

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Mouloud MAMMERY de Tizi-Ouzou
Faculté génie électrique et informatique
Département informatique



Mémoire de master

Thème :

Conception et réalisation d'un système de détection des apprenants en difficultés
cas d'étude : les forums

Réalisé par :

BOUMATI MOHAMED
CHALLAL ANIS

Proposé et encadré par :

M^{me} AIT ADDA SAMIA

Année universitaire 2014/2015

Remerciements :

La première personne que nous tenons à remercier est notre promotrice Mlle Samia Ait-Adda, pour son orientation, sa confiance et sa patience qui ont constitué un apport considérable sans lequel ce travail n'aurait pas pu aboutir. Qu'elle trouve dans ce travail notre plus humble reconnaissance.

Nos remerciements s'étendent aussi à nos chers parents qui nous ont encouragé, soutenu et qui ne cesseront de le faire pour nous voir heureux et épanouis.

Et enfin, nous tenons à gratifier tous les professeurs qui nous ont enseigné et qui par leurs compétences nous ont soutenu dans la poursuite de nos études ainsi que tous ceux qui ont, de près ou de loin, contribué à la réalisation de ce travail.

Dédicaces :

A la mémoire de ma tendre grand-mère.

Je dédie ce modeste travail à mes chers parents qui sont le symbole de patience et de tendresse et qui se sont sacrifiés pour mon bonheur et ma réussite. Vous êtes un modèle pour moi et j'aimerai un jour marcher sur vos pas.

Je fais aussi une dédicace à mes sœurs qui ont su donner du charme à ma vie et à mon binôme et ami d'enfance Mohamed. Je profite de cette occasion pour vous souhaiter à tous une longue vie et de la réussite dans vos projets.

Sans oublier les merveilleuses rencontres faites toutes au long de mon cursus universitaire, le collectif ACR-AGA qui n'ont pas hésité à me soutenir et à m'aider. En témoignage de l'amitié qui nous unie, je vous dédie ce modeste travail et je vous souhaite un bon avenir.

Anis

Dédicaces :

Je dédie ce mémoire à mes parents qui ont œuvré pour ma réussite, de par leur amour, leur soutien, tous les sacrifices consentis et leurs précieux conseils, pour toute leur assistance et leur présence dans ma vie, recevez à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui peut être fier de trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Mes frères, mon binôme et ami d'enfance Anis qui n'ont cessé d'être près de moi en m'apportant du courage et de la générosité.

Mohamed

Sommaire

Introduction	p.1
Annonce de la problématique.....	p.2
Organisation du mémoire.....	p.2
Partie 1 : Apprentissage sur internet	p.3
- Chapitre I : Introduction au E-Learning.....	p.4
I-1 : Qu'est ce que l'E-Learning.....	p.4
I-1.1 : Les plateformes E-Learning.....	p.5
I-1.2 : Les avantages de la formation à distance.....	p.6
I-1.3 : Les inconvénients de la formation à distance.....	p.6
I-2 : Les différents moyens d'apprentissage sur Internet.....	p.7
I-2.1 : Bibliothèques digitales.....	p.7
I-2.2 : Wiki.....	p.8
I-2.3 : Web et moteur de recherche.....	p.8
I-2.4 : Les outils de communication.....	p.9
- Chapitre II : Les Traces.....	p.15
II-1 : Définition.....	p.16
II-2 : Types de traces.....	p.17
II-3 : Modèle de traces.....	p.18
II-4 : Système à base de traces.....	p.19
II-5 : Formes de traces.....	p.24
Partie 2 : Ontologies et Indexation	
- Chapitre III : Les ontologies.....	p.32
III-1 : Définition.....	p.32
III-2 : Composants de l'ontologie.....	p.33
III-3 : Typologie des ontologies.....	p.34
III-4 : Construction des ontologies.....	p.36
III-5 : Formalisation et outils de représentation des ontologies.....	p.37
III-6 : Domaine d'application des ontologies.....	p.38
III-7 : Indexation.....	p.40
Partie 3 : Conception et Réalisation	
- Chapitre IV : Conception.....	p.43
IV-1 : Présentation du milieu d'apprentissage.....	p.44
IV-2 : Traçage de l'apprenant.....	p.46
IV-3 : Construction du corpus apprenant.....	p.51
IV-4 : Mise en place d'une ontologie du domaine.....	p.53
IV-5 : Indexation.....	p.55
IV-6 : Projection Activité/Ontologie.....	p.57

