

# Conception d'un dispositif d'accompagnement pour une plateforme de formation à distance en arabe

Hassina Seridi \*, Mokhtar Sellami\*, Issam Chibouni, Adlen Zenoune

\* LRI Laboratoire, Université de Badji Mokhtar

BP 12

23000 Annaba, Algérie.

{Seridi, Sari, Sellami}@lri-annaba.net

## Résumé

Cet article présente un modèle d'organisation du tutorat pour guider la conception de dispositifs d'accompagnement. Cette modélisation permet de déterminer l'activité de tutorat à mettre en œuvre en fonction d'une situation d'apprentissage donnée. Il est alors possible de spécifier et choisir les outils à mettre à disposition du tuteur pour percevoir la situation mais aussi pour interagir avec les apprenants au cours d'une session d'apprentissage.

Plusieurs travaux sont en cours pour élaborer ces outils pour les langues étrangères mais aucun projet n'est initié pour la langue arabe qui possède ses propres spécificités dans les problèmes de classification de textes. Une méthode de conception itérative et participative a permis d'élaborer un tel modèle, des outils et des interfaces de tutorat pour le dispositif de formation pour la langue Arabe.

## Mots clés

Tutorat, Accompagnement, Conception de dispositif informatique d'accompagnement, Ingénierie d'une formation à distance, Algorithme de classification de texte

## 1 Introduction

Les scénarios d'enseignement basé Web évoquent trois acteurs : auteurs, tuteurs et apprenants. Les auteurs sont généralement experts dans un domaine particulier, possèdent des connaissances appréciables sur le domaine, préparent des cours et les publient sur le système. Les apprenants accèdent aux cours publiés pour acquérir de nouvelles connaissances. Les tuteurs supervisent les apprenants durant leurs sessions d'apprentissage et répondent à leurs questions à propos du matériel d'enseignement présenté.

Le tutorat à distance entre dans le cadre des activités de formation en ligne qui permettent d'aider et d'accompagner les apprenants dans leurs démarches d'apprentissage. Ces activités de tutorat visent à atteindre un double objectif de formation et de développement de l'autonomie de l'apprenant [1].

En formation en ligne, apprenants et tuteurs sont liés par une activité de construction de connaissances qui doit être supportée informatiquement afin de pallier les contraintes liées à la distance [2]. Les interactions ont généralement lieu en différé (mode

asynchrone). Dans ce contexte de formation à distance, le tutorat se définit comme la mise en place d'une relation de modélisation construite pour l'apprenant et avec l'apprenant, ce qui crée une réelle problématique en regard aux possibilités d'interaction et de communication des dispositifs médiatisés.

A partir de ces constats, la question de la conception de dispositifs d'accompagnement des acteurs (tuteur et apprenants) se pose. Nous entendons, par accompagnement des apprenants, une mise à disposition d'outils leur permettant d'interagir avec les autres acteurs de l'activité, une perception de l'environnement d'apprentissage et leur évolution dans la formation, et un guidage dans leurs activités. L'accompagnement du tuteur recouvre les outils lui permettant de percevoir le déroulement de l'activité au cours d'une session mais aussi d'intervenir " juste à temps " et de manière adaptée auprès d'un apprenant ou d'un groupe d'apprenants. Dans cet article, nous allons nous focaliser sur un outil d'accompagnement des apprenants qui va leur permettre d'avoir une assistance dans les questions posées au tuteur en langue arabe. Cet outil assistera fortement le tuteur dans son rôle de tutorat dans un processus d'enseignement en lui offrant un environnement flexible lui permettant de personnaliser les réponses aux questions en langue arabe des apprenants automatiquement.

Nous commençons par définir la fonction de tutorat, nous énumérons nos objectifs, nous présentons le modèle de tutorat que nous avons proposé et nous détaillerons l'algorithme utilisé pour la classification du texte des questions des apprenants.

## **2 Définition du tutorat**

Au regard de la diversité des travaux, la question du tutorat est abordée selon différents points de vue. Le tutorat a une dimension individuelle (au sens relation individuelle entre le tuteur et l'apprenant) ou collective (au sens où le tuteur intervient auprès de plusieurs étudiants).

Il existe donc une diversité du tutorat du point de vue de sa forme, les acteurs jouant le rôle du tuteur dans le cadre d'une activité d'apprentissage quelle qu'elle soit, pouvant être aussi bien un tuteur humain (un enseignant ou un apprenant) qu'un dispositif informatique. Cette diversité suppose en outre une multiplicité des rôles et des tâches du tuteur ce qui implique la nécessité de se questionner lors de la mise en place d'une activité de tutorat. Plus précisément, pourquoi décider d'une activité de tutorat pour une activité d'apprentissage ? Par qui, pour qui et comment est-elle conçue ? De ces questions découlent d'autres interrogations telles que les moyens mis à disposition des acteurs pour supporter le tutorat ? Ce que nous retenons, c'est que le tutorat est une activité d'encadrement fondée sur un suivi des apprenants par un tuteur humain au cours d'une activité d'apprentissage [3] et qui a pour fonction de les guider.

