dans ce travail seulement les concepts définis au niveau A du modèle conceptuel d'IMSLD. Certains éléments du *manifest* sont définis dès le début comme par exemple : les activités qu'un apprenant doit réaliser durant

l'apprentissage, alors que d'autres doivent être ajoutés dans le *manifest* après avoir été définis par l'apprenant dans la phase de la particularisation du scénario, nous citons au titre d'exemple : l'objectif qu'un apprenant vise à atteindre et les ressources nécessaires. Les ressources sont les items proposés par le système de recommandation NPR-eL. Après la génération de recommandation, le système prend en charge l'ajout des URLs des items recommandés dans l'élément *resources* du *manifest*. La figure 3.2 résume cette phase :

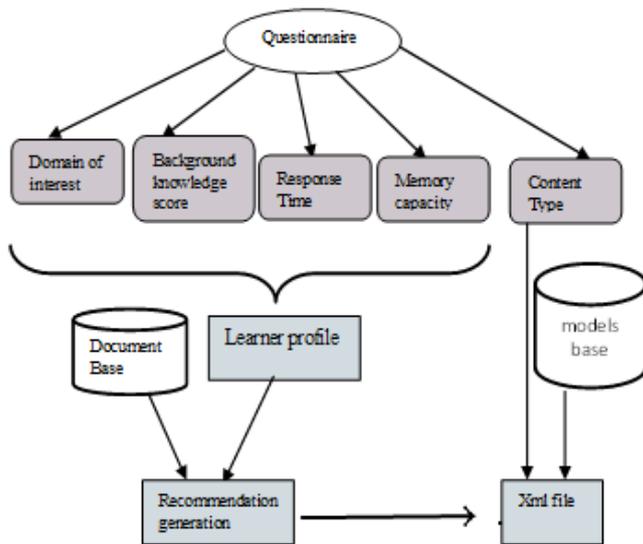


Figure 3.2 Particularisation du scénario.

La figure 3.3 montre les différents éléments ou la structure d'un *manifest* selon le langage IMSLD. Nous avons utilisé une structure simplifiée des éléments du *manifest*. La structure que nous avons proposée pour ces éléments sera détaillée par la suite.

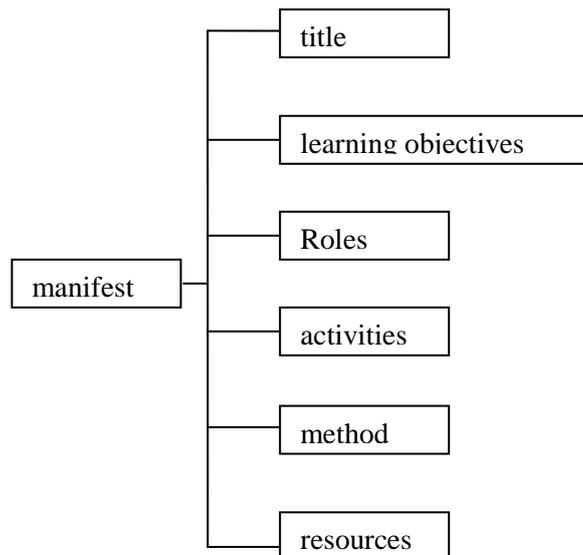


Figure 3.3 La structure d'un *manifest*.

Les rôles : roles

Cet élément permet de définir les différents rôles qui peuvent être joués par les personnes qui peuvent participer dans le processus de l'apprentissage. En effet, il existe deux types de rôle : enseignant (*stuff*) et étudiant (*learner*). On a choisis une structure très simple pour l'élément rôles, constituée d'un ou plusieurs sous éléments appelés : *learner* et un sous élément *teacher*, chacun de ces éléments est constitué à son tour d'un titre permettant de décrire son nom. Cette structure est présentée par le diagramme en arbre suivant :

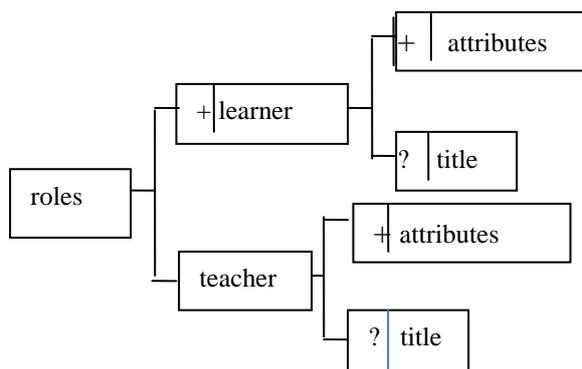


Figure 3.4 Diagramme en arbre dérivant l'élément rôles.

Les activités (activities)

Comme son nom l'indique, cet élément sert à présenter les différentes activités réalisées par les différents rôles définis par l'élément précédent. Il comporte un ou plusieurs : *learning activity*, *structure activity* ou *support activity*. *Learning activity* est une activité simple qui contient les sous éléments: *title* et une description pour donner plus de détail sur cette activité, alors que *structure activity* désigne une activité structurée, composée de deux ou plusieurs *learning activity*. Le troisième type d'activité qui est : *support activity*, est l'activité déterminée pour être effectuée par l'enseignant, les sous éléments qui la composent sont identiques à ceux des éléments *learning activity* et *structure activity*. Le diagramme suivant illustre la structure de l'élément *activities*.

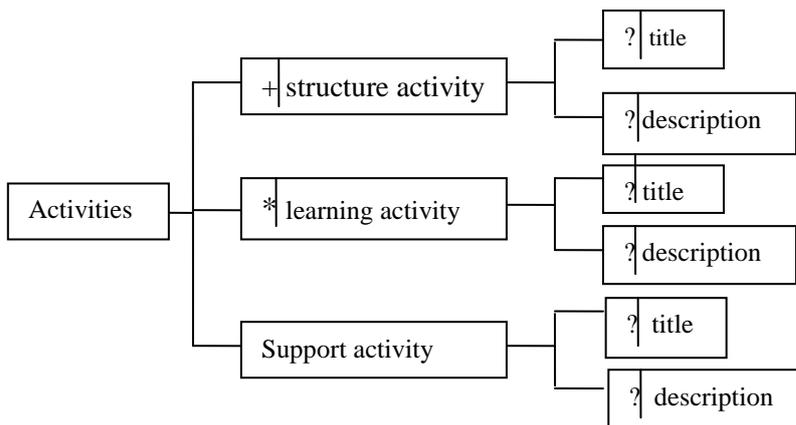


Figure 3.5 Diagramme en arbre dérivant l'élément *activities*.

La méthode (method)

C'est l'élément le plus compliqué d'un *manifest*, car il contient beaucoup d'éléments. L'élément *method* est constitué d'un ou plusieurs sous élément nommés : *play*, qui est composé d'un *title* et d'un ou plusieurs *act*. L'*act* possède lui aussi un *title*, et fait référence à un ou plusieurs *role part*. Les sous éléments: *role-ref* et l'un des sous éléments : *structure-activity-ref*, *support-activity-ref*, *learning-activity-ref*, qui sont regroupés dans l'élément :*activity-ref*. Dans ce qui suit un diagramme qui présente cet élément.

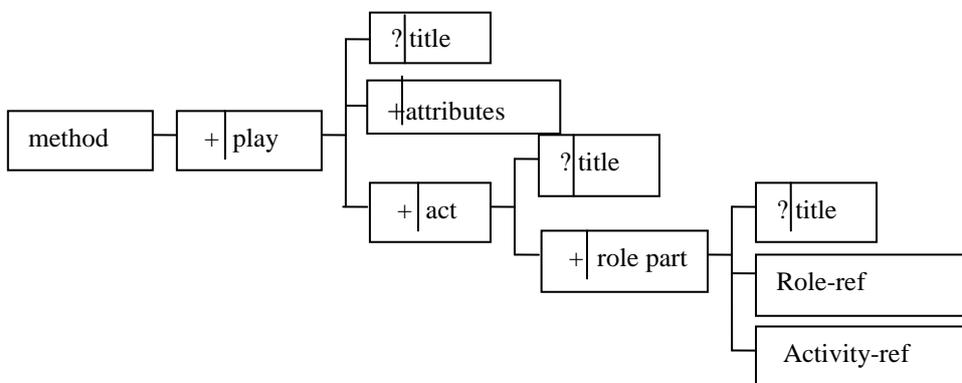
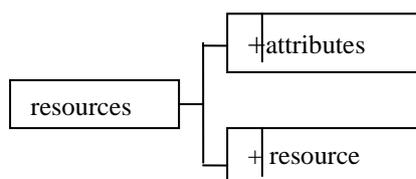


Figure 3.6 Diagramme en arbre dérivant l'élément *method*.

Les ressources (resources)

Les cours et les services de communication qui peuvent être utilisés lors de l'apprentissage, on les appelle : ressources. L'élément *resources* est structuré d'un ou plusieurs sous éléments nommés : *resource*. La structure de cet élément est présentée par la figure suivante :

Figure 3.7 Diagramme en arbre dérivant l'élément *resources*.

Giacomini (Giacomini, 2006) a proposé 16 modèles d'unité pédagogique et les scénarios qui leur sont associés, prenant en compte trois types d'unités d'apprentissage qu'elle a déterminé à partir de certains critères comme : existence des parties théoriques dans un module de formation, différents types d'exercices (QCM, problèmes à résoudre, texte à trous, etc), de projets, des types d'évaluation et/ou de l'auto évaluation,... Ces trois types sont : présentation des concepts théoriques, exercices comme application du cours et projets (travaux pratiques). D'après Giacomini (2006) l'unité pédagogique est composée d'un ou plusieurs scénarios pédagogiques. Pour notre travail ce qui nous intéresse est le type du contenu pédagogique décrit dans chaque modèle ainsi que les activités individuelles. Le tableau 3.1 permet de citer ces modèles. Le tableau 3.2 présente le modèle d'unité pédagogique (UP) 10.

No de modèle	Type d'unité pédagogique
1	Résolution de problème/exercices
2,3	concepts théoriques avec exercices
4	concepts théoriques, exercices de cours, résolution de problèmes
5, 6, 7,8	Résolution de problèmes
9	concepts théoriques, résolution de problèmes
10,11	concepts théoriques
12, 13, 14, 15	projet
16	Modélisation et résolution de problèmes

Tableau 3.1 Les modèles d'unités pédagogiques (Giacomini, 2006).

	Modèle d'UP
Nom du modèle	Le modèle 10 : Apprentissage des concepts théorique
Définition du Modèle	Le modèle d'activité est basé sur l'apprentissage des concepts théoriques et propose des activités d'apprentissage individuel comme l'analyse d'un chapitre de cours ou la recherche d'informations pertinentes. Ensuite les étudiants discutent sur les notions apprises puis rédigent une synthèse du groupe sur le sujet de chapitre étudié. Une fois que toutes ces phases sont réalisées et validées par l'enseignant, les étudiants passent à l'étude du chapitre suivant en conservant les mêmes types d'activités et le même enchaînement.

Tableau 3.2 Le modèle d'UP 10.

A. Génération de recommandation

L'objectif de tous les développeurs des systèmes de recommandation est de créer des systèmes qui produisent des recommandations de qualité. Chacun a son point de vue sur les critères qui permettent d'atteindre cet objectif. Nous pensons que la prise en considération des caractéristiques liées au système cognitif de l'apprenant permettra de générer des recommandations fiables. Avant de passer à la présentation de notre approche, nous exposons tout d'abord les différences entre notre technique de recommandation et CSHTR. La comparaison entre ces deux approches est résumée dans le tableau suivant :

	CSHTR	NPR_eL
Domaine d'application	e-commerce	EIAH
les cas traités du DàF	Nouvel utilisateur et/ou item	Nouvel apprenant et/ou item
Solution proposée au DàF	Taxonomie	Taxonomie + questionnaire
Propriétés prises en compte dans le processus de recommandation	les preferences d'utilisateurs	les préférences, le niveau et la capacité mémoire des apprenants
Création de Profil	calcul basé sur les évaluations explicites des utilisateurs (pour plus de détail veuillez référer à Weng (2008))	calcul basé sur les réponses de l'apprenant au questionnaire (l'apprenant peut directement choisir son domaine d'apprentissage)

Tableau 3.3 CSHTR et NPR-eL : comparaison.

Dans ce qui suit, nous détaillons l'approche de recommandation que nous avons proposée : NPR-eL (figure 3.8).

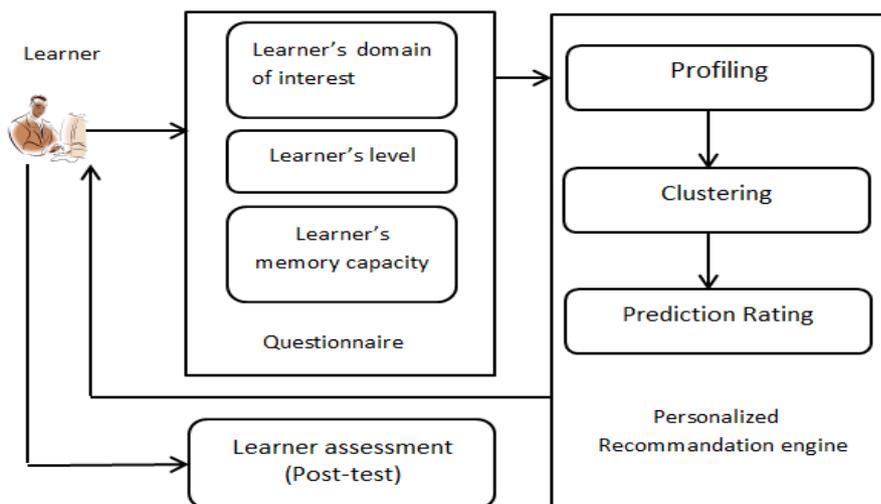


Figure 3.8 Le système de recommandation NPR-eL.

1) Représentation du monde visé

a) University Dataset :

Avant d'entrer dans le détail de l'approche de recommandation NPR-eL, nous présentons les différents concepts et notations utilisés dans ce travail, et qui seront utilisés dans tous les chapitres et sous sections dans cette thèse.

- les apprenants : $A = \{a_1, a_2, \dots, a_c\}$

Tous les apprenants qui ont reçu des recommandations sont des éléments de l'ensemble A.

-les items : $T = \{t_1, t_2, \dots, t_k\}$

Tous les items qui existent dans la base des items (les items qui sont liés aux domaines présentés par la taxonomie qu'on a proposée) appartiennent à l'ensemble T.

-les évaluations explicites :

$r_{a,i}$ est l'évaluation qu'a fait l'apprenant a sur l'item i. A la fin du processus d'apprentissage, l'apprenant évalue explicitement les items proposés par le système de recommandation. Ces évaluations seront utilisées pour calculer la similarité entre l'apprenant a et le prochain apprenant actif. Les évaluations explicites d'un apprenant sont comprises entre 0 et 1. La valeur 0 d'une évaluation indique une satisfaction minimale et 1 une satisfaction maximale.

- l'ensemble des descripteurs taxonomique d'un item: $D(t) = \{d_1, d_2, \dots, d_c\}$

$D(t)$ est l'ensemble des descripteurs d'un item t tel que $|D(t)| \geq 1$. Dans le but de décrire et classifier des items, chaque item $t \in T$, est associé à un ensemble de descripteur taxonomique. Un descripteur taxonomique (*taxonomy descriptor*) est une séquence ordonnée de thèmes (*topics*) p dénoté par : $d = \{p_1, p_2, \dots, p_q\}$ où $d \in D(t)$, $t \in T$, $p_1, p_2, \dots, p_q \in C$. les thèmes au sein d'un descripteur taxonomique sont ordonné en ordre décroissant en commençant par le thème le plus générale (racine).

La figure 3,9 présente un ensemble d'items (documents) avec leurs descripteurs taxonomiques (University Dataset). Les thèmes générales de ces descripteurs sont: GUI (Graphical User Interface), KM (Knowledge Management, TL (Theory of Langage), OP (Operating System), DB (Data Base) . Au titre d'exemple, l'item t_1 est décrits par trois descripteurs: $D(t_1) = \{d_1, d_2, d_3\}$

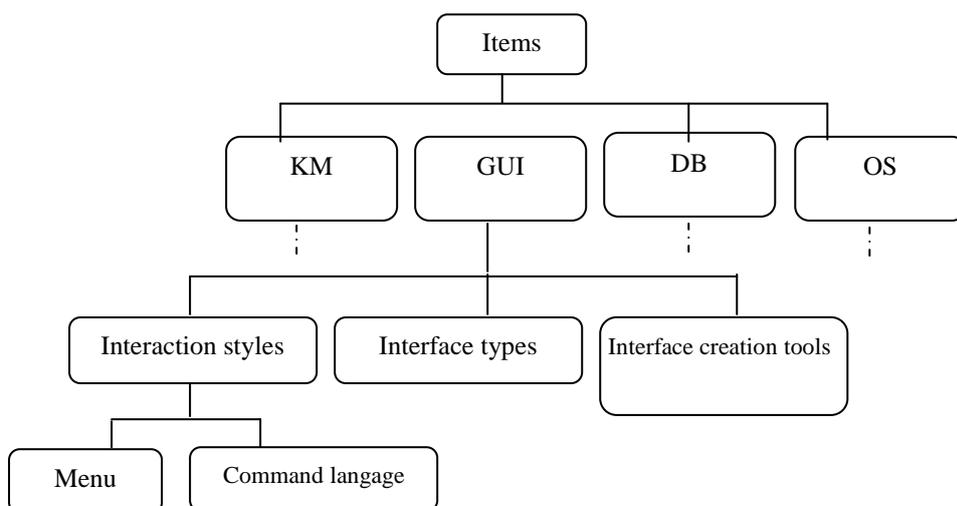


Figure 3.9 Un exemple des descripteurs taxonomiques.