- Chapitre V : Réalisation.....	p.60
V-1 : Description du matériel utilisé.....	p.60
V-2 : Logiciels utilisés.....	p.60
V-3 : Présentation du langage de programmation utilisé.....	p.61
V-4 : SGBD.....	p.61
V-5 : Connexion à la base de données.....	p.63
V-6 : Quelques interfaces graphiques de l'application conçue.....	p.64
Conclusion.....	p.70
Annexes.....	p.71

Liste des figures :

Figure 1 : Le système E-Learning.....	p.4
Figure 2 : Exemples de ressources offertes par une plateforme.....	p.5
Figure 3 : Fonctionnement du système de collecte.....	p.21
Figure 4 : Système à base de traces.....	p.24
Figure 5 : Exemple de table de traces dans une base de données.....	p.25
Figure 6 : Exemple de fichier logs.....	p.26
Figure 7 : Forme d'un document XML.....	p.27
Figure 8 : Exemple d'une trace d'interaction sous format XML.....	p.28
Figure 9 : Interactions de l'apprenant durant une session d'apprentissage.....	p.45
Figure 10 : Architecture générale du système.....	p.46
Figure 11 : Interface du MiniKeyLogger lors du traçage.....	p.48
Figure 12 : Fragment de fichier trace XML.....	p.49
Figure 13 : Algorithme de filtrage.....	p.51
Figure 14 : Etapes de construction du corpus.....	p.52
Figure 15 : Modèle entités-association.....	p.53
Figure 16 : Ontologie SKOS.....	p.55
Figure 17 : Exemple d'indexation et de projection.....	p.58
Figure 18 : Diagramme de séquence de notre système.....	p.60
Figure 19 : Capture d'écran PostgreSQL lors de la connexion à la BDD.....	p.63
Figure 20 : Capture d'écran PostgreSQL lors de la création d'une table.....	p.63
Figure 21 : Capture d'écran PostgreSQL lors de l'ajout d'une table.....	p.64
Figure 22 : Communication entre le SGBD et la base de données.....	p.65
Figure 23 : Fichier trace XML généré lors de la simulation.....	p.66
Figure 24 : Capture d'écran de l'application, authentification.....	p.67
Figure 25 : Capture d'écran de l'application, changement du mot de passe.....	p.67
Figure 26 : Capture d'écran de l'application, chargement du fichier.....	p.68
Figure 27 : Capture d'écran de l'application, sélection d'un fichier.....	p.69
Figure 28 : Capture d'écran de l'application, choix du fichier.....	p.69
Figure 29 : Capture d'écran de l'application, message de succès.....	p.70
Figure 30 : Capture d'écran de l'application, affichage global des activités.....	p.71
Figure 31 : Capture d'écran de l'application, introduction d'un seuil.....	p.71
Figure 32 : Capture d'écran de l'application, affichage des résultats individuels.....	p.72
Figure 33 : Capture d'écran de l'application, évaluation du cours.....	p.72
Figure 34 : Capture d'écran de l'application, statistiques générales.....	p.73
Figure 35 : Capture d'écran de l'application, affichage des apprenants évalués.....	p.74
Figure 36 : Capture d'écran de l'application, liste des apprenants en difficulté.....	p.74

Liste des tableaux :

Tableau 1 : Description des noms des balises.....	p.29
Tableau 2 : Modèle de représentation des concepts.....	p.56

Introduction Générale :

Depuis sa naissance jusqu'à aujourd'hui, l'Homme n'a cessé d'accroître ses compétences et l'expansion de sa maîtrise dans plusieurs domaines, tels que la médecine, l'agriculture, l'informatique...

Actuellement, avec le progrès de la technologie et des innovations, principalement avec l'explosion d'Internet, l'homme est devenu un consommateur quotidien. Il l'utilise tous les jours pour se divertir (jeux, vidéos, musiques...), s'actualiser, ou bien, faire des recherches dont le but est de s'instruire et de peaufiner ses connaissances.

En prenant comme exemple un apprenant durant son parcours d'apprentissage, on observera qu'il sera exposé à plusieurs informations concernant son domaine d'étude. Vu la grandeur d'Internet, celles-ci peuvent être bien acquises ou mal assimilées.

Afin d'améliorer le mode d'apprentissage et la facilité d'absorption de ces connaissances sur Internet, des plateformes d'apprentissage ont vu le jour.

La majorité d'entre elles sont en libre accès comme : Caroline, Chamilo, Moodle... (Etc.)

Une plateforme éducative se définit comme un support informatique rassemblant divers outils logiciels devant permettre d'organiser et de gérer (via Internet ou Intranet), des ressources pédagogiques en ligne, des inscriptions, d'animer une classe virtuelle ...