D'après ces définitions, le tutorat est la fonction du tuteur en ce sens où il va accompagner, guider un apprenant dans la réalisation de son activité d'apprentissage. Mais, pouvons-nous réellement dire que le tutorat est seulement l'activité du tuteur ? Le tutorat est aussi fondé sur toute intervention interne à l'environnement informatique susceptible d'aider un apprenant dans son activité d'apprentissage tels que le support apporté pour interagir avec les autres acteurs de la formation ou bien des informations relatives à l'avancement de son activité d'apprentissage ou de l'environnement. Autrement dit, le tutorat peut s'appuyer sur un ensemble de ressources humaines et technologiques [4]. Le tutorat ne se réduit pas non plus à des interventions unilatérales

entre tuteur et apprenants mais peut-être l'occasion d'échanges entre apprenants avec une participation éventuelle du tuteur humain [5]. Une autre manière de dire ceci consiste à définir le tutorat comme toute ressource (humaine/technique) support de l'accompagnement de l'activité d'un apprenant l'accompagnant dans son activité d'apprentissage.

Les interventions du tuteur sont nombreuses et de natures différentes. Par exemple, dans le cas du Suivi Pédagogique Synchronique (SPS) [6], le tuteur est présent de manière permanente pendant la durée de la session. Au niveau de la nature des interventions [7], indique que le tuteur est généralement sollicité au niveau des fonctions régulant l'organisation, le contrôle, l'évaluation et la vérification des acquis.

### **3 Problématique et objectifs**

Il n'existe pas a priori une manière unique d'aborder le tutorat. La façon dont va être conçue cette activité sera fonction de la situation d'apprentissage. En effet, la plupart des travaux aborde le tutorat du point de vue de l'instrumentation des activités du tuteur, ou d'une forme particulière du. Ainsi, à partir de l'ensemble des travaux abordés, nous adoptons une démarche globale en ce sens où nous considérons que le tuteur humain n'est pas toujours présent au cours d'une session d'apprentissage ce qui peut entraîner des problèmes dans la perception du déroulement d'une activité d'apprentissage et d'intervention et juste à temps et de manière adaptée auprès de l'apprenant. Plus précisément, nous définissons l'activité de tutorat comme étant la définition de l'accompagnement des apprenants réalisant une activité d'apprentissage. Tout ceci implique une description précise des rôles et tâches du tuteur, souvent peu ou mal définis dans les environnements d'apprentissage. Ceci revient aussi à apporter un soutien au tuteur par la conception de dispositifs d'accompagnement et la réalisation d'outils à sa disposition. Nos objectifs généraux se résument :

- déterminer la forme de tutorat à mettre en œuvre en fonction de la situation d'apprentissage.
- Proposer des outils pour les acteurs (tuteurs et apprenants).
- Déterminer la manière dont le tuteur et les apprenants peuvent interagir au cours d'une session.

L'objectif de la recherche est double : élaborer d'une part un modèle de la situation d'apprentissage, développer d'autre part un environnement support pour l'activité de suivi et de régulation de cette situation par le tuteur.

Le tutorat individuel est coûteux et les différentes organisations ne sont pas capables de faire face aux nouveaux coûts résultant d'une telle approche. Par exemple, les coûts du tutorat individuel aux collèges et aux universités est extrêmement élevé et cette approche ne peut pas être appliquée dans ces organisations. Par conséquent, nous suggérons de combiner l'approche de tutorat individuel avec les techniques de séquençage automatique de cours. Nous, nous proposons d'utiliser les réponses des tuteurs aux questions pour la prise en charge automatique des apprenants dans cette tâche. Le système est entraîné au début pour réagir dans certaines situations : les questions posées par les apprenants en langue arabe sont enregistrées et classées en utilisant les algorithmes de classification de textes. Quand les apprenants poseront des questions ; leur similarité sera vérifiée avec les précédentes.

L'outil que nous concevons sera greffé à la plate forme d'enseignement à distance MOODLE opérationnel au sein de notre institution.

#### 4 Proposition d'une modélisation du tutorat

Le modèle de tutorat que nous proposons s'articule autour de trois composantes (cf. figure 1) : l'acteur qui incarne le tuteur (le tuteurant), le bénéficiaire (le tutoré) et la nature du tutorat. Le contenu de ces composantes reprend les éléments retenus lors de l'étude précédente.

