

## Génération automatique de cours personnalisés dans un environnement d'e-learning

Henda Chorfi , Mohamed Jemni

Unité de recherche UTIC  
Ecole Supérieure des Sciences et Techniques de Tunis  
5, Av. Taha Hussein, B.P. 56, Bab Mnara, Tunis  
TUNISIA  
[Henda.chorfi@esstt.rnu.tn](mailto:Henda.chorfi@esstt.rnu.tn), [Mohamed.jemni@fst.rnu.tn](mailto:Mohamed.jemni@fst.rnu.tn)

**RESUME.** Les recherches menées dans le domaine de l'enseignement à distance sont nombreuses et variées. Le principal objectif de ces recherches est de donner un environnement de travail efficace aussi bien pour l'apprenant que pour le concepteur de cours. Dans ce contexte, nous présentons PERSO, un système qui modélise et construit dynamiquement des cours personnalisés pour des étudiants dans un environnement d'enseignement à distance en tenant compte de leurs pré-requis et de leurs préférences en termes de présentation et de type de média. PERSO est basé sur l'approche du Raisonnement à Base de Cas (CBR) pour déterminer la solution (un cours personnalisé). Ce papier est principalement focalisé sur le module CBR : nous décrivons l'indexation des cas, le calcul de la similarité entre deux cas et la méthode d'adaptation des solutions des problèmes précédents pour résoudre de nouveaux problèmes.

**ABSTRACT.** Many works have been done in the e-learning research domain. Their main goal is to offer to the actors an efficient work environment. In this context, we present PERSO, a system that dynamically builds personalized courses for students in an e-learning environment considering their backgrounds and their media preferences. PERSO employs Case-Based Reasoning (CBR) to determine the solution (a personalized course). This paper focuses on the CBR system, describing in detail how to index a case, how to calculate the similarity between cases, and finally how to adapt the solutions of previous problems to solve a new one.

**MOTS-CLÉS :** télé-enseignement, personnalisation, modélisation, LSA, CBR

**KEYWORDS:** e-learning, personalization, LSA, CBR

---

## 1. Introduction

Internet a révolutionné différents domaines et en particulier le domaine de l'enseignement avec l'avènement d'un nouveau mode d'apprentissage : l'e-learning [2]. En effet et dès la naissance du Web, on a vu apparaître les premiers usages pédagogiques et de nombreux enseignants se sont mis à créer des pages Web pour leurs étudiants. Assez rapidement, on voyait apparaître sur le marché des centaines de plates formes de formation à distance issues de milieux universitaires ou commerciaux. La plupart de ces systèmes sont centrés, du point de vue de leurs fonctionnalités, sur la création et la diffusion de documents et sur l'intégration d'outils de communication (messagerie, forums, chats...). Cependant, la formation à distance soulève de nombreux problèmes que nous résumons dans les trois points suivants : la modélisation de l'apprenant (ses pré-requis, ses acquis, ses préférences ...), l'adaptation dynamique de l'environnement à l'apprenant (la personnalisation du contenu et du contenant) et la conception des ressources numériques à enseigner (le découpage, la scénarisation, la réutilisation...). Néanmoins, plusieurs travaux apportent des réponses, plus ou moins complètes, à ces questions en s'appuyant sur des expérimentations réelles.

C'est dans ce cadre que s'intègrent nos travaux de recherche. Et dans ce papier, nous présentons PERSO, un système qui modélise les pré-requis et les préférences de l'apprenant à travers un questionnaire dynamique à réponses ouvertes afin de lui proposer un parcours personnalisé. La détermination de ce parcours s'appuie sur une modélisation du contenu et une réutilisation de parcours précédents basée sur le Raisonnement à Base de Cas.

---

## 2. Description de PERSO

### 2.1. Contexte de travail

Notre idée de concevoir le système PERSO est, d'une part, motivée par le fait de donner des réponses aux problèmes soulevés dans l'introduction. Et d'autre part, étayée par les résultats de l'expérience que nous avons menée à l'Ecole Supérieure des Sciences et Techniques de Tunis (ESSTT), où nous avons dispensé un cours de bureautique à distance au moyen d'une plate forme tunisienne [3] à des étudiants en 1ère année Maîtrise Informatique. Au bout de cette expérience et au vue des études que nous avons effectuées (questionnaires auprès des utilisateurs, analyses des données de la plate-forme...), nous résumons nos principaux questionnements comme suit :

- Premièrement, pourquoi proposer le même contenu du cours pour des étudiants qui peuvent avoir des pré-requis différents ?

- Deuxièmement, pourquoi proposer la même apparence du cours à un groupe hétérogène d'étudiants, ayant des préférences différentes en terme de présentation ?