Mais les notions importantes des plateformes sont leurs structurations et leurs adaptabilités aux besoins des apprenants, notamment sur le contenu des cours. Car si elles ne répondent pas aux attentes des apprenants, elles n'amélioreront pas et ne faciliteront pas leur apprentissage. Du coup, certains apprenants seront en difficulté à cause des défaillances de la plateforme

Annonce de la problématique :

Dans ce travail, nous nous intéressons aux difficultés rencontrées par un (ou plusieurs) apprenants. Pour cela, nous allons analyser les traces d'interaction de cet (ces) apprenant(s) durant sa (leurs) formation(s), c'est-à-dire, le (les) tracer pour connaître le motif de ses (leurs) recherches en dehors de la plateforme qui sont en relation avec le cours qu'il(s) étudie (ent). Ainsi, nous essayerons de trouver pourquoi l'apprenant a-t-il recours à des services externes à la plateforme d'apprentissage pour se procurer les concepts du domaine ?

En général, le contexte scientifique de l'approche proposée est celui de l'étude d'activités des apprenants à l'extérieur du système d'apprentissage.

Organisation du Mémoire :

Pour y répondre, notre travail se composera de trois parties. La première fera objet d'une description de l'apprentissage sur internet. Elle fera référence au E-Learning (l'auto-apprentissage), et présentera la modélisation utilisateur; c'est-à-dire, l'environnement lié à l'apprenant. C'est dans cette partie que nous aborderons la notion de traces. La seconde partie relatera sur les ontologies et l'indexation. Enfin la dernière partie est la partie conception de notre approche et réalisation de l'application.

La conclusion résumera la méthode et les éléments de recherche utilisés dans notre approche pour arriver à trouver une solution à la problématique annoncée.

PARTIE 1 :
APPRENTISSAGE SUR
INTERNET

Chapitre I : Introduction au E-Learning

Introduction :

L'expansion d'Internet permet aujourd'hui d'avoir accès à énormément d'informations. Cet effet a permis à de nombreuses personnes de sortir de leur ignorance en suscitant leur curiosité et leur soif d'apprentissage. L'apprentissage basé sur Internet ou en ligne peut ouvrir dans le monde entier de nouvelles voies à l'enseignement. Nous évoquerons ce concept dans les chapitres ci-dessous en définissant tout d'abord la notion d'apprentissage en ligne (E-Learning), ses plateformes ainsi que les différents moyens qui facilitent cet apprentissage.

I-1. Qu'est ce que l'E-Learning ?

Le E-Learning, ou formation en ligne, est une technique de formation reposant sur la mise à disposition de contenus pédagogiques via un support électronique : Cd-rom, Internet, intranet, extranet...

Le E-Learning désigne les outils, les applications et l'ensemble des contenus mis à disposition d'un stagiaire dans le but d'une formation pédagogique. Longtemps réduites à des supports cd-rom, le E-Learning a évolué et utilise dorénavant le web et différentes applications. De ce fait, des plateformes ont vu le jour et ont apporté une nouvelle façon d'apprendre (gain de temps, facilité, meilleure recherche) à la personne qui se forme.

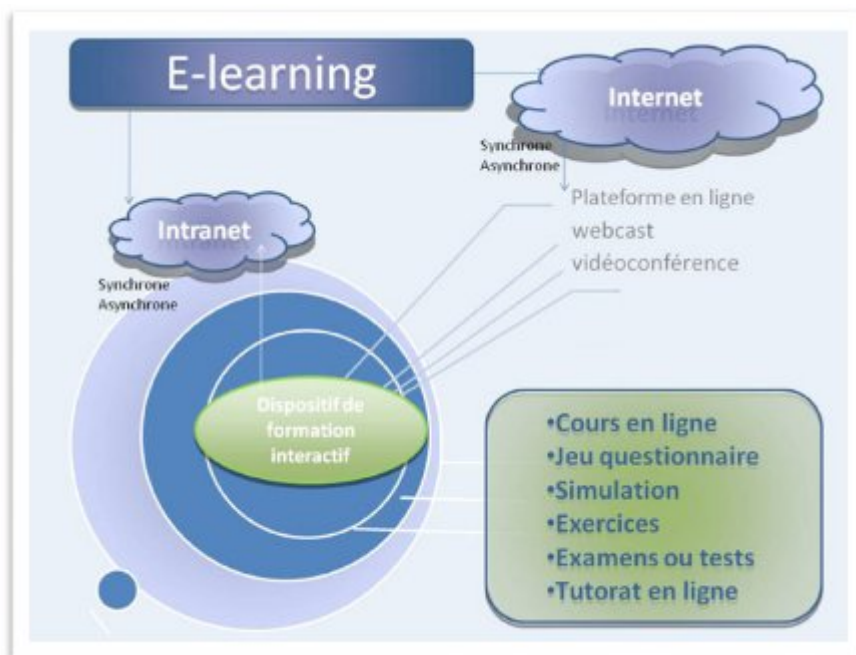


Figure 1: Le système E-Learning.