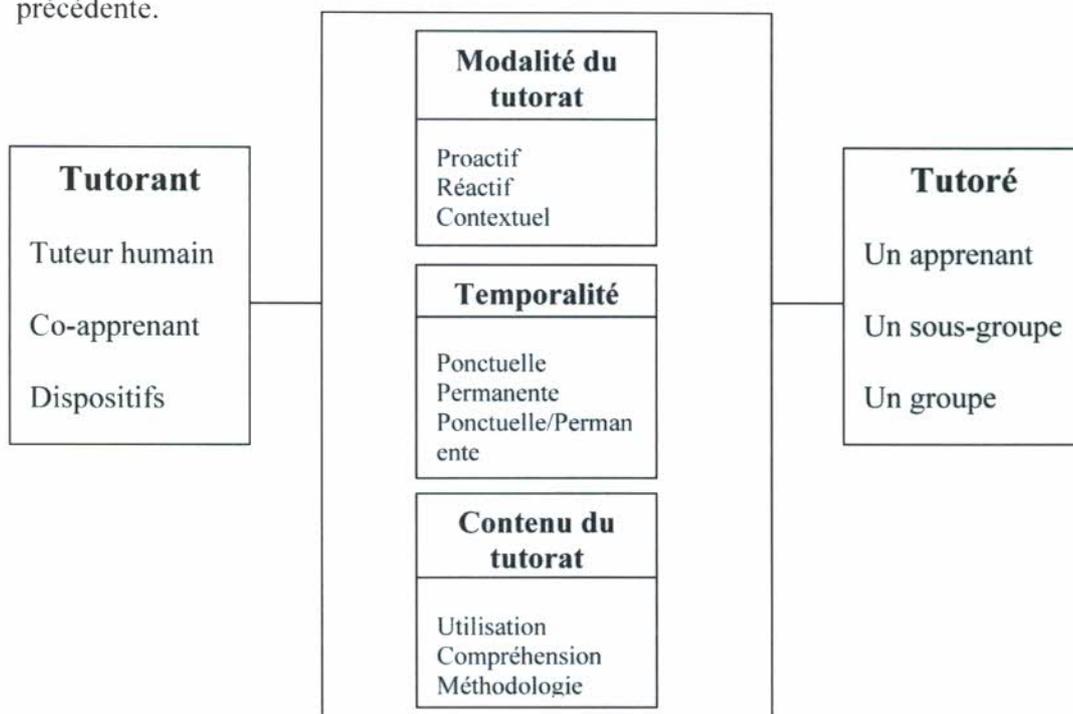


Figure 1 Le modèle descriptif de l'activité de tutorat

La première composante nous permet de différencier les acteurs, qui dans un dispositif de formation, peuvent jouer le rôle de tuteur. Un premier acteur est le tuteur humain, souvent incarné par un enseignant. Suivant les organisations pédagogiques, peuvent être présents un ou plusieurs tuteur(s) humain(s), chacun étant en charge d'un groupe ou ayant un rôle spécifique (animateur, évaluateur, ...). Il peut être aussi un co-apprenant. Le tuteurant peut être un dispositif informatique apportant des informations telles que des repères à l'apprenant par rapport au déroulement d'une activité d'apprentissage ou bien en le guidant dans l'utilisation des ressources pédagogiques et prenant en charge les réponses aux questions.

La deuxième composante du tutorat définit le bénéficiaire d'une action de tutorat au cours d'une session d'apprentissage. Les trois principaux acteurs répertoriés sont un apprenant, un groupe d'apprenants ou un sous groupe d'apprenants.

La dernière composante permet de situer la stratégie de tutorat et les outils associés. Décrire la nature du tutorat revient à déterminer (1) les formes d'intervention du tuteur, (2) le contenu du tutorat apporté à un ou plusieurs apprenants et (3) la temporalité des actions de tutorat au cours de la session d'apprentissage.

La modalité la plus classique du tutorat est qualifiée de tutorat réactif [8]. Elle se définit en deux temps : une demande explicite d'un apprenant ou d'un groupe d'apprenants puis une réponse apportée par le tuteur. La modalité réactive du tutorat permet à l'apprenant d'exprimer une difficulté rencontrée au cours d'une session d'apprentissage. La réponse fournie par le tuteur (humain ou dispositif informatique) utilise donc les informations relatives au parcours de l'apprenant et les informations concernant les interventions du tuteur auprès d'un apprenant.

De manière symétrique, le tutorat peut être qualifié de proactif [8], lorsque l'initiative d'intervenir auprès du tuteur est prise par le tuteur. Il sollicite l'apprenant car il estime que ce dernier est en difficulté par rapport à l'activité qu'il est en train d'effectuer. Les interventions proactives contribuent, et plus particulièrement lors d'interventions planifiées dans le scénario, à assurer un rythme à l'activité et donner des repères à l'apprenant par rapport à l'activité en cours.