## 2.2. Architecture de PERSO

PERSO, la solution que nous proposons pour répondre à nos questionnements, est un système de génération de cours personnalisés dans un environnement d'enseignement à distance. Il doit :

- fournir une structure de représentation des connaissances à enseigner
- déterminer le niveau de connaissance (pré-requis) de l'apprenant ainsi que ses préférences en termes de type de média.
- fournir des cours multimédia personnalisés à chaque apprenant.

Comme le montre la figure 1, l'architecture de notre système se rapproche globalement de l'architecture des hypermédias adaptatifs dynamiques c'est à dire qu'il est fondé sur les quatre composantes que sont le modèle du domaine, le modèle de l'apprenant, le générateur de cours et une base de données multimédia [4]. Cependant, l'originalité de PERSO réside dans ses deux nouvelles composantes qui sont l'analyseur et le module CBR.

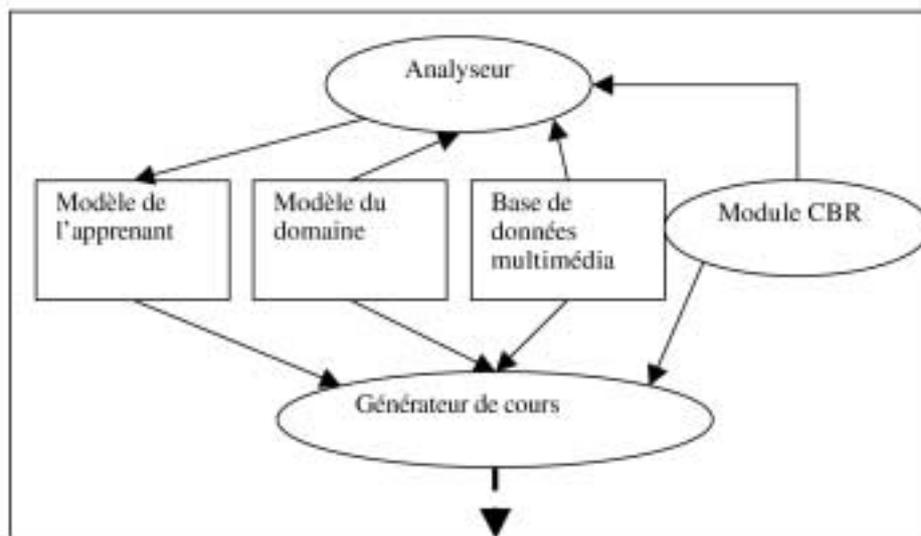


Figure 1 : Architecture de PERSO

Dans ce qui suit, nous allons étudier les principales composantes de PERSO. Au cours de cette présentation, certaines composantes de PERSO vont être illustrées par des exemples qui utilisent un cours de WORD.

---

### 3. Description des composantes

#### 3.1. Le module Analyseur

Quand l'étudiant se connecte au système pour suivre pour la première fois un nouveau cours, il est soumis à un questionnaire pour déterminer ses pré-requis et ses préférences multimédia. En ce qui concerne ces dernières, l'apprenant est appelé à ordonner par ordre de préférence différents types de média. Nous avons retenu quatre type : texte/image, son, vidéo et illustration/animation.

La détermination du pré requis de l'apprenant se procède avec un questionnaire par catégorisation [1],[5]. Nous retenons trois catégories d'apprenant : débutant, intermédiaire et avancé. Le questionnaire s'effectue en deux étapes : d'abord, il détermine la catégorie de l'apprenant. Ensuite, il détermine ses pré-requis dans cette catégorie.

Pour la détermination de la catégorie, le système aura à poser une ou deux questions à l'apprenant, portant sur les points d'entrée du réseau sémantique du cours (voir le paragraphe suivant).

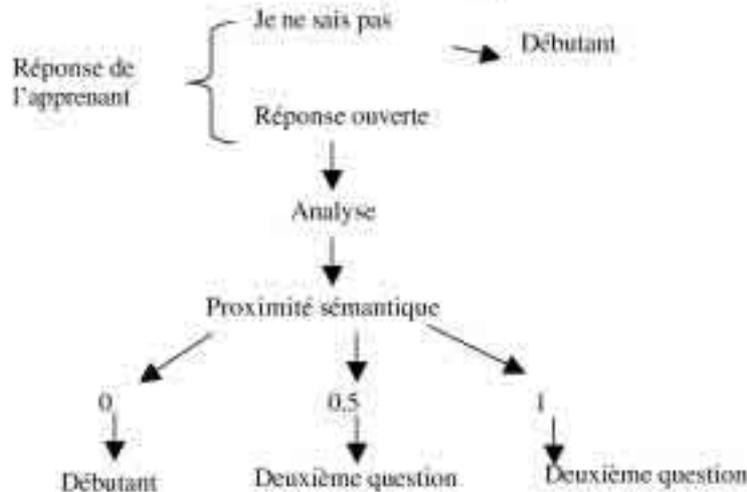
Pour modéliser les connaissances de l'apprenant dans sa catégorie, le questionnaire utilise une approche ouverte et qui consiste à laisser l'utilisateur s'exprimer librement pour répondre à une question à l'aide d'une saisie manuelle. Cette réponse est ensuite représentée dans la machine pour la comparer avec « la bonne réponse » (celle du professeur sauvegardée au préalable au système). La représentation et la comparaison des deux réponses sont données par le modèle de *l'Analyse de la Sémantique Latente (LSA)* [6]. LSA est un modèle statistique entièrement automatisé, basé sur l'analyse factorielle, qui permet d'extraire automatiquement des proximités sémantiques entre mots à partir d'une réduction de matrice d'occurrences dans un espace de très grande dimension.