1.1. Les plateformes E-Learning :

Définition d'une plateforme :

Une **plateforme de formation** est un logiciel installé sur un serveur informatique. Pour l'utilisateur, celui-ci se présente sous la forme d'un site web auquel il peut se connecter et accéder notamment aux différentes formations prévues par le formateur. Côté formateur, la plateforme de formation permet de nombreuses actions, par exemple, effectuer simplement le suivi détaillé de la formation. [1].

L'apprentissage en ligne, c'est la souplesse de temps et de lieu. La solution E-Learning s'adresse à tous : aux étudiants comme aux enseignants en quête de savoir, aux cadres comme aux employés, quel que soit leur niveau de qualification et leur âge et peut être choisie par les personnes handicapés. Ce sont toutes ces personnes et les formateurs qui supervisent les plateformes qui sont les acteurs participant à l'animation de ces dernières.

Le choix de suivre une formation en ligne est offert à tous et n'exclut personne. Il existe plusieurs plateformes en libre accès sur internet. Nous pouvons citer Caroline, Chamilo, Moodle..(etc.) Elles offrent plusieurs services tels que l'organisation et la gestion des ressources pédagogiques en ligne(apporter des cours, des exercices), des inscriptions, permettre le chat et l'interactivité dans les forums, d'animer une classe virtuelle...Nous

présenterons dans ce qui suit, les avantages et les inconvénients du E-Learning.

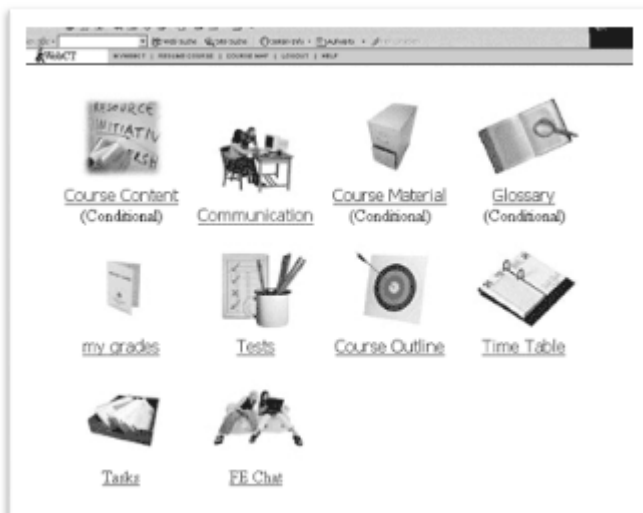


Figure 2 : Exemples de ressources offertes par une plateforme.

1.2. Les avantages de la formation à distance :

L'enseignement à distance est apparu au 19ème siècle. Il est né de la volonté de rendre accessible la formation pour tous. Ce public se compose principalement :

- Des personnes éloignées géographiquement des établissements de formation
- Des personnes voulant concilier activité salariale et formation
- Des hommes et femmes au foyer
- Des personnes handicapées
- Des expatriés
- Des personnes préférant l'apprentissage par correspondance par choix

Elle s'est imposée aujourd'hui comme une méthode de formation à part entière car elle possède de nombreux avantages :

- Elle s'adapte aux contraintes de temps et à l'organisation de chacun
- Elle permet d'alterner formation pratique et théorique (formation en alternance, stages en milieu d'application)
- Elle permet de travailler de façon autonome et indépendante
- Elle développe l'esprit de curiosité et de recherche
- Le coût de la formation est amoindri (locaux, déplacement des formateurs)
- Elle permet d'éviter l'inconvénient du transport et l'éloignement géographique

1.3. Les inconvénients de la formation à distance :

Malgré des avantages indéniables, la formation à distance a également des inconvénients tels que :

- L'appréhension de l'outil informatique et les éventuelles difficultés d'utilisation
- La mauvaise gestion de l'autonomie due à une absence d'encadrement
- L'obligation de motivation et de forte implication.
- Avoir pas ou peu de contacts directs avec les formateurs.

Dans le (sous)chapitre précédent, nous avons défini la notion du E-Learning en passant par les différents avantages et inconvénients. Dans ce qui suit, nous décrirons les autres moyens qui permettent aussi l'apprentissage mais en dehors de la plateforme.