$d_1 = (\text{"GUI"})$

$d_2 = (\text{"GUI"}, \text{"Interaction style"}, \text{"Menu"})$

$d_3 = (\text{"GUI"}, \text{"Interaction style"}, \text{"Command languages"})$

Nous présentons maintenant quelques items que nous avons adopté dans notre travail ainsi que leurs descripteurs:

- t1: GUI<General

GUI<Interaction styles<Menu

GUI<Interaction styles<Command languages

- t2: KM<General

KM< Knowledge types<Explicit

KM< Knowledge types<Implicit

- t3: KM< Knowledge creation process< General

KM< Knowledge creation process<Nonaka's model

- t4: GUI<General

GUI< Interface types< Command Interfaces

GUI< Interface types< Graphique Interfaces

- t5: KM< KM Models< measurement's Models

KM< KM Models<evaluation's Models

- t6: DB<General

DB<Conceptual Model

b) BookCrossing Dataset:

- Les utilisateurs: $U = \{u_0, u_1, \dots, u_c\}$.

- Les items: $I = \{i_0, i_1, \dots, i_k\}$.

- Nous présentons maintenant les descripteurs de quelques items:

- t1: livre anglais et etrangers< boutique

livre anglais et etrangers<Children's books< literature< humorous

- t2 : livre anglais et etrangers < Biographies &Memoirs

livre anglais et etrangers <Law

livre anglais et etrangers<Nonfiction<Crime &Criminals

- t3 :livre anglais et etrangers < Children"s books < Humorous

livre anglais et etrangers < Children's books<People & Places

- t4:livre anglais et etranger < Home & Garden <Animal Care & Pets

- t5:livre anglais et etrangers <Children"s Books < Action &Adventure
Livre anglais et etrangers <Children's books < Explore the World
livre anglais et etrangers <Children's books < Multicultural Stories

- t6 : livre anglais et etrangers < Literature& Fiction<Genre Fiction

2) création du profil d'apprenant

Tous les systèmes de recommandation ont besoin de certains types d'information sur les utilisateurs pour leur générer des recommandations. Ces informations constituent ce qu'on appelle : profil. Les types d'information que doivent être présentes dans un profil sont fortement liés au domaine d'application des systèmes de recommandation. Par exemple, dans le domaine du e-commerce, le seul type d'information nécessaire pour proposer des produits aux clients est leurs goûts ou préférences. Dans le domaine de L'EIAH, les préférences est les caractéristiques cognitives des apprenants ont une importance cruciale, pour cela nous avons introduit non seulement les préférences des apprenants, mais aussi leur niveau et la capacité de stockage de leur mémoire.

La première partie du questionnaire proposé aux apprenants est la nouvelle taxonomie que nous avons créée. Cette taxonomie représente les matières étudiées dans notre département, par exemple : interface homme-machine (GUI), gestion de connaissance (KM), base de données (BDD), théorie de langage (TL),.... la deuxième partie constituée de cinq pré-tests de différents niveaux (très faible, faible, moyen, fort, très fort) afin d'évaluer son niveau. La dernière partie de ce questionnaire est le test qui permet de mesurer la capacité mémoire de l'apprenant en se basant sur la méthode RSI. Le rappel sériel immédiat (RSI) est sans contredit l'un des paradigmes les plus répandus en psychologie (Page et Norris, 1998). RSI consiste à rappeler dans l'ordre de courtes séquences d'items verbaux immédiatement après leur présentation. Le RSI se distingue du rappel libre, qui consiste à rappeler des listes d'items verbaux sans contrainte d'ordre particulière.

Après avoir choisi son domaine d'intérêt, le modèle du contenu pédagogique et passer les cinq pré-tests ainsi que le test du RSI, le profil taxonomique de l'apprenant sera établi. Les conversions des évaluations d'un utilisateur à des préférences taxonomique est une technique qui se réfère à Ziegler (Ziegler, 2004). Dans notre travail, un profil taxonomique est

représenté par un vecteur dont les éléments sont les poids des termes (thèmes) qui représente le domaine d'intérêt sélectionné par l'apprenant, et les scores de différents tests :

$$\vec{v}_{ax} = (S_1, \dots, S_m, \dots, S_{t_1}, \dots, S_{t_5}, t_{s1}, \dots, t_{s5}, mc)$$

$$s_i = \frac{\sum_{j=1}^q p_i(t_j)}{|nb|} \quad (3.1)$$

S_{ix} représente le score du pré-test x . t_{s1} est le temps de réponse au premier pré-test. mc reflète la capacité mémoire de l'apprenant (dans la section suivante nous présentons la formule qui permet de mesurer la capacité mémoire).

La formule (3.1) calcule le score du thème i (S_i). Rappelons qu'un nouvel apprenant n'a aucune évaluation, pour cela nous avons remplacé les évaluations explicites d'un apprenant par les scores des thèmes qu'il a choisis. Le score du thème i est calculé en faisant sortir, tout d'abord, tous les descripteurs de tous les items relatifs à son domaine d'intérêt (dans ce travail, il y'a six items), et calculer le poids du thème (sa fréquence d'apparition dans les descripteurs de ces items). Le score du thème reflète la moyenne de ses poids. $p_i(t_j)$ est le poids du thème i dans l'item j . $|nb|$ est le nombre des items contenant le thème i . La valeur de ce score indique à quel point cet apprenant préfère ce thème. Par exemple si l'apprenant a choisis : GUI/styles d'interaction/menu, les scores des thèmes: menu, styles d'interaction et GUI, seront calculés comme suit :

$$s_{GUI} = \frac{P_{GUI}(t_1) + P_{GUI}(t_4) + P_{GUI}(t_7)}{6} = \frac{3+3+3}{6} = 1.5$$

$$s_{InteractionStyle} = \frac{P_{InteractionStyle}(t_1) + P_{InteractionStyle}(t_4) + P_{InteractionStyle}(t_7)}{6} = \frac{2+0+0}{6} = 0.33$$

$$s_{Menu} = \frac{P_{Menu}(t_1) + P_{Menu}(t_4) + P_{Menu}(t_7)}{6} = \frac{1+0+0}{6} = 0.66$$

Les scores des thèmes dans un vecteur taxonomique sont normalisés tels que :

$$\forall a_i \in A: \sum_{k=1}^D v_{ik} = 1$$

où v_{ik} est le $k^{i\text{eme}}$ composant du vecteur taxonomique, et D est le nombre des composants de ce vecteur.

Un exemple d'un questionnaire en utilisant la nouvelle taxonomie (Benhamdi et al., 2016):

Quel est votre domaine d'interet?

GUI

Interaction styles

Menu

Command languages

KM

Knowledge types

Explicit

Implicit

KM

Knowledge creation process

Nonaka's model

GUI

Interface types

Command Interfaces

Graphique Interfaces

Quel type de contenu vous préférez?

Theoretical concepts.

Theoretical concepts and exercises.

Projects.

La base Bookcrossing utilise la taxonomie proposée par le site d'Amazon (<http://www.Amazon.com>). Un exemple d'un questionnaire en utilisant cette taxonomie est présentée par la suite :

Quel est votre domaine d'intérêt?

livre anglais et étrangers

boutique

Children's books

Literature

humorous

Action & Adventure

Explore the World

Multicultural Stories

Literature & Fiction

Genre Fiction

People & Places

Biographies & Memoirs

Law

Nonfiction

Crime & Criminals

Home & Garden

Animal Care & Pets

DVD & Blu-ray

Action, Aventure, Policier et Thriller

Action

Films catastrophe

Policier et Thriller

quell type de contenu vous préférez?

Theoretical concepts.

Theoretical concepts and exercices.

Problem resolution.

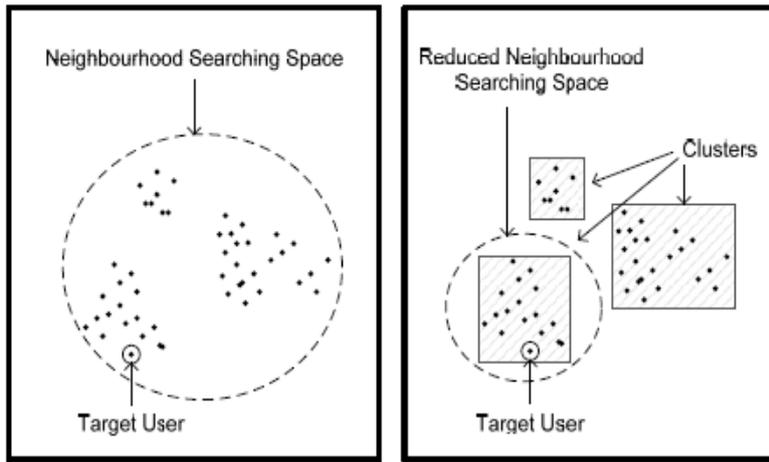
Problem resolution with exercices.

Theoretical concepts and problem resolution

Projects.

3) Regroupement

Le *clustering* est largement utilisé dans les systèmes de recommandation notamment les systèmes basé sur le filtrage collaboratif pour améliorer la qualité de recommandation (Sarwar et al., 2002). Dans les systèmes de recommandation basés sur le filtrage collaboratif, la prédiction des évaluations pour un utilisateur actif se fait à la base des opinions des utilisateurs similaires en termes de préférences, ce qui nécessite la formation des groupes des similaires de cet utilisateur. Selon Weng (Weng, 2008), la prédéfinition des groupes des similaires (clusters) permet de réduire l'espace de recherche (figure 3.10).



(a) : finding neighbours from the entire user set. (b) : finding neighbours from the clustered user set.

Figure 3.10 Réduction de l'espace de recherche en utilisant *clustering* (Weng, 2008).

La prédéfinition des clusters a l'inconvénient d'augmenter les calculs et le temps de réponse, ce qui nous amené à adopter une autre technique permettant de pallier à ce problème. Le principe de cette technique est le calcul de similarité entre l'apprenant actif et tous les autres apprenants. Ces calculs vont déterminer pour l'apprenant actif les membres de son groupe. Un nombre n des apprenants les plus proches taxonomiquement va être choisi tel que la somme des similarités de ces apprenants soit égale au moins à 1. Bien évidemment, les apprenants seront ordonnancés selon leurs similarités taxonomiques avant le calcul de n . Figure 3.11 illustre le principe de cette technique.

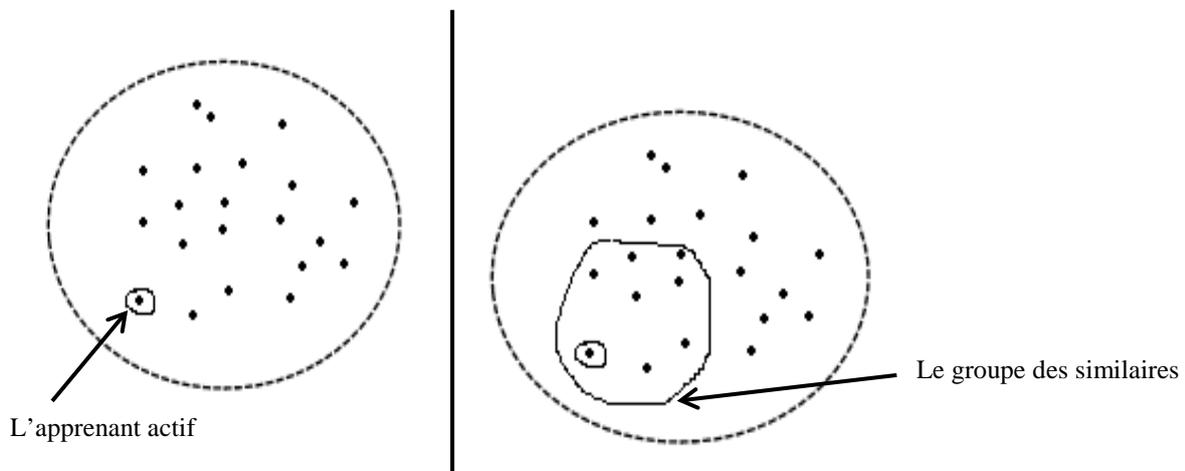


Figure 3.11 La formation des groupes des similaires (NPR-eL).

Dans le but de définir les plus proches voisins en se basant sur la similarité des préférences entre les utilisateurs, Weng a choisit la mesure la plus utilisée : Corrélation de *Pearson* (Breese et al., 1998). Comme nous avons déjà mentionné, nous avons mesuré la similarité entre l'apprenant actif et les autres apprenants prenant en compte non seulement leurs préférences ou centres d'intérêts, mais aussi plusieurs d'autres facteurs liés au système cognitif de l'apprenant. Donc, nous avons proposé une nouvelle mesure de similarité pour introduire ces facteurs (équation 3.2). Le premier terme de cette similarité permet de mesurer la similarité entre apprenants en termes de niveau scientifique, tandis que le deuxième (3.3) calcule la différence entre le temps de réponse aux différents pré-tests de l'apprenant x et l'apprenant y . Le troisième terme (3.4) reflète la similarité entre apprenants en termes de préférences d'item en utilisant la mesure de corrélation de *Pearson*. Le dernier terme (3.5) retourne la similarité entre la capacité de mémoire de l'apprenant x et l'apprenant y . La nouvelle formule de similarité est la suivante :

$$new_sim(x, y) = \sum_{n=1}^L n\alpha(Cx_n - Cy_n) \quad \alpha \in [0..1] \quad (3.2)$$

$$+ \frac{1}{\varphi(tx, ty)} + sim(x, y) + (mc_x - mc_y)$$

où C_{xn} est la note qu'a eu l'apprenant x après avoir passé le pré-test n . L est le nombre total des pré-tests ($L=5$). α est le poids attribué à chaque pré-test. mc_x est la capacité mémoire de l'apprenant x exprimé en *span*.

$$\varphi(tx, ty) = \frac{\sum_{z=1}^L |tx_z - ty_z|}{L} \quad (3.3)$$

$\varphi(tx, ty)$ retourne la différence entre le temps de réponse de l'apprenant x et celui de l'apprenant y . tx_z est le temps de réponse de l'apprenant x au pré-test z .

$$sim(a, i) = \frac{\sum (r_{a,j} - \bar{r}_a)(r_{i,j} - \bar{r}_i)}{\sqrt{\sum_{j \in I} (r_{a,j} - \bar{r}_a)^2 (r_{i,j} - \bar{r}_i)^2}} \quad (3.4)$$

$sim(a,i)$ retourne la similarité entre l'apprenant a et i en utilisant la corrélation de *Pearson*. ra,j est l'évaluation qu'a fait l'apprenant a sur l'item j . \bar{r}_a est la moyenne des évaluations explicites de l'apprenant a .

$$mc_x = \frac{l}{g} \quad (3.5)$$

mc_x est la capacité mémoire de l'apprenant x . l est la taille de la plus longue liste ou le nombre de mots répétés en ordre. g dénote le nombre de mots de la plus longue liste présentée à l'apprenant x .

Après avoir calculé les similarités entre l'apprenant actif a et les autres apprenants, nous pouvons maintenant déterminer son groupe des similaires (uc). Les membres de ce groupe seront choisis selon la formule suivante :

$$t_cluster(a) = \operatorname{argmax}_{i \in A} \text{new_sim}(a,i) \quad (3.6)$$

4) prédiction

Après l'affectation de l'apprenant actif à son groupe des similaires, il va normalement recevoir des recommandations en se basant sur les préférences des membres de son groupe ou ses plus proches voisins. Selon Weng, l'importance d'un item peut être identifiée par sa popularité et la moyenne des évaluations qu'a eues au niveau de ce groupe. Un item qui a été évalué par un nombre significatif d'apprenants, ceci ne signifie pas que cet item est populaire, car il peut recevoir des notes faibles. Pour cela, nous n'avons pris en compte que les valeurs égales ou dépassent la moitié de la note maximale (la note maximale est 1 sur l'échelle [0-1], donc la moitié est : 0.5) dans le calcul de la popularité (3.7).

$$\sigma(uc, t_j) = \frac{\sum_{u \in uc} r(a, t_j)}{|uc|}, \quad r(a, t_j) \in [0.5.. 1] \quad (3.7)$$

$\sigma(uc, t_j)$ renvoie la popularité de l'item j dans le groupe uc . $r(a, t_j)$ est l'évaluation qu'a fait l'apprenant $a \in uc$ sur l'item $j \in T$. uc est le groupe des similaires. $|uc|$ est le nombre des membres de ce groupe. La formule (3.8) calcule la moyenne des évaluations explicites qu'a eues l'item j au niveau du groupe uc ($\psi(uc, t_j)$).

$$\psi(uc, t_j) = \frac{\sum_{a \in uc} r(a, t_j)}{|uc(t_j)|} \quad (3.8)$$

$|uc(t_j)|$ signifie le nombre des apprenants similaires qui ont évalué l'item j .

Pour prédire la préférence d'un item t par un apprenant ($rank(a,t)$), NPR-eL (ainsi que CSHTR) vérifie tout d'abord si cet item est préféré dans son groupe ($cpref(uc,t_j)$), puis il vérifie la similarité taxonomique entre l'apprenant et l'item.

$$cpref(uc,t_j) = \beta \times \psi(uc,t_j) + (1 - \beta) \times \sigma(uc,t_j) \quad (3.9)$$

$$rank_{a,t} = \delta \times cpref(uc,t) + (1 - \delta) \times t_sim(\vec{v}_a, \vec{t}) \quad (3.10)$$

$$t_sim(\vec{v}_a, \vec{v}_t) = \vec{v}_a \times \vec{v}_t \quad (3.11)$$

$t_sim(\vec{v}_a, \vec{v}_t)$ mesure la similarité taxonomique entre l'apprenant a et l'item t . Le seuil β dans la formule 3.9, sert à ajuster le poids dans le calcul de la préférence de l'item t_j , dans le groupe uc entre les évaluations qu'il a eues et sa popularité.