La proximité sémantique entre les deux réponses est donnée par le cosinus des deux vecteurs représentant les réponses dans l'espace sémantique. Nous avons fixé trois intervalles de valeurs :  $[-1; 0.5]$  : la notion est très peu ou pas du tout maîtrisée,  $[0.5; 0.7]$  : la notion est assez maîtrisée et  $[0.7; 1]$  : la notion est parfaitement maîtrisée. Les trois intervalles sont représentés respectivement par les valeurs suivantes : 0, 0.5 et 1.



Dans l'exemple ci-dessous, nous retraçons une étape de la détermination de la catégorie.

Donnez les différentes instructions pour la copie d'un bloc ?



### 3.2 Le modèle du domaine

Dans ce module nous représentons les cours à enseigner sous forme de réseau sémantique : un graphe où les nœuds sont des notions et les arcs sont des relations entre les nœuds. Une notion peut être composée de plusieurs briques élémentaires : une introduction, un théorème, une formule, une illustration, une définition, une conclusion un exemple, une explication, une preuve et un algorithme (étapes de résolution). Dans cette liste, nous nous sommes inspirés des travaux de METADYNE [4] tout en essayant de la rendre la plus complète possible. Les notions qui sont introduites dans le cours sont reliées par différents types de relation. Dans notre réseau sémantique, nous utilisons quatre types de relations : «Composition», «Equivalence», «Pré requis» et «Analogie».

Dans le réseau sémantique des notions, nous définissons trois points d'entrée correspondant chacun à une catégorie (débutant, intermédiaire, avancé). Un point d'entrée est la notion la plus avancée dans une catégorie. La figure suivante schématise ces trois points d'entrée.



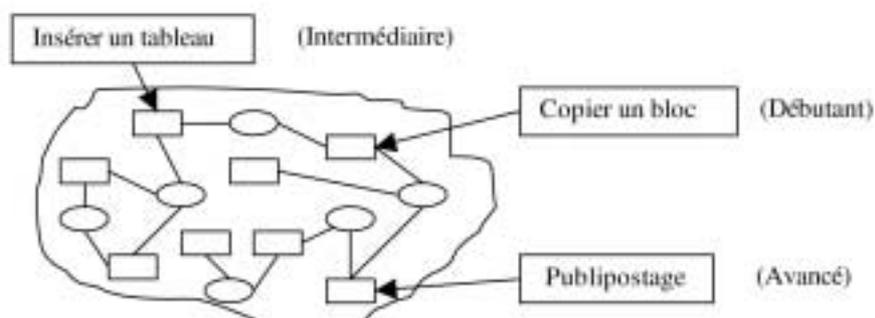


Figure 2. Les points d'entrée du réseau sémantique des notions

### 3.3. Le modèle de l'apprenant

La modélisation de l'apprenant est basée sur une approche ouverte consistant à acquérir les informations d'interactions entre l'apprenant et le système par le biais d'un questionnaire. Deux types de connaissances sont modélisés : les préférences de l'apprenant en terme de type de média et ses pré-requis. Ils sont représentés sous forme d'entités-valeurs. Les valeurs pour les préférences en terme de type de média sont le classement que fournit l'apprenant pour chaque type. Par contre, les valeurs pour les pré-requis sont fournis par le système et concernent les proximités sémantiques calculées par LSA suite au questionnaire. Dans ce qui suit un exemple :

*Préférences en terme de type de média :* {(textefimage, 1), (son, 3), (vidéo, 2), (illustration/animation, 4)}

*Pré-requis :* {(Démarrer word, 1), (Créer nouveau document, 1), (Selection bloc, 0), (Mise en forme, 0), (Alignement paragraphe, 0), (Imprimer un document, 0), (Insérer un tableau, 0), (options impressions, 0), (entête/pied de page, 0), (trier un tableau, 0), (Convertir un tableau, 0), (Dessiner un tableau, 0), (Insertion objet, 0), (Dessiner forme simple, 0)}

### 3.4. Le module CBR

#### 3.4.1. Introduction

Les fonctionnalités de ce module sont basées sur l'approche du Raisonnement à Base de Cas (CBR). Nous rappelons que la principale hypothèse derrière cette approche est la suivante « vous pouvez réutiliser la solution d'un problème similaire pour résoudre votre problème actuel » [7].