Pour compléter ces deux modalités, nous en avons défini une troisième dite " contextuelle ". Cette modalité fournit des indications ou repères aux apprenants au cours d'une session d'apprentissage. Ces repères permettent, par exemple, aux apprenants de mieux se situer dans une activité, par rapport à un groupe ou à l'activité même auprès d'un apprenant.

La composante temporalité du tutorat caractérise le moment et la durée des échanges. Nous considérons, d'une part des interventions ponctuelles correspondant à la présence du tuteur et/ou toute information au sein de l'environnement pendant une durée limitée. En revanche, les actions persistantes concernent les éléments perceptifs présents en permanence au cours de la session d'apprentissage tels que les connexions des apprenants connectés pendant la session, un indicateur précisant que le tuteur humain est en ligne actuellement et permettre ainsi à un apprenant de le contacter à tout moment.

Nous caractérisons le contenu du tutorat de trois manières : l'*utilisation* de l'environnement, la *compréhension* et la *méthodologie*. L'*utilisation* porte sur l'organisation de l'activité d'apprentissage ainsi que la navigation au sein de l'environnement. La *compréhension* du domaine d'apprentissage concerne le contenu des diverses activités pédagogiques (par exemples la compréhension d'un énoncé, la définition d'un terme, la proposition d'un ou plusieurs exemples de résultats à une question posée dans une activité pédagogique). La *compréhension* consiste également à donner à l'apprenant une analyse du résultat issu d'une activité en terme notamment d'explication de ces erreurs. La *méthodologie* concerne la manière dont l'apprenant doit s'organiser pour résoudre un problème donné, pour aborder la formation.

## **5 Utilisation du modèle proposé**

Le modèle de tutorat précédemment décrit se décline et s'instancie lors des différentes phases du cycle de vie d'une formation. Au niveau de la phase de conception, le modèle donne les moyens de décrire la forme, la quantité et la qualité du tutorat souhaité en fonction du contexte d'apprentissage. Au cours de la phase de production du dispositif d'apprentissage, la description du tutorat nous permet d'identifier les spécifications ou les caractéristiques des outils contribuant à la gestion du tutorat et à l'accompagnement des acteurs (tuteur, apprenants) dans le cadre du tutorat.

Le modèle de tutorat contribue à définir les observables permettant de fournir au(x) tuteur(s) humain(s) ainsi qu'au(x) artefact(s) informatique(s) des éléments leur

permettant d'arbitrer entre différentes stratégies de tutorat et d'apporter des éléments de réponses pour intervenir " juste à temps " et percevoir l'activité des apprenants.

Une fois les acteurs tutorant et tutoré identifiés, il est possible à l'aide du modèle de tutorat de définir pour chaque tutorant, la nature du contenu de l'intervention et à qui celle-ci peut être adressée. Par exemple, le tutorant enseignant intervient au cours de cette activité de manière proactive et réactive, sur un contenu d'ordre de la compréhension et méthodologique. Le bénéficiaire de cette intervention peut concerner un apprenant, un sous-groupe d'apprenants ou bien un groupe d'apprenants.

Une fois déterminée la nature et le type d'intervention qu'un tutorant peut effectuer sur un tutoré, une autre étape consiste à définir plus précisément le contenu de ces interventions. Pour ce faire, nous décrivons les règles d'intervention des tuteurs au cours d'une session d'apprentissage. Pour chaque règle identifiée, le modèle de tutorat est à nouveau appliqué de manière à pouvoir sélectionner ou spécifier les outils utiles au tutorant. Voyons cela grâce aux deux règles d'intervention du tutorant enseignant.

**Règle 1 :** *l'enseignant a pour tâche d'informer de manière permanente les étudiants dans le but de rythmer l'activité. Cette règle est générale à l'activité d'apprentissage étant donnée qu'elle concerne l'organisation et le déroulement d'une étape.*

Cette intervention du tuteur auprès des étudiants est une modalité proactive planifiée puisqu'elle est à l'initiative du tuteur et non à une réponse d'un apprenant. Le contenu est d'ordre méthodologique et l'information est permanente pendant le déroulement de la session. Étant donnée que cette information est générale et qu'elle permet de rythmer l'activité, elle est adressée à l'ensemble des étudiants de l'activité en cours.

**Règle 2 :** *le tuteur a pour tâche de répondre aux questions des étudiants de manière ponctuelle lors d'une activité pédagogique. Cette règle, contrairement à celle énoncée précédemment, est relative à une activité pédagogique présente au sein d'une séquence pédagogique.*

Cette intervention du tuteur auprès des étudiants est une modalité réactive puisqu'elle est une réponse à une sollicitation d'un apprenant. Le contenu est de l'ordre de la compréhension et est ponctuel pendant le déroulement de la session puisqu'elle s'effectue à un moment donné lors d'une période synchrone où le tuteur humain est présent et est limité à une réponse à un tutorant. Étant donnée que cette information est spécifique à une difficulté rencontrée par un apprenant en particulier ou un sous-groupe impliqué dans une activité pédagogique, le tutoré est donc à la fois un apprenant ou bien une partie du groupe d'apprenants participant à l'activité d'apprentissage.