Le vecteur taxonomique de l'apprenant a est: $\vec{v}_a = (S_1, \dots, S_m)$.

$\vec{v}_t = (X_1, \dots, X_h)$ est le vecteur taxonomique de l'item t , S_i le score du thème i calculé par la formule (3.1). Par exemple le vecteur taxonomique de l'item t_1 est :

$$\vec{v}_{t1} = (X_{GUI}, X_{interaction\ style}, X_{menu}, X_{command\ language}).$$

Le score d'un thème i (X_{ti}) dans l'item t est calculé par la formule (3.12).

$$X_{ti} = f_i / nb_topics(t) \quad (3.12)$$

Où f_i est la fréquence d'apparition du thème i dans les descripteurs de l'item t . $nb_topic(t)$ dénote le nombre des thèmes qui composent les descripteurs de cet item (les descripteurs de l'item t_1 sont composés de 7 thèmes). Nous calculons par la suite, le vecteur taxonomique de l'item t_1 :

$$\vec{v}_{t1} = (\frac{3}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7}) = (0.42, 0.28, 0.14, 0.14)$$

B. l'algorithme du NPR-eL

Après avoir calculé leurs rangs ($rank$) ou leurs notes, les items seront ordonnancés selon leurs notes et le système va proposer à l'apprenant actif les k premiers items. L'algorithme de recommandation détaillé de la technique NPR-eL est listé ci-dessous :

Algorithm NPR-eL recommend (a, k)

Where a \in A is a given target student

K is the number of items to be recommended

- 1) SET T, the candidate items list
- 2) FOR EACH $t \in T$
- 3) SET $uc = t_cluster(a)$
- 4) SET $rank_{a,t} = \varphi \times cpref(uc,t) + (1-\varphi) \times t-sim(\vec{v}_a, \vec{v}_t)$
- 5) END FOR
- 6) Return the top k items with highest $rank_{a,t}$ score to a.

Comme nous avons déjà mentionné, le système va maintenant ajouter les URLs des items recommandés dans le *manifest* (un exemple d'un *manifest particulier* est présenté dans le chapitre suivant).

C) Les caractéristiques principales du NPR-eL :

Notre technique de recommandation NPR-eL est caractérisée par :

- contrairement aux autres systèmes de recommandations basés sur le filtrage collaboratif et basé sur le contenu, l'apprenant n'est pas obligé de lire et évaluer quelques documents au début pour déterminer ses préférences. Pour gagner du temps et d'effort, l'apprenant doit seulement répondre au questionnaire proposé et choisir directement son domaine d'intérêt.
- NPR-eL est plus personnalisée ou plus proche de l'apprenant par rapport aux systèmes de recommandations existants parce que les recommandations sont générées prenant en compte plusieurs caractéristiques importantes liées à l'apprenant notamment la capacité de sa mémoire.
- NPR-eL calcule les poids (fréquence d'apparition) des termes, la similarité entre les apprenants et la popularité d'un item de façons plus rigoureuse.

3.4.1.2 la création d'une unité d'apprentissage (UoL)

Nous commençons tout d'abord par la définition d'une unité d'apprentissage (UoL), ça sert à quoi et quelle est sa structure ? Une unité d'apprentissage représente tout simplement un cours, autrement dit, c'est le résultat d'implémentation d'une situation d'apprentissage individuelle ou collaborative. Selon (Koper et Tattersall, 2005), une unité d'apprentissage (*unit of learning*) est un terme abstrait pour exprimer un cours, un module d'enseignement, un leçon, etc. Les unités d'apprentissage peuvent varier en fonction du contexte, des pratiques et des approches théoriques utilisées. Du point de vue d'IMS une unité d'apprentissage

(représentant un cours) est composée d'une partie « organisation du contenu » et une partie « ressources » (figure 3.12).

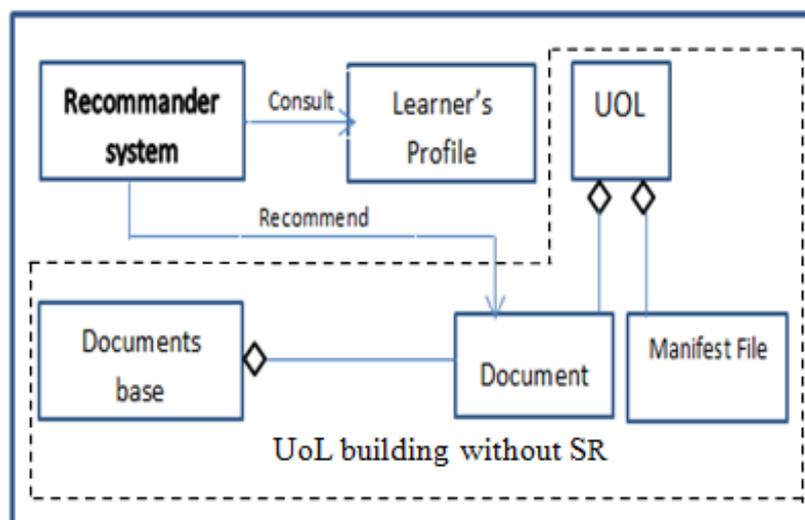


Figure 3.12 Structure d'une unité d'apprentissage.

L'unité d'apprentissage est englobée dans une structure IMS intitulée « content package ». L'organisation du contenu (éléments pédagogiques) est décrite par un fichier XML appelé : *manifest*. Les ressources sont des liens sur des contenus pédagogiques qui peuvent être : des documents (divers formats, Word, pdf), des pages web, des images, mais aussi des fichiers instances d'autres standards (comme des tests d'évaluation en IMS Question & Test Interoperability (Giacomini, Trigano et Alupoie, 2003)). Dans ce travail, ces documents (items) ne sont pas choisis par l'enseignant, mais ceux sont proposés par le nouvel système de recommandation NPR-eL.

3.4.2 La phase d'apprentissage

Dans cette phase, l'unité d'apprentissage créée dans l'étape précédente sera exécutée par l'un des environnements d'exécution ou *players*, nous avons choisi l'environnement *coppercore*. Au début, le système de recommandation NPR-eL consulte le profil d'apprenant actif pour lui recommander les documents appropriés. L'apprenant commence maintenant à exécuter les différentes activités citées dans le *manifest*. Après la lecture des documents recommandés, l'apprenant est invité à évaluer explicitement ces documents.

L'exécution d'une UoL en utilisant *coppercore* passe par trois étapes : la première consiste à valider cette UoL. La préparation du processus d'apprentissage se fait en seconde étape. La troisième étape est celle de l'apprentissage.

- a) validation de l'UoL : en premier lieu, le système vérifie si le *manifest* est bien formé, c.-à-d, il possède un fichier DTD (Document Type Data) ou schéma. Ce type de fichiers permet de décrire la structure d'un document XML. Puis, le *player* vérifie si le *manifest* est valide. Un fichier XML est dit valide s'il dispose d'une feuille de style (*steel sheet*) qui définit sa mise en forme. le *player* vérifie aussi si les ressources mentionnées dans le *manifest*, existent dans le package de l'unité d'apprentissage.
- b) préparation de l'apprentissage : l'apprenant doit créer son session en précisant son rôle (*learner*). après l'ouverture de la session de l'apprenant, les activités d'apprentissage et les différentes ressources seront affichées.
- c) l'étape d'apprentissage : Dans cette étape l'apprenant peut réaliser les activités affichées et exploiter les différentes ressources.

Le diagramme d'activité (l'un des diagrammes du langage UML) suivant résume les différentes étapes d'exécution d'une UoL :

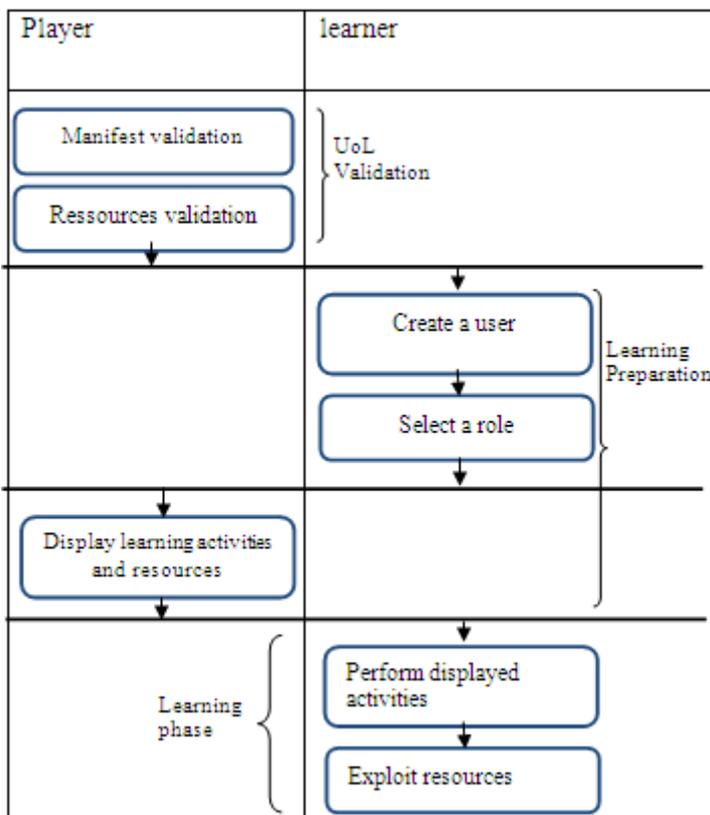


Tableau 3.4 Les étapes d'exécution d'une UoL.

Le processus d'apprentissage peut être résumé par le diagramme d'activité suivant :

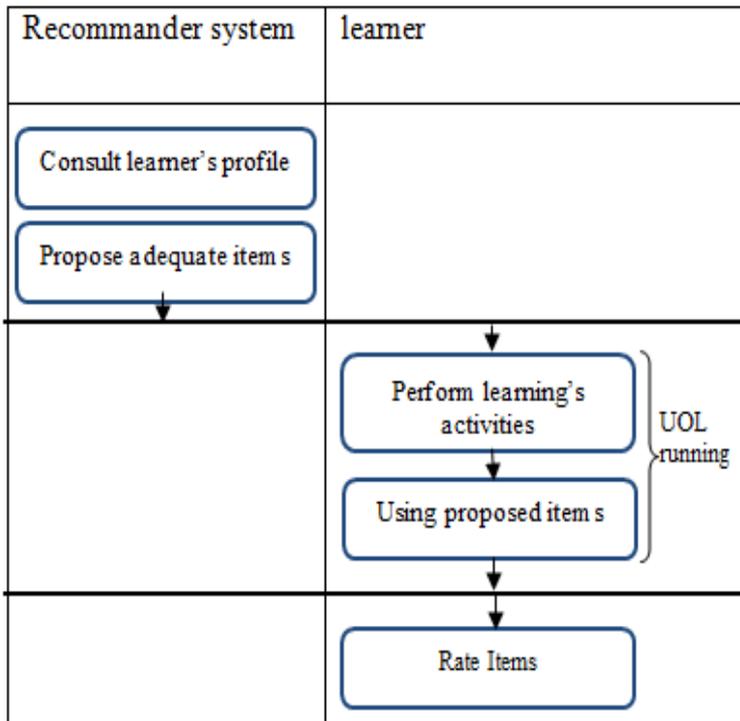


Tableau 3.5 Le processus d'apprentissage.

3.5 Conclusion

Nous voulons, à travers la technique NPR-eL décrite dans ce chapitre, profiter des avantages de l'information taxonomique, prouvée par l'expérimentation faite pour la validation de CSHTR (Weng, 2008). Nous avons exploité aussi l'effet positif des caractéristiques cognitif de l'apprenant sur la qualité de recommandation et de l'apprentissage, sans oublier la bonne solution proposée pour le traitement du problème de démarrage à froid souvent rencontré dans les systèmes de recommandation. Mais on ne peut juger la puissance de ces choix stratégiques qu'après la phase d'expérimentation qui sera détaillée dans le chapitre suivant.

Chapitre 4

Expérimentation et évaluation

4.1 Introduction

Nous arrivons dans cette partie à la validation de notre approche NPR-eL sur deux corpus de données : the University Dataset et Book-Crossing Dataset. Nous avons utilisé la même base qui a été utilisée pour valider la technique CSHTR (Book-Crossing) afin de permettre une comparaison, qui reste- pour le moment - partielle, entre ces deux techniques. Nous avons testé notre système de recommandation prenant en compte deux situations. La première situation consiste à générer des recommandations sans la prise en compte de la capacité de mémoires des apprenants, pour cela nous avons utilisé le système: Boosted_CSHTRL (Boosted Cold Start Hybrid Taxonomic Recommender System for e-Learning) (Benhamdi et Séridi, 2011). Dans la deuxième situation, nous avons introduit cette caractéristique (NPR-eL), calculée via la méthode Rappel Sériel Immédiat (RSI) (Benhamdi et al., 2016). Nous mesurons sa performance en utilisant les métriques d'évaluation : *Rappel*, *Précision* et *F1*. Afin de démontrer les résultats de cette expérimentation, nous avons créé un prototype de la nouvelle plateforme d'apprentissage personnalisé: NPLE (New Personal Learning Environment), au sein duquel le nouvel système de recommandation : NPR-eL, est intégré. Figure 4.1 illustre la page d'accueil du NPLE.



Figure 4.1 L'environnement NPLE.

La page d'accueil du NPLE offre un certain nombre de fonctionnalités aux apprenants et enseignants. Dans cet environnement les enseignants peuvent créer et ajouter des documents (figure 4.2). L'apprenant via l'accès à son espace, peut répondre au questionnaire, passer les tests, recevoir des recommandations et réaliser les activités d'apprentissage (figure 4.3).

Figure 4.2 Espace enseignant.

nce/abdallah/apprenant.php



Figure 4.3 Espace apprenant.

4.2 Données exploitées

4.2.1 University Dataset

Comme nous n'avons pas trouvé une taxonomie appropriée au domaine de e-Learning, nous avons proposé une taxonomie qui représente quelques modules enseignés au niveau du département de l'informatique, Université 08 mai 1945, Guelma, Algérie (une partie de cette taxonomie, qui représente quelques items du Dataset, est décrite dans la section 3.5). Nous avons créé un corpus de données (items, apprenants et leurs évaluations explicites) en se basant sur la nouvelle taxonomie que nous avons proposée pour e-Learning (Bnehamdi et al.,

2016b). Nous avons proposé 15 items à 10 étudiants en troisième année licence (Université 08 mai 1945, Guelma, Algérie) afin de les évaluer. La figure 4.4 montre les évaluations explicites de l'apprenant a1. Le tableau 4.1 représente les évaluations explicites qu'ont faites les dix étudiants.

the next table represents ratios

id learner	id paper	value
1	1	0.7
1	2	1
1	3	0.1
1	4	0.4
1	5	0.5
1	6	0.9

Figure 4.4 Les évaluations réelles de l'apprenant a1.

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10
t1	0.7	0.8	0.9	0.0	1.0	0.2	0.0	0.9	0.2	0.9
t2	1.0	0.2	0.1	0.5	0.0	0.1	0.5	0.9	0.1	0.1
t3	0.1	0.1	0.1	0.1	0.5	0.5	0.2	0.1	0.2	0.9
t4	0.4	0.5	0.7	0.3	0.2	0.8	0.3	0.1	0.1	0.7
t5	0.5	0.8	0.6	0.1	0.3	0.2	0.9	0.1	0.4	1.0
t6	0.9	0.2	0.1	0.5	0.9	0.4	0.9	0.4	0.1	0.5
t7	0.0	0.4	0.8	0.8	0.9	0.1	0.1	0.0	0.3	0.8
t8	0.4	0.1	0.2	0.2	0.1	0.3	1.0	0.9	0.4	0.9
t9	0.8	0.3	0.1	0.1	0.1	0.4	0.0	0.1	0.2	0.8
t10	0.9	0.4	0.5	0.5	0.1	0.5	0.5	0.1	0.3	0.9
t11	0.7	0.5	0.8	0.8	0.4	0.8	0.2	0.4	0.0	0.7
t12	1.0	0.9	0.2	0.2	0.0	0.2	0.3	0.0	0.5	1.0
t13	0.0	0.8	0.4	0.4	0.9	0.4	0.9	0.9	0.2	0.6
t14	0.5	0.5	0.1	0.1	0.7	0.1	0.3	0.7	0.3	0.4
t15	0.8	0.4	0.3	0.3	1.0	0.3	0.9	1.0	0.0	0.5

Tableau 4.1 Les évaluations explicites des apprenants.

4.2.2 Book-Crossing Dataset

Nous avons restreint la validation à un petit échantillon de la base Book-Crossing (<http://www.informatik.uni-reiburg.de/~ziegler/BX/>), contenant 278,858 utilisateurs engendrant 1, 149, 780 évaluations sur 271, 379 livres. L'échantillon choisi est composé de

10 utilisateurs : $U = \{u_0, \dots, u_{10}\}$, et 15 items : $I = \{i_0, i_1, \dots, i_{14}\}$, les six premiers items ont été décrits dans la section 3.5.

4.3 Métriques d'évaluation

Se rappelons encore des métriques que nous avons utilisées pour l'évaluation de la recommandation: *Précision*, *Rappel* et *F1*. *Précision* est définie par le rapport entre le nombre d'items pertinents sélectionnés et le nombre total d'items sélectionnés. La précision représente la possibilité qu'un item sélectionné soit pertinent. *Rappel* c'est le rapport entre le nombre d'items pertinents sélectionnés et le nombre total d'items pertinents disponible. La métrique F1 est une combinaison des deux précédentes.