Les différentes étapes du système CBR sont traitées dans les paragraphes suivants.

### 3.4.2. Initialisation et indexation

Un cas est composé d'un numéro de cas, de la catégorie de l'apprenant, son identification, son profil (les pré-requis et les préférences en type de média de présentation) et de la solution. Le profil de l'apprenant permet d'indexer le cas.

La solution consiste dans l'agencement des briques élémentaires dans un parcours donné.

La base de cas est initialisée par des cas initiaux typiques dans chaque catégorie.

### 3.4.3. Similarité

Une fois un nouveau cas est construit (sans la solution), nous calculons sa similarité avec les autres cas de sa catégorie dans la base. La fonction de similarité que nous allons appliquer est la distance euclidienne définie comme suit:

$$\text{soit } u = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ et } v = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \|u - v\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Cette fonction est appliquée sur les attributs du profil cognitif du cas. Nous considérons que deux cas sont similaires si leur similarité est inférieure à un certain seuil.

### 3.4.4. Adaptation

Les résultats de la mesure de similarité permettent au système de décider de l'adaptation à appliquer. PERSO utilise l'adaptation structurelle [7].

PERSO construit la solution d'un nouveau cas en retenant la solution du problème le plus similaire et en lui appliquant un ensemble de règles. Dans ce qui suit, nous citons quelques unes de ces règles.

*Exemples de règles :*

1) Si  $similar.attribut(i) = 0$  et  $nouveau.attribut(i) = 1$  alors supprime  $attribut(i)$  à  $similar.solution$

2) Si  $similar.attribut(i) = 1$  et  $nouveau.attribut(i) = 0$  alors ajoute  $attribut(i)$  de la  $similar.solution$

Notons que  $similar.attribut(i)$  est un attribut du cas similaire retenu et  $nouveau.attribut(i)$  est l'attribut correspondant du cas nouvellement construit et pour lequel nous cherchons à adapter la solution.

---

## 4. Conclusion

Le système PERSO que nous avons présenté dans ce papier est centré sur la génération de cours personnalisés. La personnalisation se fait en amont de la session de formation pour éviter le problème de la surcharge cognitive. Nous jugeons que ceci est nécessaire pour permettre un meilleur suivi de l'apprenant : quand l'apprenant dispose d'un cours où toutes les notions lui sont nouvelles il va se focaliser sur l'apprentissage et sa navigation devient un meilleur indicateur. Pour cela, PERSO est basé sur une approche ouverte pour modéliser l'apprenant et sur le Raisonnement à Base de Cas pour réutiliser les parcours personnalisés déjà proposés.

---

## Bibliographie

- [1] Aïmeur, E., Brassard, G., Dufort, H., Gambs, S., 2002 "CLARISSE: A Machine Learning Tool to Initialise Student Models", *Sixth International Conference on Intelligent Tutoring Systems*, pp. 718-728, France, June 2002.
- [2] Brusilovsky P., Eklund J., Schwartz E., 1998 "Web-Based Education for All: A Tool for Developing Adaptive Courseware", *Computer Networks and ISDN Systems*, 30, 1-7, pp. 291-300.
- [3] Chorfi H. & Jemni M., 2002 "Evaluation and Perspectives of an Innovative Tunisian e-Learning Experimentation ", *International Conference Advances in Infrastructure for e-business, e-education, e-science and e-medicine on the Internet, l'Aquila, Italy, July 29 – August. 04*.
- [4] Nicolas Delestre, Stéphane Frénot, Stéphane Mottelet, Michel Vayssade, 2002 « Distributed PolyTeXML, une nouvelle plateforme de partage d'items didactiques » TICE'2002, Lyon, p149-159.
- [5] Kay, J., 2000 "Stereotypes, Students Models and Scrutability", *Intelligent Tutoring Systems ITS 2000*, Montréal, Berlin: Springer-Verlag, pp. 19-30.
- [6] Landauer, T.K., Foltz, P., & Laham, D., 1998 "Introduction to Latent Semantic Analysis". *Discourse Processes*, 25, pp. 259-284.
- [7] Wilke, W., Bergmann, R., 1998 Techniques and Knowledge Used for Adaptation During Case-Based Problem Solving, *Proceedings of IEA-98-AIE*, Springer-Verlag, Berlin Heidelberg New York.