Par ailleurs, si le tuteur n'est pas présent, nous proposons d'utiliser les réponses des tuteurs aux questions posées lors des sessions d'enseignement précédentes. Un mécanisme similaire à celui utilisé dans les documents actifs peut être appliquée. Dans le cas de la pré-construction des questions/réponses (Q/A), les tuteurs doivent recenser toutes les questions susceptibles d'être posées par les utilisateurs d'un système particulier et prévoir les réponses à ces questions. Un mécanisme est implémenté permettant de répondre automatiquement aux questions après une réponse manuelle à quelques questions par le tuteur. Ces mécanismes utilisent les algorithmes de classification de textes pour déterminer les similarités entre les questions, les algorithmes d'analyse

sémantique de texte et de filtrage collaboratif afin d'utiliser les similarités entre les profils utilisateurs. Nous exposerons dans ce qui suit ces algorithmes.

## 6 Problèmes de classification du texte des questions

Quel que soit son contexte d'utilisation, chaque méthode de classification génère des classes ayant des formes et/ou un type qui lui sont spécifiques, car les caractéristiques des classes sont induites par le critère de classification optimisée par la méthode. De plus, chaque méthode de classification génère un ensemble de classes qui est organisé selon une structure (partition, hiérarchie, classification pyramidale,...) qui lui est propre également. Lorsque l'on applique une méthode de classification afin d'effectuer une indexation de cas [9], on cherche à ce que les classes obtenues soient clairement identifiées, dans le but de pouvoir affecter de nouveaux objets (cas) à chaque classe obtenue.

Le problème de la classification en général est de construire une procédure permettant d'associer une classe à un objet. Ce problème se décline en deux variantes : l'approche *supervisée* et l'approche *non-supervisée*.

Dans la première, on connaît les classes possibles et on dispose d'un ensemble d'objets déjà classés, servant d'ensemble d'apprentissage. Le problème est alors d'être capable d'associer à tout nouvel objet sa classe la plus adaptée, en se servant des exemples déjà étiquetés.

Dans la seconde (la classification non-supervisée), les classes possibles ne sont pas connues à l'avance, et les exemples disponibles sont non étiquetés. Le but est donc de regrouper dans un même *cluster* (ou groupe) les objets considérés comme similaires, pour constituer les classes (créer, par exemple, différents groupes de textes, à partir d'un ensemble de textes de tous genres, selon la similarité de leur contenu).

Dans ce cas, le problème est alors de définir cette similarité entre objets. Typiquement, la similarité entre objets est estimée par une fonction calculant la distance entre ces objets. Une fois cette fonction distance définie, la tâche de clustering consiste à réduire au maximum la distance entre membres d'un même cluster, tout en augmentant au maximum la distance entre clusters.

De nombreuses techniques peuvent alors être envisagées: clustering hiérarchique, clustering par partition, clustering statistique, fuzzy clustering, clustering par réseaux de neurones, clustering par algorithmes génétiques, ou clustering par recherche Tabou [10].

Le choix de la mesure de distance entre objets est très important. Malheureusement, trop souvent, il s'agit d'un choix arbitraire, sensible à la représentation des objets, et qui traite tous les attributs de la même manière.

Une solution pour pallier à cette limitation est celle de la prise en compte de la connaissance d'un expert, qui identifiera certains attributs, considérés comme plus pertinents que d'autres pour la tâche de classification, et leur attribuera un poids plus important lors du calcul des distances entre objets.

Mais cette solution devient très difficile à mettre en oeuvre lorsque le nombre d'attributs décrivant les données est trop grand.

Dans notre cas, une classification non supervisée de textes a été retenue vu que les tuteurs n'ont pas idée précise sur les questions qui vont être posées.

## 6.1 Classification non supervisée de textes

De nouvelles difficultés apparaissent en classification non supervisée lorsque l'on tente de prendre en compte le fait que certains attributs composant les données ont plus ou moins d'importance dans la construction de certains clusters. Cette difficulté est depuis longtemps connue en linguistique arabe ou autres : il s'agit du problème de la prise en compte du contexte dans les textes.

Typiquement, on représente un texte par un vecteur de mots, et on associe à ceux-ci leur nombre d'occurrences dans le texte. Ainsi, pour établir une distance entre textes, il suffit alors d'évaluer la similarité de ces vecteurs. L'ambiguïté des mots, problème bien identifié en linguistique, réapparaît alors en classification non supervisée [11].