4.4 Expérimentation en utilisant University Dataset

Un apprenant parmi les dix, soit a_1 , représente notre apprenant central, c'est-à-dire sur lequel se déroule le test. Nous avons choisis l'un des utilisateurs passés du système parce qu'il a fait, sur le plan réel, des évaluations sur les items du Dataset, ce qui nous permet de savoir son avis sur n'importe quel item que le système va recommander pour lui.

L'apprenant commence à répondre au questionnaire proposé (l'apprenant choisit son domaine d'intérêt, le modèle du contenu et passe les cinq tests). Figure 4.5 présente un exemple d'un questionnaire en utilisant la nouvelle taxonomie. Il doit aussi passer le test RSI (pour plus de détails sur cette méthode voir la section 3.5) afin de mesurer la capacité de sa mémoire (figure 4.6). Dans la suite, nous montrons les différentes étapes à suivre pour appliquer cette méthode.

- consigne : « je vais vous présenter des mots que vous connaissez et quatre liste: une liste de lettres, liste de chiffres et deux listes de mots. Vous allez répondre ces mots après moi, dans le même ordre que moi»
- « si pour une certaine position vous savez qu'il y'a un mot, mais que vous ne savez plus lequel, vous dite je ne sais pas pour ce mot »
- si le sujet commence à avoir des difficultés, et qu'il risque de se décourager, dire : « c'est tout à fait normale que vous ne savez plus répéter tous les mots que je vous ai présenté. »
- On calcule le nombre de mots correctement répétés dans le bon ordre. Puis on divise ce nombre sur la taille de la plus longue liste présentée à l'apprenant.

- Il est également indispensable d'enregistrer toutes les réponses afin de garantir la validité de la cotation : en effet, il est parfois difficile de juger du degré d'exactitude d'une réponse pendant l'administration de l'épreuve.

Dans quel domaine choisissez-vous ?

INTERFACE GRAPHIQUE

Styles d'interaction

Menu

langages de command

Interface types

Command Interfaces

Graphique Interfaces

KM (Gestion des connaissances)

knowledge types

implicit

explicit

TL (theorie de langage)

Regular Languages

Regular expression

Turing machine

Valider



NIVEAU 1

0 minute(s) 40 secondes 1 dixièmes

1- Les trois parties de logiciels d'un programme de GUI sont:

Fenêtres, boutons, Souris

Composants d'interface graphique, graphiques, codes

Composants d'interface graphique, les écouteurs d'événement, code de l'application

Cadres, Code, Événements

2- Quelle est la seule composante que presque tous les programmes graphiques auront?

Frame

Souris

Bouton

Moniteur

suivant

Figure 4.5 Le questionnaire proposé et un exemple de test (niveau 1).

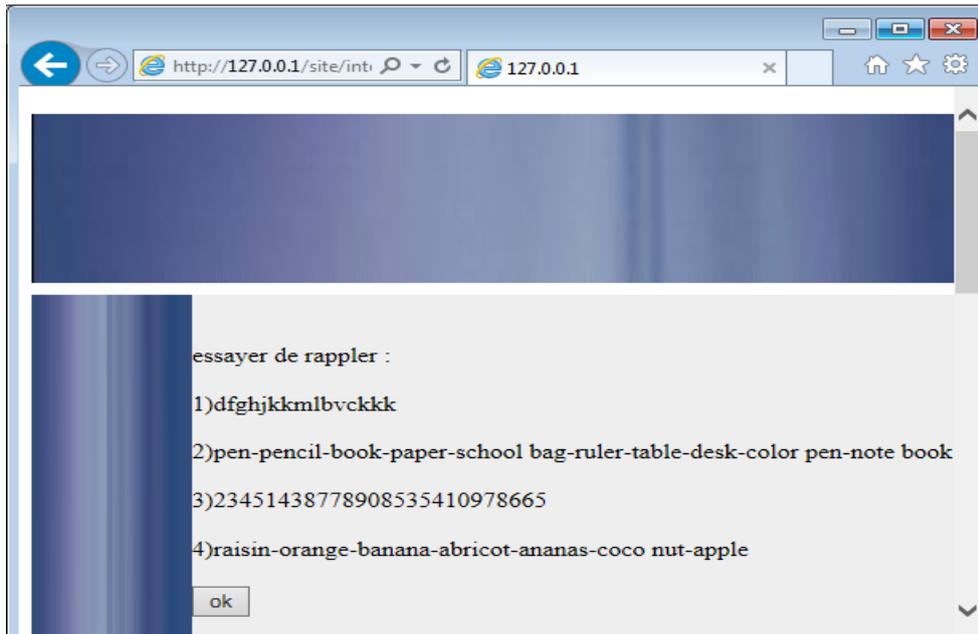


Figure 4.6 Le test de RSI.

Les résultats des tests passés par les étudiants de 'University Dataset' sont montrés dans le tableau 4.2. On peut noter que l'apprenant central a choisi le domaine : GUI/Styles d'interaction/Menu. Les cinq tests vont proposer des questions sur : Graphic User Interface (GUI), et plus précisément sur les menus. Le vecteur (normalisé) qui représente le profil de l'apprenant central a_1 est le suivant :

$$\vec{v}_{a_1} = (S_{GUI}, S_{Interaction\ Styles}, S_{Menu}, S_{t_1}, \dots, S_{t_5}, t_{s_1}, \dots, t_{s_5}, mc).$$

$$\vec{v}_{a_1} = (0.03, 0.01, 0.0, 0.04, 0.02, 0.0, 0.04, 0.01, 0.13, 0.08, 0.21, 0.25, 0.17, 0.03).$$

$S_{GUI} = 0.03$ est le score du thème (*topic*) 'GUI', calculé par la formule (3.1) (section 3.5). $S_{t_1} = 0.04$ est le score du test 1. $t_{s_1} = 0.13$ est le temps de réponse du test 1. $mc = 0.03$ est la capacité de la mémoire de l'apprenant a_1 .

On va tester notre algorithme avec un nombre d'item égal à 5, 10 puis à 15 et comparer les résultats obtenus (figure 4.10) avec ceux de Boosted_CSHTL (prenant en compte seulement les préférences et le niveau des apprenants), en utilisant les métriques d'évaluation Rappel, Précision et F1 (figure 4.11). Le tableau suivant représente les résultats des tests.

	S _{t1}	S _{t2}	S _{t3}	S _{t4}	S _{t5}	t ₁	t ₂	t ₃	t ₄	t ₅	mc
a1	0.9	0.4	0.1	0.9	0.2	3	2	5	6	4	0.7
a2	0.5	0.4	0.1	0.7	0.3	5	7	6	8	5	1.0
a3	0.2	0.7	0.1	0.3	0.1	2	5	7	9	6	0.4
a4	0.6	0.3	0.5	0.2	0.9	6	3	8	9	7	0.6
a5	0.1	0.4	0.2	0.0	0.0	1	2	9	9	8	0.9
a6	0.9	0.9	0.1	0.1	0.1	2	5	6	3	7	0.7
a7	0.2	0.1	0.5	0.3	0.2	2	5	7	1	2	1.0
a8	0.7	1.0	0.2	0.5	0.3	2	9	9	5	7	0.3
a9	0.9	0.1	0.1	0.1	0.4	2	5	6	2	5	0.5
a10	1.0	0.5	0.8	0.9	0.8	1	2	1	3	3	0.3

Tableau 4.2 Les résultats des tests.

S_{t1} : le score du test 1, t₁ : le temps de réponse du test 1, mc : représente la capacité de la mémoire des apprenants. La figure 4.7 montre les items recommandés triés en ordre décroissant selon leurs notes prédites (calculé par NPR-eL). Comme nous l'avons mentionné, Le système ajoute les URLs des items recommandés dans le *manifest* pour pouvoir créer une unité d'apprentissage (UoL) (voir la section 3.5).

The screenshot shows the NPR-eL Recommender System interface. The main content area displays the following information:

- Recommendation results using Guelma dataset**
- learners ordered by similarity**: a10, a4, a2, a3, a9, a8, a6, a7, a5
- uc members (n= 4)**: a10, a4, a2, a3
- Recommended items**: a table with 14 columns (t1 to t14) and 2 rows (items, predicted ratings).

items	t1	t11	t5	t7	t10	t12	t4	t13	t6	t15	t8	t3	t9	t14	
predicted ratings	0.41	0.36	0.34	0.32	0.29	0.29	0.27	0.26	0.17	0.17	0.16	0.15	0.15	0.13	0.10

Figure 4.7 Le résultat de la recommandation (University Dataset).

Nous présentons maintenant un exemple d'un *manifest* particularisé (la définition de l'objectif d'apprentissage, une description détaillée de chaque activité et les items recommandés) en montrant l'ajout des URLs des items recommandés dans l'élément *resources* :

```

<?xml version="1.0" encoding="iso-8859-1"?>
<manifest>
<learning-design identifier="LD" uri="" level="A"/>
<titleLD>learn some concepts </titleLD>
<learning-objective>learn some concepts about gui</learning-objective>
<rôles>
<learner identifier="R-etud">
<title>etudiant</title>
</ learner>
<teacher identifier="R-profesor">
<title>enseignant</title>
</teacher>
</rôles>
<activities>
<learning-activity identifier="N1 ">
<title>lire item </title>
<activity-description>read paper</activity-description>
<resrefidentfierref="cour1"/>
</learning-activity>

<support-activity identifier="activite-control">
<title>Activite control</title>
<activity-description>répondre aux questions de l'apprenant</activity-description>
</support-activity>
</activities>
<method>
<play identifier= P1>
<act identifier =A1>
<role-part identifier="RP-part1">
<title>RP1</title>
<role-ref ref="R-etud" />
<learning -activity-ref ref="N1 " />
</role-part>
<role-part identifier="RP-part2">
<title>RP1</title>
<role-ref ref="R-profesor" />
<support -activity-ref ref=" activite-control " />
</role-part>

</act>
</play>
</method>
<resources>
<resource type=" webContent" id="cour1" href="cour1.pdf" />
</resources>
</manifest>

```

Les items recommandés

Après avoir reçu des recommandations, l'apprenant commence à apprendre, donc on parle de la phase d'apprentissage. Dans cette phase l'unité d'apprentissage créée dans la phase de modélisation sera exécutée en utilisant l'environnement d'exécution (*player*) : Coppercore. Cet environnement est intégré dans notre plateforme d'apprentissage (NPLE). La figure 4.8 montre l'exécution d'un exemple d'UoL. La phase d'apprentissage se termine par l'évaluation explicite des items recommandés. Ces évaluations remplacera les score des thèmes (*si*) (dans cette expérimentation, l'apprenant central a choisi les thèmes : GUI, Styles d'interaction et Menu) dans le profil de l'apprenant. Puis, l'apprenant passe un autre test (post-test) afin d'évaluer la qualité d'apprentissage dans les deux cas : recommandation sans et avec la capacité de mémoire de l'apprenant (Figure 4.9). Les figures 4.12 et 4.13 montrent les résultats du pré-tests et post test pour les dix étudiants.

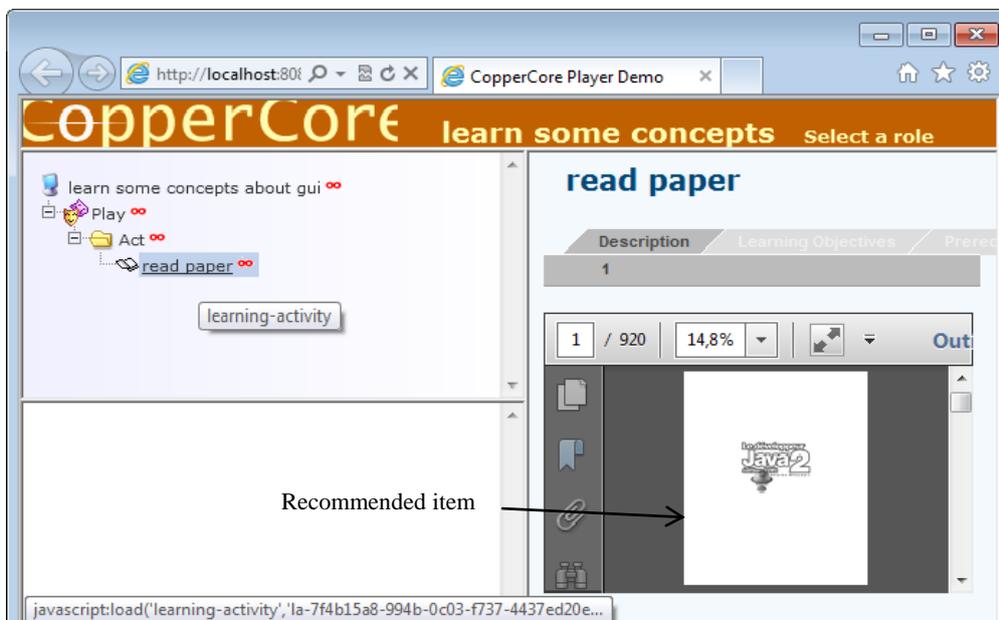


Figure 4.8 Exécution d'un exemple d'UoL.

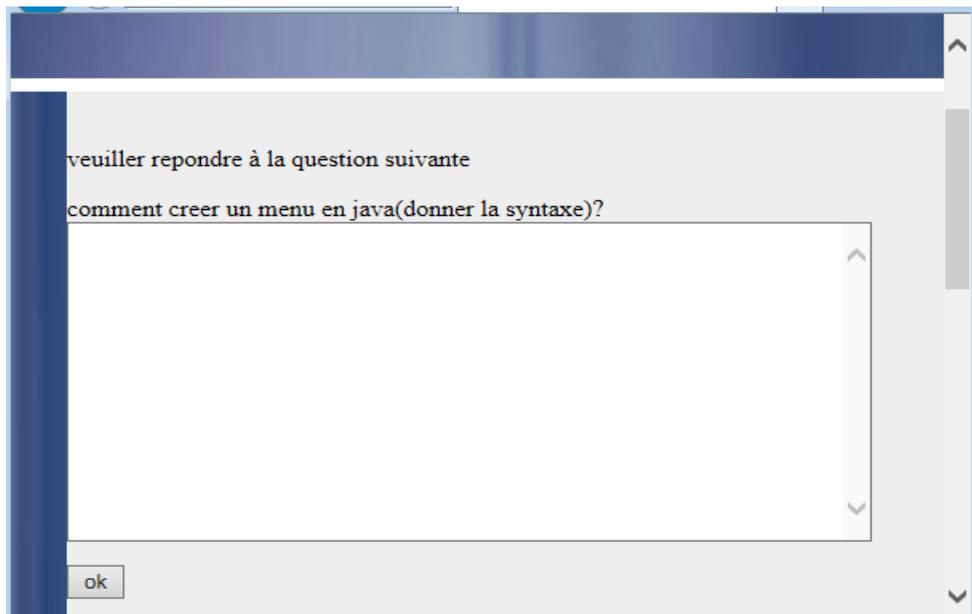


Figure 4.9 Un exemple du post test.

4.4.1 Calcul des métriques d'évaluation

Nous avons calculé les métriques d'évaluations : Rappel, Précision et F1, pour notre système comme suit :

- Nombre totale des items pertinents = $N_p = 10$
- Nombre des items pertinents trouvés = N_t
- Nombre totale des items = N

Recommandation de 5 items

$$\text{Précision} = P = \frac{N_t}{N} = \frac{4}{5} = 0.8$$

$$\text{Rappel} = R = \frac{N_t}{N_p} = \frac{4}{10} = 0.4$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * 0.8 * 0.4}{0.8 + 0.4} = 0.53$$

Recommandation de 10 items

$$\text{Précision} = P = \frac{N_t}{N} = \frac{7}{10} = 0.7$$

$$\text{Rappel} = R = \frac{N_t}{N_p} = \frac{7}{10} = 0.7$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * 0.7 * 0.7}{0.7 + 0.7} = 0.7$$

Recommandation de 15 items

$$\text{Précision} = P = \frac{N_t}{N} = \frac{10}{15} = 0.67$$

$$\text{Rappel} = R = \frac{N_t}{N_p} = \frac{10}{10} = 1$$

$$\text{F1} = \frac{2 * P * R}{P + R} = \frac{2 * 0.67 * 1}{0.67 + 1} = 0.8$$

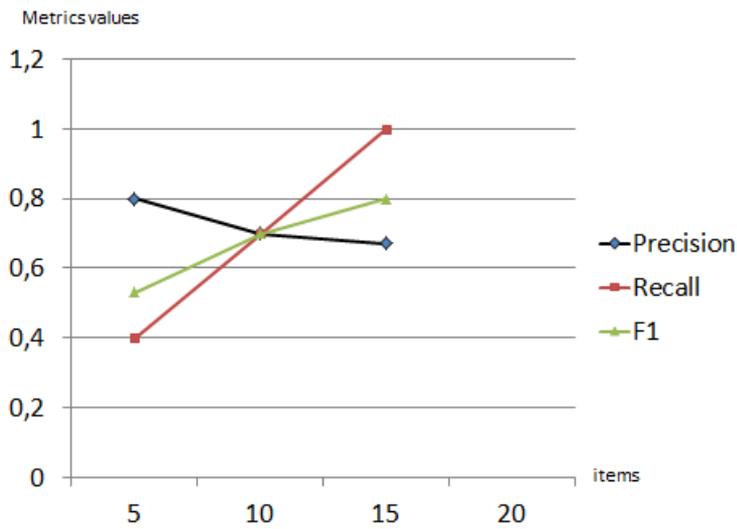


Figure 4.10 L'évaluation de NPR-eL (University Dataset).

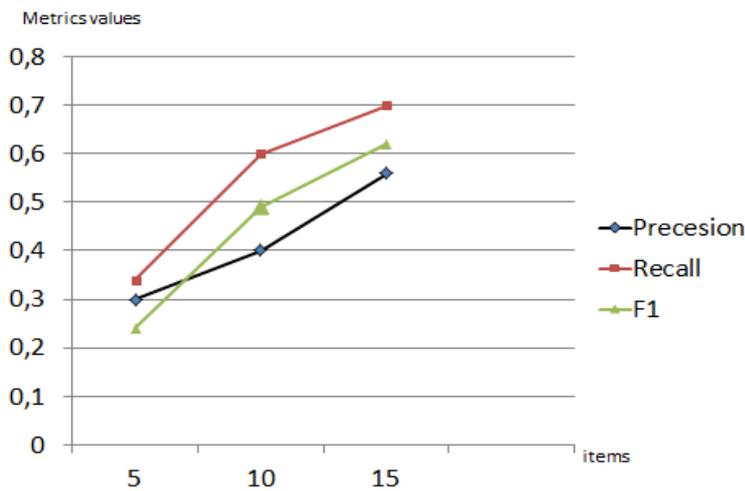


Figure 4.11 Evaluation de Boosted_CSHTTR (Benhamdi et Séridi 2011).

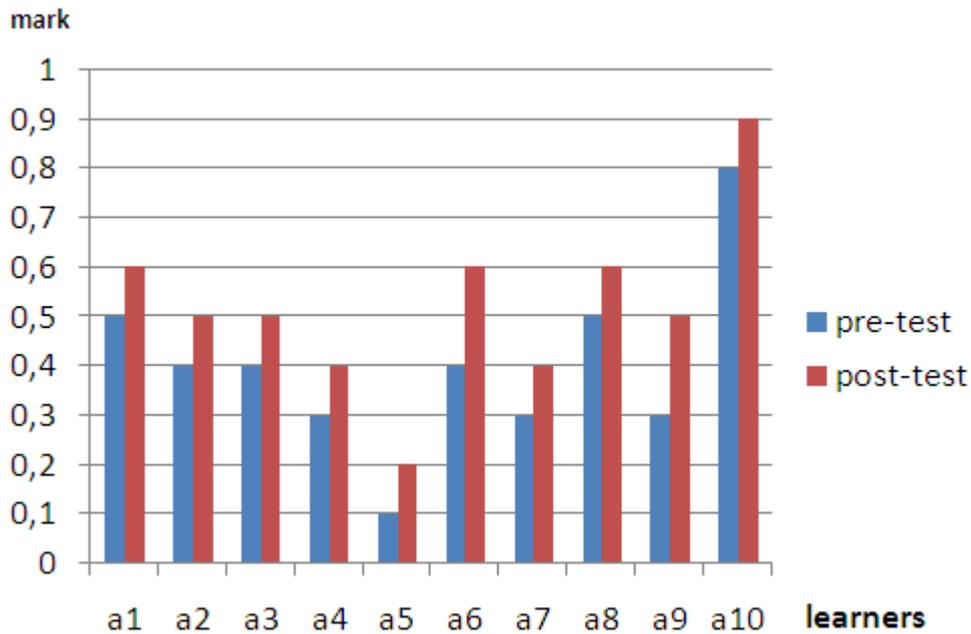


Figure 4.12 Les résultats des pré-tests et post test (Boosted_CSHTL).

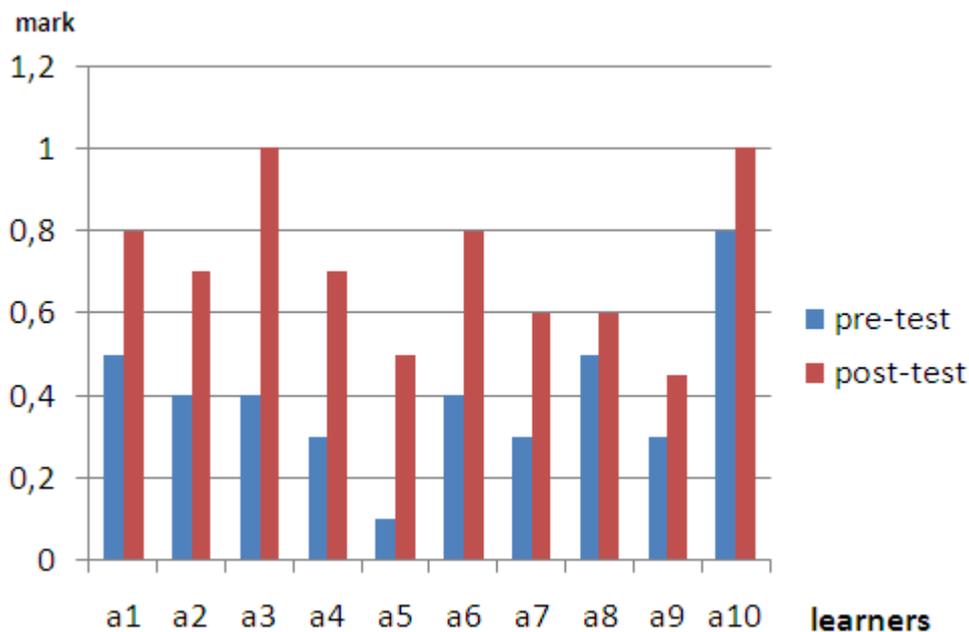


Figure 4.13 Les résultats des pré-tests et post test (NPR-eL).

4.4.2 Explication des résultats

Entrons un peu dans les détails des résultats obtenus (présentés par la figure 4.7) en regardant les différentes valeurs retournés par le système, concernant a1, et contribuant à la détermination de l'ensemble des items recommandés :

1. Présente les apprenants triés en ordre décroissant selon la valeur de similarité entre eux et l'apprenant central a1.

2. Selon les degrés de similarité, le système choisira les membres du groupe des similaires de a1 (uc), qui sont : a10, a4, a2, a3. Dans cette expérimentation, la taille du groupe (n) est égale à 4.
3. Après la construction du groupe, le système consultera tous les items qui ont été évalués par les membres de ce groupe, et il calcule leurs scores en fonction de la moyenne des évaluations qu'ils ont eus, leur similarité taxonomique avec les préférences du groupe, et avec l'apprenant a1, puis le système les ordonnancera selon leurs scores (notes estimées).

4.4.3 Discussion des résultats

Comme nous avons déjà mentionné, l'apprenant central a1 a fait suffisamment d'évaluations sur la base des items utilisés dans cette expérimentation, et les items non évalués sont considérés non appréciés. Donc, nous profiterons de ces évaluations pour comparer la liste des items qui va être recommandés pour lui avec la liste des appréciations réelles.

Analysons maintenant les résultats obtenus dans cette expérimentation. Nous commençons par commenter le tableau 4.3 des items recommandés. Nous avons testé différentes valeurs de k (nombre d'items recommandés) de 5 à 15. Parmi les cinq premiers items (k=5), un seul item est mal apprécié (t7), mais en réalité il n'a pas été apprécié par a1 sauf qu'il soit considéré mal apprécié parce qu'il n'a pas encore une note, par contre les autres sont bien appréciés. Cet item a été recommandé parce qu'il est bien apprécié par la plupart des membres du groupe des similaires (a10, a4, a2, a3). Dans le cas où k = 10 (c'est-à-dire, les dix premiers items), un seul item est mal apprécié (t4), et deux n'ont pas encore une note (t7 et t13). Donc, la plus part des items recommandés sont pertinents. Les évaluations de l'apprenant a10 sont très riche comme celles de l'apprenant central a1, par contre, celles des apprenants : a4 a2 et a3 sont pauvres, ce qui explique la recommandation de quelques items non pertinents. Ces apprenants ont été choisis comme « les apprenants les plus proches » parce que leurs scores des pré-tests sont proches de ceux de l'apprenant central a1 (la moyenne des score des cinq tests de a1 est : moyenne = 0.5, a4 : moyenne = 0.5, a2 : moyenne = 0.4, a3 : moyenne = 0,3).

Nous passons maintenant à commenter les résultats montrés dans les figures 4.12 et 4.13. On peut noter qu'il y'a une différence entre les scores des pré-tests et post test dans les deux cas, ce qui signifie que les apprenants ont acquis de nouvelles connaissances. La différence est importante dans le deuxième cas (recommandation avec la capacité mémoire « NPR-

eL »). Donc, la prise en considération d'un autre facteur dans le processus de recommandation peut améliorer la qualité de la recommandation et l'apprentissage.

nombres d'items (k)	items recommandés
k=5	t ₁ , t ₁₁ , t ₇ , t ₄ , t ₁₀
k=10	t ₁ , t ₁₁ , t ₇ , t ₄ , t ₁₀ , t ₁₂ , t ₄ , t ₁₃ , t ₆ , t ₁₅
k=15	t ₁ , t ₁₁ , t ₇ , t ₄ , t ₁₀ , t ₁₂ , t ₄ , t ₁₃ , t ₆ , t ₁₅ , t ₈ , t ₃ , t ₉ , t ₁₄ , t ₂

Tableau 4.3 Les items recommandés.

4.5 Expérimentation en utilisant Book-Crossing Dataset :

Pour recevoir des recommandations, l'utilisateur central doit suivre les mêmes étapes mentionnées ci-dessus. L'utilisateur u₀ représente notre utilisateur central. Le tableau 4.4 illustre les évaluations explicites faites par u₀ sur les 15 items de la base.

On va tester notre algorithme avec un nombre d'item égal à 5, 10 puis à 15 et comparer les résultats obtenus avec ceux de CSHTR (Weng, 2008) en utilisant la métriques d'évaluation F1 (figure 4.15).

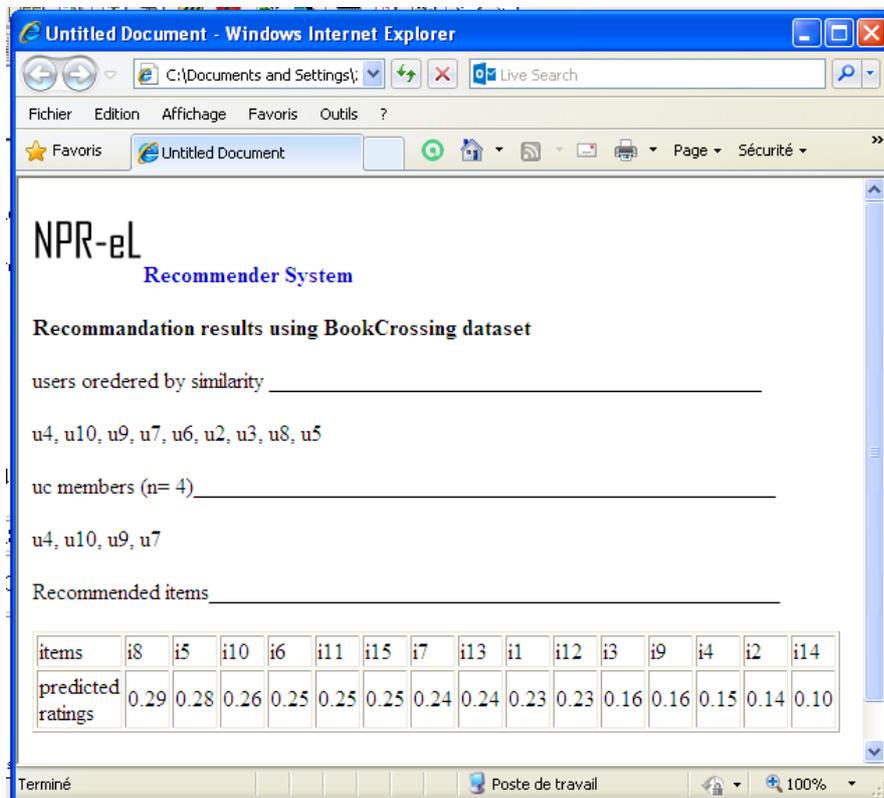


Figure 4.14 : Le résultat de la recommandation.

4.5.1 Calcul des métriques d'évaluation

Nous avons appliqué les formules définies dans la section précédente :

Recommandation de 5 items

$$\text{Précision} = P = \frac{N_t}{N} = \frac{4}{5} = 0.8$$

$$\text{Rappel} = R = \frac{N_t}{N_p} = \frac{4}{10} = 0.4$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * 0.8 * 0.4}{0.8 + 0.4} = 0.53$$

Recommandation de 10 items

$$\text{Précision} = P = \frac{N_t}{N} = \frac{7}{10} = 0.7$$

$$\text{Rappel} = R = \frac{N_t}{N_p} = \frac{7}{10} = 0.7$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * 0.7 * 0.7}{0.7 + 0.7} = 0.7$$

Recommandation de 15 items

$$\text{Précision} = P = \frac{N_t}{N} = \frac{10}{15} = 0.67$$

$$\text{Rappel} = R = \frac{N_t}{N_p} = \frac{10}{10} = 1$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * 0.67 * 1}{0.67 + 1} = 0.8$$

Nous avons obtenu les mêmes valeurs obtenues dans l'expérimentation en utilisant University Dataset (figure 4.10) parce que nous avons trouvés le même nombre des items non pertinents dans chaque cas (k = 5 et k=10).

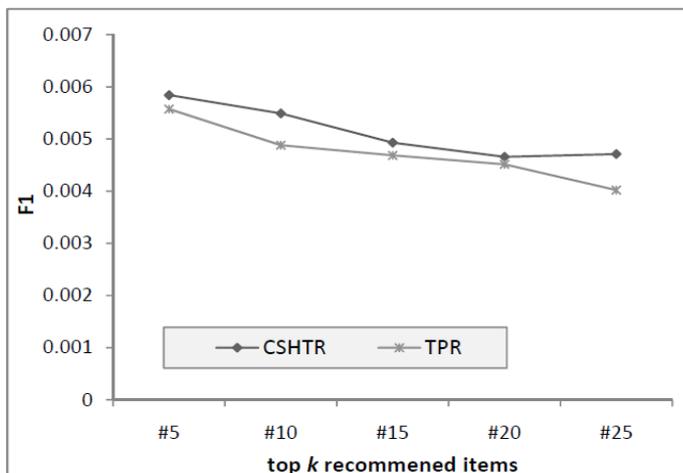


Figure 4.15 Evaluation du CSHTR en utilisant F1 (Weng, 2008).

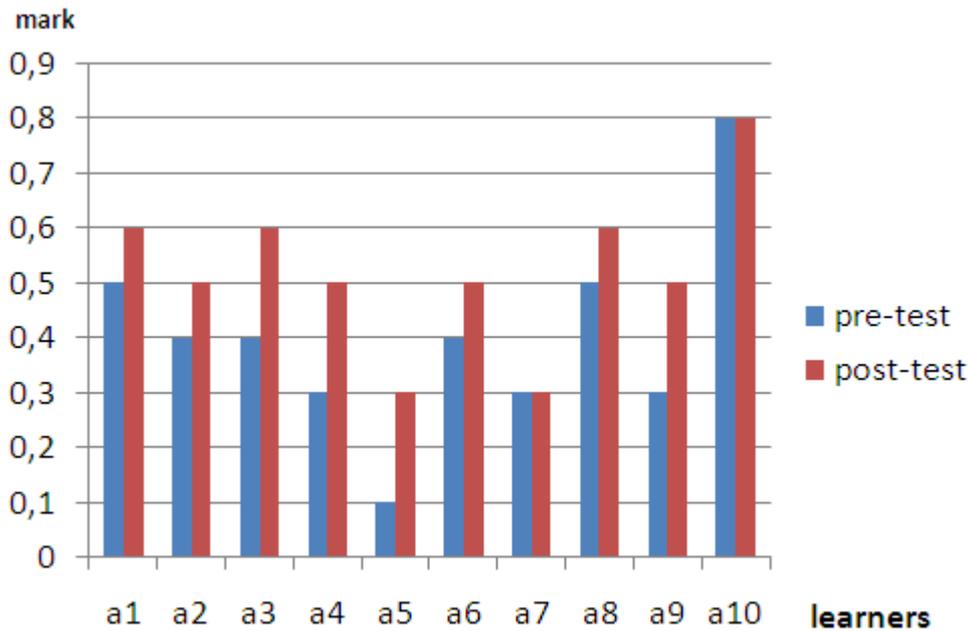


Figure 4.16 Les résultats des pré-test et posttest (Boosted_CSHTRL).

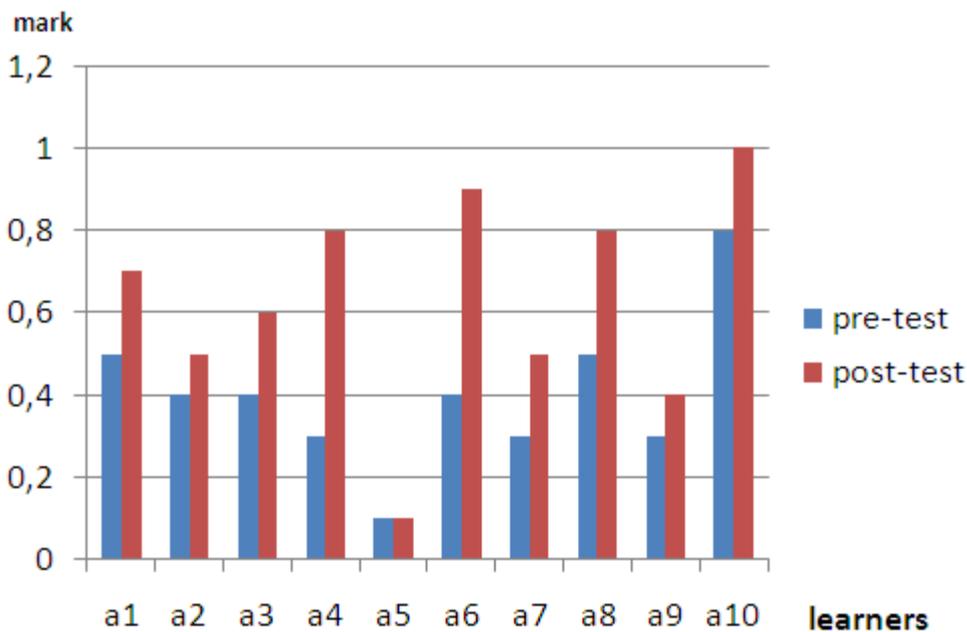


Figure 4.17 Les résultats des pré-tests et post test (NPR-eL).