Par ces exemples, on comprend bien que cette ambiguïté va poser problème lors de la comparaison de deux vecteurs dont les mots le composant n'ont pas le même sens, ni le même poids. C'est pourquoi le contexte dans lequel apparaît un mot doit intervenir lorsque l'on détermine si ce mot doit être pris en compte ou non dans la construction d'un cluster. Cette problématique, doit être prise en compte.

Ce besoin de prendre en compte le contexte pour juger de la pertinence d'un attribut semble alors naturel, même sur des données quelconques, non textuelles. Ce problème a des échos en classification non supervisée, car certains attributs peuvent être discriminants pour la formation d'un certain cluster, alors que ces mêmes attributs peuvent s'avérer peu révélateurs pour la formation d'un autre cluster. Le contexte dans lequel seront placés les attributs va donc déterminer la pertinence de leur prise en compte pour le clustering.

Or, face à cette problématique, la simple utilisation de la notion de distance entre objets va échouer. En effet, la distance entre deux objets étant définie globalement, cette technique ne pourra prendre en compte le fait que cette distance peut varier selon le contexte.

La complexité de l'algorithme envisagé étant trop importante. En effet, typiquement, ces algorithmes calculent les distances entre objets, deux à deux. La complexité de l'algorithme est alors quadratique par rapport au nombre de données en entrée, alors que sur des problèmes concernant de larges bases de données, pour que la complexité soit raisonnable, il faut qu'elle soit linéaire par rapport au nombre de données en entrée.

## 6.2 Algorithme de prise en compte de l'information textuelle

Nous avons constaté que prendre en compte la description des utilisateurs améliore les résultats. De la même façon, on peut penser que prendre en compte la description des articles aidera également. L'une des perspectives de recherche que nous envisageons est donc d'utiliser le contenu des articles lus par les utilisateurs pour aider à la caractérisation de ceux-ci.

Nous, nous retrouvons ici dans un cadre de classification, tel que nous l'avons présenté au début. En classification de textes, les objets considérés sont donc des textes, et les classes caractérisent d'une manière ou d'une autre leur contenu. Mais la nature particulière des objets textuels oblige à des traitements spécifiques.

### 6.2.1 Les différents niveaux d'analyse possible

On peut considérer le texte de la question comme un sac de mots (approche *bag of words*) en supposant tous les mots indépendants les uns des autres; ou bien on peut considérer des *n-grammes*, c'est-à-dire conserver les groupes de *n* mots ; enfin nous avons également fait appel à des informations syntaxiques (données par des grammaires) ou sémantiques

(données par exemple par un *thesaurus* : dictionnaire hiérarchique décrivant des relations sémantiques entre termes), pour cibler les concepts plus généraux émanant du texte.

## 6.2.2 Application l'approche bag-of-words

### a. Le preprocessing

La première étape dans la catégorisation d'un texte est de transformer la question qu'on veut analyser en une représentation convenable pour l'algorithme d'apprentissage et la tâche de classification. Il s'agit donc généralement (surtout pour l'approche "bag-of-words"), dans un premier temps, de *normaliser* le texte, c'est-à-dire de se débarrasser des caractères spéciaux, puis d'utiliser une *stoplist* (ou liste de mots vides) pour supprimer tous les mots qui ne sont pas porteurs de sens (pronoms, prépositions, conjonctions, articles, etc.). Puis, sur les mots qui restent, on effectue une *lemmatisation*, c'est-à-dire une analyse des formes morphologiques des mots, pour ne garder que leur racine (la forme sous laquelle on les cherche dans un dictionnaire). Enfin, on regroupe les mots identiques, en les comptant ou non, selon l'algorithme qu'on veut développer.

### b. Représentation d'une question

La représentation la plus utilisée pour stocker ensuite en mémoire la question analysée est le *modèle vectoriel*, c'est-à-dire qu'une question est représentée par le vecteur des mots qui apparaissent le plus souvent dans la question. Chaque utilisateur est alors associé à un ensemble de questions, et donc à une matrice  $A=(a_{ik})$  des mots par question. Selon l'algorithme choisi, on associera à chaque mot composant le vecteur son poids dans la question (son nombre d'occurrences, ou le fait qu'il soit mis en valeur, etc.) (*word frequency weighting*), ou bien on ne s'intéressera qu'à l'absence ou la présence d'un mot (algorithme de *Bayes*). C'est à partir de ce vecteur engendré qu'on lance l'algorithme de classification.