4.5.2 Discussion des résultats

Nous commentons maintenant le tableau 4.4. Parmi les cinq premiers items recommandés, un seul item est mal apprécié (t8). Il y'a un seul item mal apprécié (t8) et deux n'ont pas encore été appréciés (t7 et t13), parmi les dix premiers items recommandés. Quand on compare entre les valeurs de la métrique F1 obtenues dans cette expérimentation et celles obtenus dans l'expérimentation du CSHTR, on peut conclure que notre approche NPR-eL produit des

recommandations de qualité par rapport à CSHTR ($F1 < 0.006$ pour CSHTR, $F1 = 0.8$ pour NPR-eL). La différence principale entre ces deux systèmes se réside dans le calcul du profil d'apprenant ainsi que la formation des groupes des similaires (uc). CSHTR calcule le profil taxonomique des utilisateurs (cette méthode se réfère à Ziegler (Zeigler, 2004) en se basant sur leurs évaluations explicites (leurs préférences d'item). Donc, la formation des groupes des plus proches voisins se fait en comparant leurs profils. NPR-eL calcule les profils des apprenants en se basant sur leurs domaines d'intérêt et ajoute à leurs profils d'autres informations importantes : leur niveau et capacité de leur mémoire, et par conséquent, le groupe des voisins est formé prenant en considération ces informations.

D'après les résultats des pré-tests et post test, il est très clair que la déviation entre les deux types de tests est plus importante dans le deuxième cas, ce qui confirme notre hypothèse sur l'impact positive de la propriété : la capacité de la mémoire humaine, sur la qualité de recommandation et d'apprentissage.

items	t ₈	t ₅	t ₁₀	t ₆	t ₁₁	t ₁₅	t ₇	t ₁₃	t ₁	t ₁₂	t ₃	t ₉	t ₄	t ₂	t ₁₄
real ratios	0.4	0.5	0.9	0.9	0.7	0.8	0.0	0.0	0.7	1.0	0.1	0.8	0.4	1.0	0.5

Tableau 4.4 Les évaluations réelles faites par u0.

nombres d'items (k)	items recommandés
k=5	i ₈ , i ₅ , i ₁₀ , i ₆ , i ₁₁
k=10	i ₈ , i ₅ , i ₁₀ , i ₆ , i ₁₁ , i ₁₅ , i ₇ , i ₁₃ , i ₁ , i ₁₂
k=15	i ₈ , i ₅ , i ₁₀ , i ₆ , i ₁₁ , i ₁₅ , i ₇ , i ₁₃ , i ₁ , i ₁₂ , i ₃ , i ₉ , i ₄ , i ₂ , i ₁₄

Tableau 4.5 Les items recommandés.

4.6 Discussion globale

L'objectif de cette expérimentation est d'évaluer l'impact des différentes propriétés des apprenants sur la qualité de recommandation et d'apprentissage. Pour atteindre ce but, deux situations ont été étudiées. La première situation consiste à recommander des items aux apprenants prenant en compte seulement deux caractéristiques : leur préférences et niveau, en

utilisant le système de recommandation : Boosted_CSHTR (Benhamdi et Séridi, 2011), testé avec University Dataset (les valeurs des métriques d'évaluations sont montrés par la figure 4.11: $F1 = 0.62$). Le travail présent représente la deuxième situation, d'où une troisième caractéristique liée au système cognitif des apprenants a été ajoutée : leur capacité mémoire (figure 4.10 : $F1 = 0.8$, pour le système de recommandation NPR-eL). Les valeurs de la métrique F1 dans les deux cas montrent que la qualité des recommandations est améliorée quand on ajoute cette propriété. Donc, cette étude comparative prouve que le système de recommandation proposé NPR-eL produit des recommandations de qualité par rapport à Boosted_CSHTR et CSHTR. La personnalisation de l'information est appliquée dans le domaine de l'EAIH afin de motiver les étudiants, donc améliorer la qualité de recommandation conduit effectivement à l'augmentation de la qualité d'apprentissage.

4.7 Caractéristiques du NPLE

L'intégration du système de recommandation NPR-eL dans l'environnement d'apprentissage personnalisé NPLE permet de rehausser la personnalisation, et par conséquence, l'amélioration de la qualité d'apprentissage. Nous citons maintenant les différentes caractéristiques du NPLE :

- La réutilisation des ressources grâce à la modélisation des cours en utilisant le standard IMSLD.
- L'évaluation des étudiants avant et après l'apprentissage peut aider l'enseignant dans la création du contenu pédagogique.
- NPLE permet de guider les étudiants dans leurs processus d'apprentissage en proposant non seulement des documents pertinents, mais aussi les activités d'apprentissage qu'ils doivent réaliser.
- NPLE permet de rehausser la personnalisation par l'intégration de certaines propriétés liées à l'apprenant.

4.8 Conclusion

L'expérimentation faite sur un apprenant a donné de très bons résultats en comparaison avec CSHTR ($F1 < 0.006$ pour CSHTR, $F1 = 0.8$ pour NPR-eL). Sauf que cette performance nécessite encore un échantillon plus large de donnée pour qu'elle soit fixée avec assurance. Ce qui est plus que sur pour le moment, est que notre approche (NPR-eL) propose une meilleure solution au problème de démarrage à froid. C'est le point qui défendra notre approche même si sa performance se diminuera en élargissant l'échantillon de test.

Conclusion et perspectives

L'expansion de l'Internet et du nombre d'applications basées sur le Web tels que les portails d'entreprise, est associée à une prolifération d'information ou d'items dont le volume ne cesse de croître. Devant cette profusion et cette surcharge d'items, l'utilisateur peine à repérer l'information pertinente qui correspond le plus à ses besoins. Dans ce contexte, les systèmes de recommandation ont été développés en vue de faciliter l'accès à ces items pertinents. Leur objectif est d'anticiper les besoins de l'utilisateur en lui fournissant des recommandations d'items jugés pertinents par rapport à ses goûts.

L'un des enjeux pour les systèmes de recommandation est de résoudre le problème de démarrage à froid concernant la nouveauté d'un utilisateur et/ou d'un item. En l'absence des notes de la part de cet utilisateur et/ou sur cet item, il devient impossible pour le processus de filtrage de les intégrer dans les recommandations.

En outre, la précision des recommandations est un défi majeur pour tout système de recommandation dans la mesure où la pertinence des items recommandés permet de contribuer à la satisfaction des attentes de l'utilisateur.

A partir de ces questions de recherche, nous avons proposé dans cette thèse une nouvelle approche de recommandation qui combine le filtrage collaboratif et le filtrage basé sur le contenu (filtrage hybride): NPR-eL (New Personalized Recommender System for e-Learning), s'appuyant sur les préférences taxonomiques et les caractéristiques cognitives des apprenants : leur niveau et la capacité de leur mémoire. Cette approche vise à résoudre le problème de démarrage à froid en proposant un questionnaire. Ce questionnaire est constitué de deux parties : la taxonomie que nous avons développée pour représenter les domaines d'apprentissage (les modules programmés par notre département) constitue la première partie. La deuxième partie est un ensemble de tests de différents niveaux (très faible, faible, moyen, fort et très fort). Les réponses des apprenants au questionnaire permettent de construire leurs profils.

Pour valider cette approche, nous avons mené des expérimentations sur deux corpus de données. Le premier corpus est celui qu'on a créé pour le domaine du e-Learning : University

Dataset. Le deuxième est un corpus standard développé pour le domaine du e-commerce : Book-Crossing Dataset. L'étude comparative que nous avons faite prouve que le système NPR-eL produit des recommandations de bonne qualité par rapport aux systèmes de recommandation : CSHTR (Cold Start Hybrid Taxonomy Recommender (Weng, 2008)) et Boosted-CSHTRL (Boosted- Cold Start Hybrid Taxonomy Recommender for e-Learning (Benhamdi et Seridi, 2011)). Sauf que cette performance nécessite encore un échantillon plus large de donnée (corpus de données) pour qu'elle soit fixée avec assurance.

Dans le contexte du e-Learning, le but des plateformes récentes de formation est de fournir de l'information personnalisée aux apprenants. Pour cela, nous avons intégré le système de recommandation NPR-eL dans une plateforme de formation (NPLE: New Personalized Learning Environnement). Cette plate forme adopte une approche de modélisation de situation d'apprentissage individuelle en se basant sur le langage de modélisation des cours (Benhamdi et al, 2007): IMSLD (IMS Learning Design), centré sur la notion d'activité (Laforcade, 2004). Le modèle conceptuel d'IMSLD définit trois niveaux : A, B et C. Dans ce travail, nous n'avons pris en considération que les concepts du niveau A.

Perspectives

Dans une perspective d'amélioration des performances de notre plate forme d'apprentissage (NPLE), nous envisageons :

- prendre en compte d'autres caractéristiques qui peuvent améliorer la qualité des recommandations.
- faire des expérimentations sur un échantillon plus large de données afin de confirmer la performance de l'approche de recommandation qu'on a proposé.
- finaliser la plateforme NPLE en ajoutant quelques services (ressources) qui peuvent être importants pour l'apprentissage (par exemple : email, convertisseur de documents, etc).

Bibliographie

Adomavicius, Gediminas, et Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *Transactions on Knowledge and Data Engineering*. 7(6), p. 734-749.

Anand, S., et Mobasher, B. 2005. Intelligent techniques for web personalization. *Lecture Notes in Artificial Intelligence*, 3169, p. 1–36.

Anjorin, M., Rensing, C., et Steinmetz, R. 2011. Towards Ranking in Folksonomies for Personalized Recommender Systems in E-Learning. Retrieved from <http://www.crokodil.de>, <http://demo.crokodil.de>

Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., et Jose, J.M. 2010. Enriching User Profiling with Affective Features for the Improvement of a Multimodal Recommender System. *Proceeding of the 2010 ACM International Conference on Image and Video Retrieval*, p. 1–8.

Arnaud, M. 2005. La gouvernance dans le processus de normalisation. Le cas des normes ouvertes pour l'apprentissage en ligne. In Michel Mathien (dir.) *La « Société de l'information »*. Entre mythes et réalités, Bruylant, Bruxelles, p. 201-223

Balog-Crisan, R., Roxin, I., Szilagyi, I. 2009. Ontologies for a semantic quiz architecture. 9th IEE International conference on Advanced Learning Technologies (ICALT 2009), Riga, Latvia, p. 492-494.

Baltrunas, L., et Ricci, F. 2007. Dynamic item weighting and selection for collaborative filtering. In *Web mining 2.0 Workshop, ECML-PKDD 2007*. Springer-Verlag.

Baron, M., Gras, R., et Nicaud, J. F. 1991. Eds. *Actes des deuxièmes journées d'EIAO de Cachan*. Les éditions de l'école normale de Cachan, 1991.

Baron, M. 1995. *EIAO, quelques repères*. 1995.

<http://www.terminal.sgdg.org/articles/65/multimediaron.html>

- Basu, C., Haym, H., et Cohen, W. 1998. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. *AMi/IAAI*, p. 714-720.
- Beauvoir, P., et Sharples, P. 2005. ReLOAD: Reusable e-Learning Object Authoring and Delivery. This project is based at the Universities of Bolton and Strathclyde and funded by JISC, the Joint Information Systems Committee.
- Belkin, N., J., et Croft, B. 1992. In formation filtering and information retrieval: two sides of the same coin. *Commun. ACM.* 35(12), p. 29-38. En ligne. <http://doi.acm.org/10.1145/138859.138861>.
- Benhamdi S., et Séridi, H. 2007. Modélisation et standardisation de situations d'apprentissage collaboratif. 4^{ème} Symposium International en EAD : Technologies pédagogiques, Sousse, Tunisie.
- Benhamdi S., et Séridi, H. 2011. Pedagogical content personalization. In the proceeding of ITHET'11 conference, vol (10), Turkey.
- Benhamdi S., Babouri, A., et Chiky R. 2016. Personalized recommender system for e-Learning environment. *Educ Inf Technol* (2016). doi:10.1007/s10639-016-9504-y, Springer.
- Berrut, C., et Nathalie D. 2003. Filtrage collaboratif. In *Assistance intelligente à la recherche d'informations*, Hermes Science Publications, p. 30. En ligne. <<http://www-mrim.imag.fr/publications/2003/CBOO1/berrut03b.pdf>>.
- Bertrand-Pierron, Y. 2006. Transfert de technologies sur le filtrage collaboratif : intégration des techniques de filtrage collaboratif sur un portail de gestion de contenu. Mémoire de D.E.A., UHP University Nancy 1.
- Betbeder-Matibet, M.L. 2003. *Symba: un environnement malléable support d'activités collectives en contexte d'apprentissage*. Thèse de doctorat de l'Université du Maine, France
- Billsus, D. ; Brunk, C. ; Evans, C. ; Gladish, B. et Pazzani, M. 2002. Adaptive interfaces for ubiquitous web access. *Communications of ACM*, 45(5), p. 34–38.
- Bobadilla, J., Seradella, F., et Hernando, A. 2009. Collaborative filtering adapted to recommender systems of e-Learning. *Knowledge-Based Systems*, 22, p. 261–265.
- Boone, G. 1998. Concept features in Re: Agent, an intelligent Email agent. 2nd international conference on Autonomous agents. Minneapolis, Minnesota, USA.

- Bouzeghoub, M., et Kostadinov, D. 2005. Personnalisation de l'information : Aperçu de l'état de l'art et définition d'un modèle flexible de définition de profils. Actes de la 2nde Conférence en Recherche d'Information et Applications CORIA, 2005, p. 201-218.
- Bradley, K., Rafter, R., et Smyth, B. 2000. Case-Based User Profiling for Content Personalisation. Intl. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems (AH), p. 62–72.
- Breese, J. S., Heckerman, D., Kadie, C. 1998. Empirical Analysis of Predictive Algorithms for collaborative Filtering. Proceeding of 14th conference on Uncertainty in Artificial Intelligence. Madison, MI.
- Broisin, J. 2006. Un Environnement Informatique Humain au Service de la visualisation et de la gestion des objets pédagogiques. Thèse de Doctorat, Université Paul Sabatier-Toulouse III, France.
- Broisin, J., Brut, M., Butoianu, V., Sedes, F., et Vidal, P. 2010. A personalized recommendation framework based on CAM and document annotations. *Procedia Computer Science, Science Direct*, 12(2), p. 2839-2848.
- Bruillard, E. 1997. Les machines à enseigner. Ed. Hermès, Paris.
- Brusilovsky, P. 2001. Adaptive hypermedia. *User Modeling and User Adapted Interaction*, p. 87-110.
- Brusilovsky, P. 2003. Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology*, p. 487-497.
- Budzik, J., et Hummond, K. 2000. User interactions with every applications as context for just-in-time information access. *Proceedings of the 5th international conference on intelligent user interfaces, Mars 2000*, p. 44-51.
- Burke, R. (2000). Knowledge-based recommender systems. *Encyclopedia of Library and Information Systems*, 69(32).
- Burke, R. 2002. Hybrid recommender systems: survey and experiments. *User Modeling and User-Adapted Interaction*, 2002, 12(4), p. 331–370.
- Butoianu, V., Vidal, P., Verbert, K., Duval, E., et Broisin, J. 2010. User context and personalized learning: a federation of Contextualized Attention Metadata. *Journal of Universal Computer Science, John Wiley and Sons*, 16(16), p. 2252-2271.

- Butoianu, V., Catteau, O., Vidal, P., et Broisin, J. 2011. Un Système à Base de Traces pour la Recherche Personnalisée d'Objets Pédagogiques : le cas d'Ariadne Finder. Atelier "Personnalisation de l'apprentissage : quelles approches pour quels besoins ?", EIAH 2011, Mons, Belgique.
- Carbonell, J. 1970. AI in CAI: an artificial intelligence approach to computer -assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), p.190-202.
- Caron, P. A. 2006. Implantation de scénarios pédagogiques selon une approche orientée modèle : Une application de l'IDM aux EIAH, 1ères Rencontres Jeunes Chercheurs en EIAH, RJC-EIAH'2006, p. 1-8.
- Cedefop (Centre européen pour le développement de la formation professionnelle) 2008, Terminology of European Education and Training Policy, Bureau des publications officielles des Communautés européennes, Luxembourg.
- CETIS. 2006. The personal learning environments reference model project. JISC CETIS Wiki. Retrieved from <http://wiki.cetis.ac.uk/ple>
- Chalubert, A. 2001. Formalisation des données anatomique. Realisation d'un outil de génération de sites Web d'exploration anatomique. JOBIM Comité organisateur.
- Chen, K, et Sycara, L. 1998. Webmate: A personal agent for browsing and searching. In proceedings of the 2nd international conference on autonomous agent and multi agent systems, Minneapolis, p. 10-13.
- Cheung, R., Wan, C., et Cheng, C. 2010, An Ontology-based Framework for Personalized Adaptive Learning. In: Proceedings of the 9th International Conference on Web-based Learning, Shanghai, China, p. 52–61.
- Cheung R., Kazemian, H. B., Fong, S. 2011. An Adaptive Framework for Personalized E-Learning. In: Fong S. (eds) Networked Digital Technologies. NDT 2011, CCIS 136, p. 292–306. Springer-Verlag Berlin Heidelberg, 2011.
- Ciuciu, I., et Tang, Y. 2010. A personalized and collaborative elearning materials recommendation scenario using ontology-based data matching strategies. (Eds): OTM 2010 workshop, LNCS 6428, p.575-584. Springer-Verlag Berlin
- Claës, G. 1988. Contribution à l'application de l'intelligence artificielle pour l'enseignement assisté par ordinateur. Thèse de Doctorat, Université de Paris-Sud, Octobre 1988.