## 6.2.3 La notion de profil

Une fois la classification thématique de question effectuée, on peut envisager d'intégrer les classes et mots majoritaires à un profil d'apprenant. La représentation la plus simple d'un profil d'apprenant serait alors un vecteur des thèmes et mots les plus fréquents dans les lectures de celui-ci. Mais une autre représentation possible d'un profil peut être la suivante: celui-ci serait tout d'abord constitué d'un *profil initial* rempli par l'apprenant lui-même; puis d'un *profil thématique à long-terme*, représenté sous la forme de vecteurs (un positif pour les textes considérés comme pertinents et un négatif pour les non-pertinents) des mots et classes les plus fréquents dans l'ensemble des lectures de l'apprenant; et enfin d'un *profil thématique à court-terme*, représenté sous la forme de vecteurs (positifs et négatifs) des mots et classes les plus fréquents dans les  $N$  dernières lectures de l'apprenant. Avec cette représentation, l'apprenant ne serait donc pas enfermé dans un profil figé, mais l'ensemble des thèmes abordés serait toujours conservé en mémoire. Enfin, si l'on veut intégrer le profil apprenant dans un groupe d'apprenants similaires, il s'agit encore ici de classification, et on se retrouve donc de nouveau confronté au choix entre la *classification non-supervisée* (faire du clustering sur les profils apprenants pour former des groupes d'intérêt), et la *classification supervisée* (en caractérisant les groupes attendus a priori)

## 6.2.4 Limites de cette approche

Les points sensibles de l'analyse basée sur le contenu sont les suivants:

- Quel niveau d'analyse appliquer au texte de la question?
- Quelles sont les informations les plus discriminantes du texte de la question?
- Comment caractériser les classes auxquelles sont associés les textes (ensemble de classes fixé à l'avance/déterminé par clustering) ?
- Les textes appartiennent-ils à une seule classe ou à plusieurs (classification monoclasse/multiclasse) ?
- Comment valider les procédures de classification ?
- Quel est le lien entre les classes de textes et les profils ? Quelle représentation adopter pour un profil ?

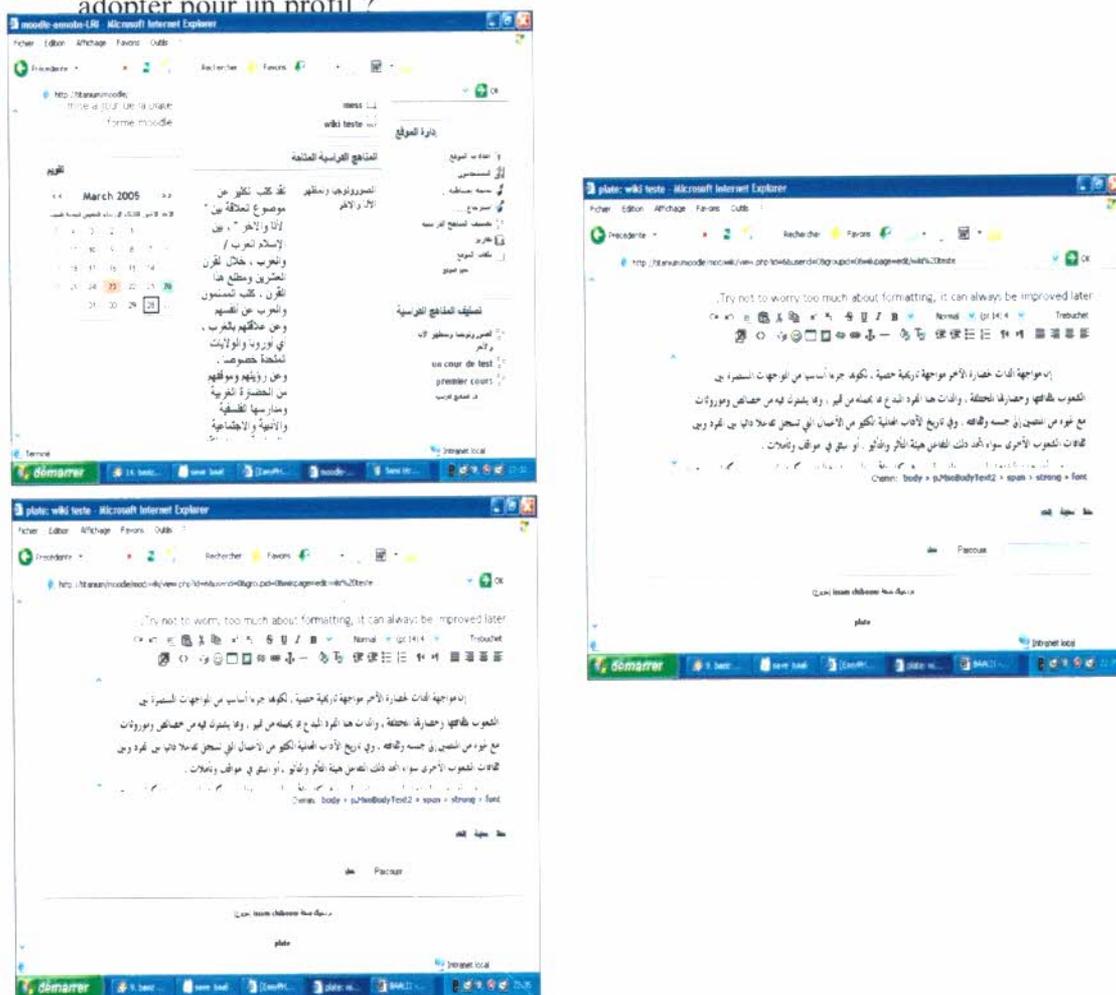


Figure 2 Les interfaces d'un cours en arabe de la plate forme

Après l'utilisation de cette approche, nous nous orientons actuellement vers les algorithmes de filtrage collaboratif versus analyse de contenu où chaque apprenant est associé à un vecteur de notes sur les pages qu'il a parcouru. Aussi, une combinaison des méthodes pourrait être très bénéfique. Le profil pourrait alors consister en des groupes d'utilisateurs considérés comme similaires pour un thème donné.

### **Mise en œuvre**

Le modèle de description de situation d'apprentissage a été instancié et expérimenté dans la plate forme d'enseignement à distance open source MOODLE. La plate forme est installée dans notre serveur local et des cours en langue arabe y ont été intégrés. Des apprenants sont actuellement entraînés à utiliser cette plate forme et dialoguer surtout avec leurs tuteurs (cf. figure 2). L'outil greffé à la plate forme est entraîné à capitaliser les questions posées par les apprenants et les réponses des tuteurs avec une application de la première approche de classification de textes arabes précédemment exposée.

### **Conclusion**

Nous avons présenté dans cet article une méthodologie pour la conception de dispositifs permettant l'accompagnement des apprenants au cours d'une activité d'apprentissage. Pour cela, nous utilisons un modèle d'organisation d'une activité de tutorat nous permettant d'organiser les tâches entre les acteurs tuteur (dispositif informatique, tuteur humain et co-apprenant) et d'assurer auprès des tutorés un accompagnement adapté à la situation d'apprentissage. Un outil pour la prise en charge des questions des apprenants en langue arabe a été implémenté, toutefois l'algorithme utilisé est en phase de validation et d'autres approches sont envisageables en regard des résultats obtenus.

### **Références**

- [1] Jezegou A. 1998. La formation à distance : enjeux perspectives et limites de l'individualisation. L'Harmattan Paris.
- [2] Hotte R.; et Leroux P. Eds. 2003. Technologies et formation à distance. Revue STICEF Vol 10.
- [3] Deschênes, A.-J. and Paquette, D. 1996. *Cahier d'étude. Programme de formation continue des enseignants (Première année)*, sous la direction de A.-J. Deschênes, Québec, Centre Romand d'Enseignement à Distance et Télé-Université.
- [4] Power, M. ; Dallaire S. ; Dionne M. ; and Théberge C. 1994, *L'encadrement des étudiantes et des étudiants en situation d'apprentissage à distance à l'Université du Québec à Rimouski*. Département des sciences de l'éducation, Monographie n°42, collection l'autre, Editions GREME.
- [5] Dionne, M. ; Mercier J. ; Deschênes A.-J. ; Bilodeau H. ; Bourdages L. ; Gagné P. ; Lebel C. ; and Raba-Donath A. 1999. Profil des activités d'encadrement comme soutien à
- [6] Després C. and Leroux P. 2003. Tutorat synchrone en formation à distance - Un modèle pour le suivi pédagogique synchrone d'activités d'apprentissage à distance, 1ère Conférence sur les Environnements Informatiques d'Apprentissage Humain, EIAH 2003, Strasbourg, 15-17 avril, 2003, p. 139-150.
- [7] Barnier, G., 2001. *Le tutorat dans l'enseignement et la formation*, L'Harmattan.
- [8] De Lièvre B. and Depover C. 2001. Apports d'une modalité de tutorat proactive ou réactive sur l'utilisation des aides dans un hypermédia de formation à distance, *Hypermédiat et apprentissages* 2001, 9-11 avril, Grenoble.

- [9] BENEDEK A. and TROUSSE B. (2002) : *Adaptation of Self-organizing Maps for CBR Case Indexing*. Symbolic and Numeric Algorithms for Scientific Computing. 9-12 Octobre, Timisoara, Roumanie.
- [10] Jain, A., Murty, M., and Flynn, P. (1999). Data clustering: a review. *ACM Computing Surveys*.
- [11] Slonim, N. and Tishby, N. (2000). Document clustering using word clusters via the information bottleneck method. In *Research and Development in Information Retrieval*.