- Cohen, W.W. 1995. Fast Effective Rules Induction. IN PRIEDITIS, A. & RUSSELL? S. (Eds.) 12th international conference on machine learning. Tahoe City. CA, Morgan Kaufmann.
- Cohen, W.W. 1996. Learning rules that classify e_mail/ AAAI Spring Symposium on Machine Learning in information Access.
- Cooly, R., Tan, P.-N, et S rivastava, J. 1999. Websift: the web site information filter system. 1999 KDD workshp on Web mining. Sazn Diego. CA, Springer-Verlag.
- Corlett, D., Chan, T., Ting, J., Sharples, M., et Westmancott, O. 2005. Interactive Logbook: amobile Portfolio and Personal Developement Planning Tool. Paper presented at the meeting of mLearn. Cape Town, South Africa.
- Cosley, D., Lawrence, S. et Pennock, D. 2002. Referee : Anopen framework for practical testing of recommender systems using research index. In Proceedings of the 28th international conference on Very Large Data Bases, p. 46. VLDB Endowment.
- Cunningham, P., Bergmann, R., Schmitt, S., Traphoner, R., Breen, S., et Smyth, B. 2001 WEBSSELL: Intelligent sales assistants for the World Wide Web. KI – Zeitschrift fr Knstliche Intelligenz.
- Dagger, D., Wade, V., et Conlan, O. 2004. Developing Active Learning Experiences for Adaptive Personalised eLeraning. International Conference Adaptive Hypermedia 2004.
- Daneman, M., et Carpenter, P.A. 1980. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, p. 450-466.
- Daoud, M. 2009. Accès personnalisé à l'information: approche basé sur l'utilisation d'un profil utilisateur sémantique dérivé d'une ontologie du domaines à travers l'historique des sessions de recherches. Thèse de doctorat de l'université de Paul Sébastien de Toulouse, France.
- Daudigney, L., Geist, M., et Pietquin, O. 2011. Apprentissage par renforcement pour la personnalisation d'un logiciel d'enseignement des langues. Atelier "Personnalisation de l'apprentissage : quelles approches pour quels besoins ?", EIAH 2011, Mons, Belgique.
- Downes, S. 2006. Learning Networks and Connective Knowledge. Retrieved from <http://it.coe.uga.edu/itforum/paper92/paper92.html>.
- Drachsler, R., Drachsler, H., Pecceu, D. 2009. R.: ReMashed - An Usability Study of a Recommender System for Mash-Ups for Learning. 1st Workshop on Mashups for Learning at the International Conference on Interactive Computer Aided Learning, Villach, Austria.

Durao, F., Dolog, et P. 2010. Extending a Hybrid Tag-Based Recommender System with Personalization. In: Proc. 2010 ACM Symposium on Applied Computing, SAC 2010, Sierre, Switzerland, p. 1723–1727.

ePortConsortium. 2003. Electronic Portfolio White Paper, ePortConsortium.

Esslimani, I. 2010. Vers une approche comportementale de recommandation : apport de l'analyse des usages dans le processus de personnalisation. Thèse de doctorat en informatique. Université de Nancy 2, France.

Felder, R.M., et Silverman, L.K. 1988. Learning and Teaching Styles in Engineering Education. *Engineering Education*. 78(7), p. 674-681.

Galip, K., et Altun a. (2011), A Learner Model for Learning Object Based Personalized Learning Environments. E. García-Barriocanal et al. (Eds.): MTSR 2011, CCIS 240, p. 349–355, 2011. © Springer-Verlag Berlin Heidelberg 2011.

Gauch, S., Chaffe, J., et Pretschner, P. 2003. Ontology based user profiles for search and browsing. Volume Special issue on user modelling for web and hypermedia information retrieval.

Garlatti, S. et Prié, Y. 2004. Adaptation et personnalisation dans le web sémantique. Revue i3.org. Grand dictionnaire terminologique de l'Office québécois de la langue française: disponible en ligne http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp

Giacomini, E. P., Trigano, P., Alupoie S. 2006. Knowledge base for automatic generation of online IMS LD compliant course structures. *Educational Technology & Society*, 9(1), p. 158-175.

Giacomini, E. P. 2005. Une plate forme basée sur IMSLD pour la conception de cours en ligne dans le cadre de projet CEPIAH. Thèse de doctorat de l'université de technologie de Compiègne, France.

George, S. 2001. Apprentissage collaboratif à distance SPLACH : un environnement informatique support d'une pédagogie de projet, Thèse de doctorat en informatique université de Maine.

Goldberg, D. , Nichols, D. , Oki, B. et Terry, D. 1992. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), p. 61–70.

- Goldberg, K., Roeder, T., Gupta, D. et Perkins, C. 2001. Eigentaste : A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2), p.133–151.
- Good, N., Schafer, B. J. , Joseph, A., Konstan, A.I., Borchers, Badrul, S., Herlocker, J. 1999. Combining collaborative filtering with personal agents for better recommendations. *MAI '99/IAAI '99: Proceedings (JI the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, p. 439-446.
- Harrathi, R., et Calabretto, S. 2006. Un modèle de qualité de l'information. In G. Ritschard and C. Djeraba, editors, *EGC*, volume RNTI-E-6 of *Revue des Nouvelles Technologies de l'Information*, p. 299–304. Cepaduès-Editions, 2006.
- Herlocker, J. ; Konstan, J. ; Borchers, A., et Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*.
- Herlocker J. L., Konston J. A., Terveen L. G., et Riedl J. 2004. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, vol 22, p. 5-53.
- Hibou, M., et Py, D. 2006. Représentation des connaissances de l'apprenant. In M. Grandbastien & J.M. Labat (Eds.), *Environnements informatiques pour l'apprentissage humain*, p. 97-116. Paris: Lavoisier.
- IEEE 1484.12.1-2002. 2002. Draft Standard for Learning Object Metadata, http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
- IMC 1998. vCard, The Electronic Business Card. Internet Mail Consortium. <http://www.imc.org/pdi/>.
- IMS. IMS Learning Design Best Practice and Implementation Guide. Technical report, 2003-January-2003 2003.
- IMS. IMS Content Packaging Version 1.1.3 Final Specification. Technical report, June-12 2003.
- IMS Global Learning Consortium. 2002. IMS Reusable Definition of Competency or Educational Objective –Information Model, Version 1.0 Final Specification. IMS Global Learning Consortium.

- IMS Global Learning Consortium. 2004. IMS ePortfolio Best Practice and implementation Guide, Version 1.0 Final Specification. IMS Global Learning Consortium.
- Jäschke, R. ; Marinho, L. ; Hotho, A. ; Schmidt-Thieme, L., et Stumme, G. 2007. Tag recommendations in folksonomies. *Knowledge Discovery in Databases (PKDD 2007)*, p. 506–514.
- Keegan, D. 1996. *Foundations of distance education* (3rd ed.). London: Routledge.
- Kim, H. H. 2011. A Personalized Recommendation Method Using a Tagging Ontology for a Social E-Learning System. N.T. Nguyen, C.-G. Kim, and A. Janiak (Eds.): *ACIIDS 2011*, LNAI 6591, p. 357–366, 2011. © Springer-Verlag Berlin Heidelberg 2011.
- Kirshenman, U. et al. (2010). The demands of modern PLEs and the ROLE approach. *M. Wolpers et al.* (Eds.): *EC-TEL 2010, LNCS 6383*, p. 167–182.
- Koper, Rob. 2001. Modeling units of study from a pedagogical perspective: the pedagogical meta-model behind eml. first draft, version 2. Technical report, Educational Expertise Technology Centre, Open University of the Netherlands, june 2001.
- Kostadinov, D. 2003. La personnalisation de l'information, définition de modèle de profil utilisateur. rapport de dea. Master's thesis, Université de Versailles, France, 2003.
- Koutrika, G. , et Ioannidis, Y. E. 2004. Personalization of queries in database systems. *Intl. Conf. on Data Engineering (ICDE)*, IEEE Computer Society, p. 597–608.
- Koutrika, G., et Ioannidis, Y. E. 2005. Personalized queries under a generalized preference model. *Intl. Conf. on Data Engineering (ICDE)*, IEEE Computer Society, p. 841–852.
- Krulwich, B. et Burkey, C. 1996. Learning user information interests through extraction of semantically significant phrases. In *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access*. Stanford, CA.
- Johnson, M., et Liber, O. 2008. The personal learning environment and the human conditions: From theory to teaching practice. *Interactive Learning Environment*. 16(1), 3-15.
- Laforcade, P. 2004. « Meta modelisation pour la conception et la mise en œuvre de situations _problèmes cooperatives », thèse de doctorat l'université de PAU et des PAYS de l'ADOUR, 2004.

- Lam, S. et Riedl, J. 2004. Shilling recommender systems for fun and profit. In Proceedings of the 13th international conference on World Wide Web (WWW'04), p. 393–402, New York, NY, USA. ACM.
- Lameul, G. 2000. Former et échanger par les réseaux, ingénierie de formation a distance. IUFM de Bretagne, Intervention au séminaire inter-iufm de Nantes 6 et 7/4/00 (adaptation pour RESAFAD).
- Lasnier, F. 2000. Réussir la formation par compétences. Montréal: Guérin. Learning Technology Standards Committee of the IEEE.(2002). Draft Standard for Learning Object Metadata (LOM).[en ligne]. New York: IEEE.
- Lefevre, M., Jean-Daibia, S., Guin, N. 2009. Generation of pencil and paper exercises to personalize learners' work sequences: typology of exercises and meta-architecture for generators. E-Learn 2009, Vancouver, Canada, p. 2843-2848.
- Lefevre, M., Mille, A., Jean-Daibia, S., et Guin, N. 2009. A Meta-Model to Acquire Relevant Knowledge for Interactive Learning Environments Personalization. Adaptive 2009, Athènes, Grèce, p. 78-85.
- Lefevre, M. 2010. GEPPETO : une approche générique permettant d'adapter les activités des apprenants aux intentions pédagogiques de chaque enseignant. 3ème Rencontres Jeunes Chercheurs en Environnements Informatiques pour l'Apprentissage Humain (RJC EIAH 2010), Lyon, France, p. 35-40.
- Lefevre, M., Jean-Daibia, S., et Guin, N. 2011. Adapte, un logiciel pour aider l'enseignant à proposer des activités personnalisées à chacun de ses apprenants. Atelier "Personnalisation de l'apprentissage : quelles approches pour quels besoins?", EIAH 2011, Mons, Belgique.
- Lefevre, M., Jean-Daibia, S., et Guin, N. 2011. PERSUA2, un modèle pour unifier le processus de personnalisation des activités d'apprentissage. EIAH 2011, Mons, Belgique, p. 369-380.
- Leo, O., Howard, H. L., Robert, E. W. 2003. Ontologies for corporate web applications. AI Mag., 24, p. 49-62.
- Levy, T. 2004. The state and value of taxonomy standards. The Seybold Report July 21, 2004.
- Leung, C. W., Chan, S. C., et Chung, F. 2006. A collaborative filtering frame work based on fuzzy association rules and multiple-level similarity. Knowl. Inf. Syst., 10(3), p. 357–381.

- Lieberman, H. 1995. Letizia: an agent that assists web browsing. In IJCAI95 : Proceedings of International Joint Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence., p. 924–929.
- Lichtnow, D., Gasparini, I., Bouzeghoub, A. 2011. Recommendation of Learning Material through students' Collaboration and user Modelling in an Adaptive E-Learning Environment. Technology enhanced systems and tools, SCI, 350, p. 258-278.
- Linden, G., Smith, B. et York, J. 2003. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet computing, 7(1), p.76–80.
- Liu, F., et Yu, C. 2004. Personalized Web search for improving retrieval effectiveness. IEEE Transactions on knowledge Data Engineering, 16, 2004, p. 28-40.
- Lops, P., de Gemmis, M., et Semeraro, G. 2010. Content based recommender systems: State of the art and trends. In Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B., editors, Recommender Systems Handbook. Springer.
- Lorteau, C. , Peralta, S 2004 . « Schema2XMLs ». Rapport de projet, 2004.
- Lubensky, R. 2006. The present and future of personal learning environment (PLE). Retrieved from, <http://members.optusnet.com/rlubensky/2006/12/present-and-future-of-personal-learning.html>.
- Malone, T., Grant, K., Turbak, F., Brobst, S., et Cohen, M. 1987. Intelligent information-sharing systems. Communications of the ACM 30, 5, p. 390–402.
- Marty, J-C., et Mille, A. 2009. Analyse de traces et personnalisation des environnements informatiques pour l'apprentissage humain. Traité IC2 : Informatique et systèmes d'information, Hermès Sciences.
- Marot, J.C., et Darnige, A. 1996. La téléformation, Que sais-je ? (Presses Universitaires de France, Paris, 1996).
- Mbala, H. A. 2003. Analyse, conception, spécification et développement d'un système multi-agents pour le soutien des activités en formation à distance. Thèse de Doctorat en informatique, Université de Franche-Comté. 2003.
- MC Gowan, J. 2003. A multiple model approach to personalized information access. Master thesis in computer science, Faculty of science, University College Dublin, February 2003.

- McLaughlin, M. et Herlocker, J. 2004. A collaborative filtering algorithm and evaluation metric that accurately model the user experience. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'04), p. 329–336, New York, NY, USA. ACM
- McNee, S. ; Albert, I. ; Cosley, D. ; Gopalkrishnan, P. ; Lam,S. ; Rashid, A. ; Konstan, J. et Riedl, J. 2002. On the recommending of citations for research papers. In Proceedings of the 2002 ACM conference on Computer supported cooperative work, p. 125. ACM.
- Middleton, S. E., Alani, H., Shadbolt, N. R., et DeRoure, D. C. 2002. exploiting synergy between ontology and recommender system. In the semantic web workshop, World Wide Web Conference, 2002, p. 41-50.
- Middleton, S., Shadbolt, N. et Roure, D. D., 2004. Ontological user profiling in recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), p. 54–88.
- Mladenic, D. (1999). Text-learning and related intelligent agents: A survey. *IEEE Intelligent Systems*, 14(4):44–54.
- Montaner, M. 2003. Collaborative Recommender Agents Based on Case-Based Reasoning and Trust. Girona, Spain, Universitat de Girona, Girona.
- Moulet, L. 2006. ePortfolio, compétences et personnalisation d'apprentissage. Colloque CIRTA au congrès international de l'ACED et de l'AMTEC., 23-26 mai 2006, Montréal, Canada.
- Nicaud, J.F., et Vivet, M. 1988. Les Tuteurs Intelligents : réalisations et tendances de recherche, *Technique et Science Informatiques*, 7(1), p.21-45.
- Moulet, L. 2006. Revue de littérature de l'ePortfolio : Définitions, contenus et usages. Visant à l'intégration d'un ePortfolio dans le modèle de l'apprenant d'un système d'apprentissage en ligne. Note de recherche LICEF2006002, LICEF, Télé-université, Montréal.
- Oubahssi, L. 2005. Conception de plates-formes logicielles pour la formation à distance, présentant des propriétés d'adaptabilité à différentes catégories d'utilisateurs et d'interopérabilité avec d'autres environnements logiciels. Thèse de doctorat de l'Université René Descartes – Paris V, 2005.
- OCDE (2007), *Systèmes de certification : des passerelles pour apprendre à tout âge*, OCDE, Paris.

- Olivier, B., et Liber, O. 2001. Lifelong Learning: The need for portable Personal Learning Environments and supporting interoperability standards, Bristol, The JISC Centre for Educational Technology Interoperability Standards, Bolton Institute, Rapport de recherche, 2001.
- O'Mahony, M. ; Hurley, N. et Silvestre, G. 2006. Detecting noise in recommender system databases. In Proceedings of the 11th international conference on Intelligent user interfaces (IUI'06), p. 109–115, New York, NY, USA. ACM.
- Oxford, R. (2001). Language learning styles and strategies. In M. Celce- Murcia (Ed.), Teaching English as a second or foreign language (3rd ed.). Boston: Heinle & Heinle.
- Rashid, A.; Karypis, G., et Riedl, J. 2008. Learning preferences of new users in recommender systems: an information theoretic approach. SIGKDD Explor. Newsl., 10(2), p. 90–100.
- Page, M. P. A., et Norris, D. 1998. The primacy model: A new model of immediate serial recall. Psychological Review, 105, p. 761-781.
- Papagelis, M. ; Plexousakis, D. et Kutsuras, T. 2005. Alleviating the sparsity problem of collaborative filtering using trust inferences. In iTrust. Springer-Verlag Berlin Heidelberg.
- Paquette, G. 2002. Modélisation des connaissances et des compétences: Un langage graphique pour concevoir et apprendre, Presse de l'Université du Québec, Sainte-Foy, Québec.
- Pass, G. ; Chowdhury, A. et Torgeson, C. 2006. A picture of search. In Proceedings of the 1st international conference on Scalable information systems.
- Pazzani, M. J. 1999. A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Revue, 13(5-6), p. 393–408.
- Pazzani, M. et Billsus, D. 2007. The Adaptive Web, volume 4321/2007 de Lecture Notes in Computer Science, chapitre Content-Based Recommendation.
- Poirier, D. 2011. Des textes communautaires à la recommandation. Thèse de doctorat de l'université de Pierre et Marie Curie-Paris 6, France.
- Prévôt, P. 1992. Un tuteur intelligent pour la formation industrielle ; application à l'intégration d'un didacticiel Cimentier Thunder Bay. Proceeding of the Workshop de Thunder Bay, p. 20-39.

- Rafter, R. ; Bradley, K. et Smyth, B. 2000. Adaptive Hypermedia and Adaptive Web-Based Systems, volume 892/2000 de Computer Science, chapitre Automated Collaborative Filtering Applications for Online Recruitment Services, p. 363–368. Springer Berlin Heidelberg.
- Rasseneur , D. 2006. ‘Coffinet : Saafir : un environnement support à l’appropriation d’une formation à distance’. Thèse de doctorat de l’université du Maine.
- Resnick, P. ; Iacovou, N. ; Suchak, M. ; Bergstrom, P. Et Riedl, J. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In Proceedings of the ACM conference on computer-supported cooperative work.
- Resnick, P., et Varian, H. 1997. Recommender systems. Communications of ACM, 40(3), p. 56–58.
- Rawlings, A. , van Rosmalen, P. , Koper, R., Rodriguez-Artacho, M. et Lefrere, P. 2002. Survey of educational modelling languages (emls). Technical report, September 19st 2002.
- Salehi, M., kamalabadi, I. N., et Ghouschi, M.B.G. 2014. Personnalized recommandation of learning material using sequential patterns mining and attribute based collaborative filtering. Educ Inf Tech, 19, p. 713-735.
- Salton, G. et McGill, M. 1983. Introduction to modern information retrieval. McGraw-Hill, New York.
- Salton, G. 1989. Automatic text processing: the transformation, analysis, and retrieval of information by computer. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Self, J.A. 1988. Student models: what use are they?, In P. Ercoli and R. Lewis, Artificial Intelligence Tools in Education, Amsterdam, North-Holland.
- Severance, C., Hardin, J., et Whyte, A. 2008. The coming functionality mash-up in personal learning environments. Interactive Learning Environments, 16(1), p. 47-62.
- Shardanand, U., et Maes, P. 1995. 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, NewYork, NY, USA. ACM Press/Addison-Wesley Publishing Co.
- Sarwar, BadruJ, George Karypis, Joseph Konstan et T. John Riedl. 2000a. Analysis of recommendation algorithms for e-commerce. In Proceedings of the 2nd ACM conference on Electronic commerce: ACM.

- Sarwar, B. ; Karypis, G. ; Konstan, J. et Riedl, J. 2000b. Application of dimensionality reduction in recommender system - a case study. In ACM Web KDD 2000 Web Mining for E-Commerce Workshop.
- Sarwar, Badrul, George Karypis, Joseph Konstan et John Reid 1. 2001. «Item-based collaborative filtering recommendation algorithms». In Proceedings of the 10th international conference on World Wide Web: ACM.
- Sarwar, B., Karypis, G., et Konstan, J. 2002. Recommender system for large scale e-commerce: Scalable neighborhood formation using clustering. Fifth International Conference on Computer and Information Technology.
- Schafer, J. B., Konstan, J., et Riedi, J., 1999. Recommender systems in e-commerce. In EC '99: Proceedings of the 1st ACM conference on Electronic Commerce, p. 158–166, New York, NY, USA. ACM.
- Schwab, I., Pohl, W., et Koychev, I. 2000. Learning to recommend from positive evidence. AAAI 2000 Spring Symposium: Adaptive User Interface.
- Settouti, L.S., Prie, Y., Mille, A., et Marty, J.C. 2006. Systèmes à base de traces pour l'apprentissage humain. CEN Workshop Agreement (CWA 15454).
- Settouti, L., Guin, N., Luengo, V., Mille, A. 2010. A Trace-Based Learner Modelling Framework for Technology-Enhanced Learning Systems. The 10th IEEE International Conference on Advanced Learning Technologies (ICALT 2010), Sousse, Tunisie, p. 73-77.
- Settouti, L., Guin, N., Luengo, V., Mille, A. (2011). Adaptable and Reusable Query Patterns for Trace-Based Learner Modelling. Sixth European Conference on Technology Enhanced Learning: Towards Ubiquitous Learning (EC-TEL 2011), Palerme, Italie.
- Sieg, A., Mobasher, B., Burke, R., Prabu, G., et Lytinen, S. 2004. Using concept hierarchies to enhance user queries in web-based information retrieval. In The IASTED International Conference on Artificial Intelligence and Applications. Innsbruck, Austria, 2004.
- Skinner, B.F. 1954. The Science of Learning and the Art of teaching. Harvard Education Revue, 24, p. 86-97.
- Sollenborn, M. et Funk, P. 2002. Category-based filtering and user stereotype cases to reduce the latency problem in recommender systems. In Proceedings of the 6th European Conference on Advances in Case-Based Reasoning (ECCBR'02), p. 395–420, London, UK.Springer-Verlag.

- Sun, Y., Li, H., Councill, I.G, Huang, J., Lee, W-C., et Giles, C. L. 2008. Personalized ranking for digital libraries based on log analysis. Intl. Workshop on Web Information and Data Management (WIDM), ACM, p. 133–140.
- Sutton, R.S., et Barto, A.G. 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
- Svensson, M. ; Höök, K. et Cöster, R. 2005. Designing and evaluating kalas : A social navigation system for food recipes. ACM Transactions on Computer-Human Interactions (TOCHI), 12(3), p. 374–400.
- Szilagyi, I., Balog-Crisan, R., et Roxin, I. 2010. Kernel for a Semantic Learning Platform with adapted suggestions. 10th IEE International conference on Advanced Learning Technologies (ICALT 2010), Sousse, Tunisie, p. 400-402.
- Szilagyi, I., Greffier, F., et Domenget, J-C. 2011. Apprentissage personnalisé via le web sémantique. Atelier "Personnalisation de l'apprentissage : quelles approches pour quels besoins ?", EIAH 2011, Mons, Belgique.
- Tamine, L., Boughanem, M., et Daoud, M. 2009. Evaluation of contextual information retrieval : overview of issues and research. Knowl Inf Syst (Kais), in press.
- Taraghi, B., Ebner, M., Till, G., et M'uhlburger, M. 2009. Personal Learning Environment – a Conceptual Study. In: International Conference on Interactive Computer Aided Learning, p. 25–30.
- Tchounikine, P. 2002. Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain. Revue I3 information-interaction-intelligence 2(1), www.revue-i3.org
- Verpoorten, D., Glahn, C., Kravcik, M., Ternier, S., Specht, M. 2009. Personalisation of Learning in Virtual Learning Environments. Lecture Notes in Computer Science, Learning in the Synergy of Multiple Disciplines, Vol. 5794, p. 52-66. Web Based Enterprise Management. (1999). <http://www.dmtf.org/standards/wbem/> (consulté le 12 septembre 2011).
- Utecht, J. 2006. Creating a learning community. TechLEARNING. Retrieved from http://www.techlearning.com/blog/2006/10/creating_a_learning_community.php.
- Vladoiu M., Constantinescu Z., et Moise G. 2013. QORECT – A Case-Based Framework for Quality-Based Recommending Open Courseware and Open Educational Resources, C.

- Bădică, N.T. Nguyen, et M. Brezovan (eds.): ICCCI 2013, LNAI 8083, p. 681–690, © Springer-Verlag Berlin Heidelberg 2013.
- Water, J. 2008. Unleashing the power of Web 2.0. *Capmus Technology*, 21(10), p. 44-49.
- Wen, J., Lao, N., et Ma, W. Y. 2004. Probabilistic model for contextual retrieval. *Proceedings of the 27th annual international ACM SIGIR Conference on Research and development in Information retrieval*, August 2004, p. 57-63.
- Weng, L. T., Xu, Y., Li, Y., et Nayak, R. 2008. Exploiting Item Taxonomy for Solving Cold-Start Problem in Recommendation Making. *ICTAI* (2), p. 113-120.
- Werquin, P. 2007. Terms, Concepts and Models for Analysing the Value of Recognition Programmes. étude préparée pour l'activité de l'OCDE sur la reconnaissance des acquis des apprentissages non formels et informels, www.oecd.org/dataoecd/33/58/41834711.pdf.
- Werquin, P. 2010. Reconnaître l'apprentissage non formel et informel, résultats, politiques et pratiques. OCDE, Paris, France.
- Wild, F., Ullmann, T., Scott, P., Rebedea, T., et Hoisl, B. 2011. Applicability of the Technology Acceptance Model for widget-based Personal Learning Environments", *Proceedings of the 1st Workshop on Exploring Fitness and Evolvability of Personal Learning Environments*, La Clusaz.
- Yang, T.C., Hwang, G.J., Chiang, T.H.C., et Yang1, J.H.. 2013. A Multi-dimensional Personalization Approach to Developing Adaptive Learning Systems. A. Holzinger and G. Pasi (Eds.): *HCI-KDD 2013, LNCS 7947*, p. 326–333.
- Zaier, Z. 2010. Modèle multi-agents pour le filtrage collaboratif de l'information. Thèse de doctorat de l'université de Québec, Montréal, Canada.
- Ziegler, C.-N., Lausen, G., et Schmidt-Thieme, L. 2004. Taxonomy-driven computation of product recommendations. In *Proceedings of the 2004 ACM CIKM Conference on Information and Knowledge Management* Washington, D.C., USA, ACM Press, p. 406–415.
- Ziegler, C. ; McNee, S. ; Konstan, J. et Lausen, G. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web (WWW'05)*, p. 22–32, New York, NY, USA. ACM.

Annexe A

XML (eXtensible Markup Language)

1. Introduction :

XML (eXtensible Markup Language), signifie « langage de balisage extensible », développé par W3C, ce langage est devenu rapidement un standard de description et d'échange de données sur le Web. XML est un langage balisé comme l'HTML. La différence entre les deux langages est que l'HTML utilise des balises prédéfinies (qui conditionneront la manière dont le texte sera affiché dans le navigateur), alors que dans XML les balises sont utilisées seulement pour délimiter les éléments et l'interprétation des données est laissée à l'application qui les lit (Chalubert, 2001). Il nous permet de décrire le processus d'apprentissage de façon formelle.

Les avantages de XML :

- Il est caractérisé par la définition de ses propres balises et ses propres attributs. Il possède un nombre élevé de balises contrairement à l'HTML, d'où se réside sa flexibilité.
- Un document XML peut être validé en utilisant le DTD ou des schémas qui permettent de définir sa structure.
- La séparation entre les informations et le traitement de la mise en forme.

2. Structure d'un document XML:

Un document XML peut être structuré en trois éléments : Le prologue, Un arbre d'éléments et les commentaires.

2.1. Prologue :

2.1.1 Déclaration XML :

C'est la ligne d'introduction d'un document XML :

```
<?xml version="1.0" encoding='ISO-8859-1' standalone='yes' ?>
```

Cette déclaration permet de fournir trois informations: la version du XML utilisée dans le document, le jeu de codage du caractère utilisé (*encoding*) et dépendance du document par rapport à une déclaration de type de document (*standalone*). Si *standalone* a la valeur 'yes', le

processeur de l'application n'attend aucune déclaration de type de document extérieure au document. Sinon, le processeur attend une référence de déclaration de type de document. La valeur par défaut est 'no'.

2.1.2 Instructions de traitement :

Les instructions de traitement ne font partie du document (facultatif), elles sont interprétées par l'application afin de traiter le document XML. Un exemple de ces instructions est l'instruction qui sert à déclarer une feuille de style.

2.1.3 Déclaration de type de document (DTD) :

La DTD (Document Type Definition) sert à définir la structure d'un document XML. Elle peut être vue comme une grammaire algébrique dont les non terminaux sont les éléments balises et les terminaux sont les différents types de données présents dans l'arborescence du document. Un document XML est dit valide s'il est conforme à une DTD donnée.

2.2 Les commentaires :

En XML, les commentaires se déclarent de la même façon qu'en HTML. Ils sont encadrés par les marques de début et de fin de commentaire :

```
<!-- Ceci est un commentaire -->
```

2.3 L'arbre d'éléments :

Un document XML peut être structuré en arborescence d'éléments. Cette arborescence est constituée d'un élément racine (unique), des branches et des feuilles. Chaque élément peut contenir un ou plusieurs attributs. Un attribut est composé d'un nom et d'une valeur (« nom = 'valeur' »).

Un document XML peut contenir des entités. Elles peuvent être utilisées aussi bien dans la DTD que dans le document XML. Certains caractères ayant un sens précis en XML, il est nécessaire de leur trouver un remplaçant lorsque l'on a besoin de les insérer dans un document. On a recours dans ce cas à des entités prédéfinies. Ces entités sont :

Caractère	Entité
&	&
<	<
>	>
"	"
'	'

Table A.1. Liste des entités prédéfinies

3. Schéma XML :

Il est conçu pour pallier les inconvénients du DTD, c'est un langage qui permet de définir le format d'un document XML. Un schéma XML est lui-même un fichier XML (XML Schema Description en anglais, ou fichier XSD). La connaissance de la structure d'un document XML permet notamment de vérifier la validité de ce document.

4. Les feuilles de styles :

Pour qu'un fichier XML soit affichable il faut lui ajouter une « feuille de style » servant à définir une présentation physique (marges, polices et tailles de caractères, couleurs, etc.) pour celui-ci. Les feuilles de styles CSS (Cascading style Sheets) permettent de contrôler la mise en page des différents éléments qui composent un document HTML ou XML. XSL (eXtended Style Language), comme CSS, est un complément « feuille de style » de XML.

5. L'API JDOM :

Java Document Object Model (JDOM) est une API écrite en Java afin de manipuler des fichiers XML. Cette API passe outre les limites de DOM dont elle s'inspire. Il existe également une API pour le traitement des données XML nommée : SAX, qui est très limitée quand il s'agit d'effectuer des modifications au sein du document. L'intérêt principal de JDOM est sa simplicité.

Annexe B

UML (Unified Modeling Language)

1. Introduction :

L'UML est utilisé pour visualiser, spécifier, construire et documenter les artefacts d'un système à fortes composantes logicielles. Ce langage est issue de la fusion de plusieurs d'autre langage de modélisation orientés objets. Ce formalisme modélise le système en utilisant neuf diagrammes selon deux modes de représentation : le premier est dédié pour la structure du système (diagramme de cas d'utilisation, de classe, etc.) le deuxième concerne sa dynamique de fonctionnement (diagramme d'états transitions, de séquence, d'activité, etc.).

2. Les diagrammes de l'UML (Laforcade, 2004) :

Nous présentons par la suite seulement les deux diagrammes qu'on a utilisés : diagramme de classe et diagramme d'activité.

2.1. Diagramme de classe :

Ce diagramme sert à modéliser la structure statique du système en utilisant des éléments tels que : classe, paquetage,...Il montre cette structure comme un ensemble de classes et de relations entre ces classes. Ce diagramme est utilisé pour représenter le modèle d'information de l'IMSLD.

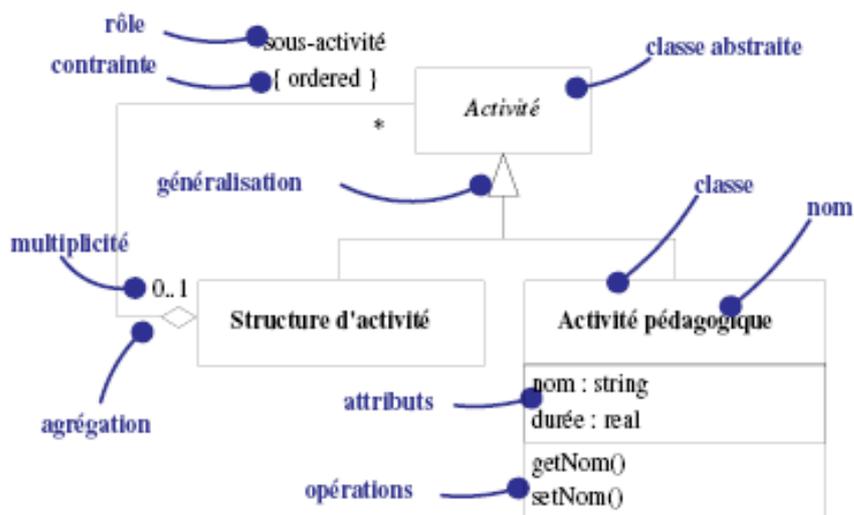


Figure A.1: Un exemple simplifié et annoté de diagramme de classe.

2.2. Diagramme d'activité :

Ce diagramme est utilisé pour représenter le comportement du système. Il permet également de décrire les flux d'informations (workflows) comme les flux de données (data flow). Le diagramme d'activité sert à spécifier les scénarios de nos situations d'apprentissage collaboratif.

La structure de ces scénarios ou organisation des activités est décrite en termes de *act* et *play* dans IMS-LD.

Annexe C

The University's Taxonomy

- t₁: GUI<General
 - GUI<Interaction styles<Menu
 - GUI<Interaction styles<Command langages
- t₂: KM<General
 - KM< Knowledge types<Explicit
 - KM< Knowledge types<Implicit
- t₃: KM< Knowledge creation process< General
 - KM< Knowledge creation process<Nonaka's model
- t₄: GUI<General
 - GUI< Interface types< Command Interfaces
 - GUI< Interface types< Graphique Interfaces
- t₅: KM< KM Models< measurement's Models
 - KM< KM Models<evaluation's Models
- t₆: DB<General
 - DB<Conceptual Model
- t₇: GUI<General
 - GUI<Interface creation<tools<API
 - GUI<Interface creation< tools<Interface generator
- t₈: KM<General<nul
 - KM<Company strategies<Encoding
 - KM<Company strategies<Personnalization
- t₉: OP<General
 - OP<Processor management<nul
 - OP<I/O management<I/O type
- t₁₀: OP<General
 - OP<File managment
- t₁₁: LT<General
 - LT<Regular languages
 - LT<Regular languages<Regular expression
- t₁₂: LT<General

LT<Turing machine

- t₁₃: OP<General

OP<Central memory management<Memory protection

- t₁₄: DB<General

DB<DB Managment System

DB<SQL

- t₁₅: DB<General

DB<DB update