I-2. Les différents moyens d'apprentissage sur internet :

Quand l'apprenant étudie sur une plateforme et que vient le moment où il ne comprend pas un terme (concept), il fait des recherches en dehors de la plateforme. Nous allons citer quelques moyens que cet apprenant peut utiliser sur internet pour mener ses recherches.

2.1. Bibliothèques Digitales :

Attirés par la souplesse et l'accessibilité qu'offre l'apprentissage en ligne, les apprenant(e)s qui étudient en ligne ont besoin aussi de services de soutien pédagogique en ligne, notamment : orientation, conseils, évaluation et perfectionnement de l'aptitude aux études, accès et aide à l'égard des bibliothèques.

Une pluralité de définitions : Les bibliothèques digitales (plus souvent appelées bibliothèques numériques) sont des collections organisées de contenu numérique mises à la disposition du public. Il s'agit d'un contenu qui a été soit préalablement numérisé (copies de livres ou autres documents), soit initialement produit dans un format numérique [2].

Autre définition : « Les bibliothèques numériques sont des organisations qui offrent des ressources, y compris en personnel, pour sélectionner, structurer, offrir un accès intellectuel, distribuer et conserver l'intégrité de documents sous une forme numérique. Une bibliothèque numérique garantit également un accès sur la durée aux œuvres électroniques dans le but d'être aisément et à un moindre coût disponibles à un ou plusieurs publics successifs » [3].

Dans la pratique, les bibliothèques numériques restent un concept fuyant et sont extrêmement diverses, du point de vue :

- du nombre de documents et du type de documents mis à disposition.
- des standards techniques adoptés.
- des fonctionnalités proposées.
- de l'accompagnement des contenus.
- des partenariats mis en place.
- des modèles économiques retenus, etc.

Il existe plusieurs grandes bibliothèques numériques telles que : Gallica, bibliothèque numérique de la BnF, « Online Gallery », de la British Library et Wikipédia qui reste la plus accessible et la plus sollicitée sur Internet.

2.2. Wiki

Wikis et Wikipédia est une encyclopédie collaborative en ligne à laquelle chacun peut contribuer librement en fonction de ses connaissances et compétences. L'équipe de Wikipédia définit ainsi cette encyclopédie:

Wikipédia (API [wikipe'dja] ou [vikipe'dja]) est une encyclopédie universelle et multilingue écrite collaborativement sur Internet avec la technologie wiki. Ceci signifie que tous les internautes peuvent écrire dans l'encyclopédie ou réutiliser son contenu librement distribuable et gratuit par le biais d'un navigateur Web. Sa politique éditoriale est basée sur la neutralité de point de vue. (Wikipedia 2006).

2.3. Web et moteur de recherche :

Anthony Harmant définit le moteur de recherche comme un logiciel permettant de retrouver des ressources (pages Web, forums Usenet, images, vidéo, etc.) associées à des mots quelconques. Certains sites Web offrent un moteur de recherche comme principale fonctionnalité ; on appelle alors moteur de recherche le site lui-même. C'est un outil constitué de "robots", encore appelés spiders, crawlers ou agents qui parcourent les sites à intervalles réguliers et de façon automatique (sans intervention humaine, ce qui les distingue des annuaires) pour découvrir de nouvelles adresses (URL). Ils suivent les liens hypertextes (qui relient les pages les unes aux autres) rencontrés sur chaque page atteinte. Chaque page identifiée est alors indexée dans une base de données, accessible ensuite par les internautes à partir de mots-clés.]

Nous pouvons citer des exemples tels que Google, Yahoo, Bing, Ask...

2.4. Les outils de communication et de collaboration :

Les outils de communication ne sont pas nécessairement initialement conçus pour l'apprentissage. Mais ce qui nous intéresse ici, c'est leur application pédagogique, c'est-à-dire

leur utilisation en contexte d'apprentissage, leur intégration dans un dispositif ou dans un parcours de formation.

Les outils de communication et de collaboration jouent un rôle essentiel en E-Learning puisqu'ils rendent possible le tutorat, facilitent le travail des tuteurs et permettent la construction d'une véritable communauté d'apprentissage. Ils se subdivisent en deux catégories:

- **les outils synchrones.** Ils permettent de communiquer en temps réel entre des personnes distantes géographiquement: messagerie instantanée (chat, clavardage, etc.), téléphonie par le net, audioconférence et vidéoconférence, etc. En effet, pour "chatter", il faut que les "chatteurs" se trouvent en même temps face à leur ordinateur respectif;
- **les outils asynchrones.** Ils permettent des échanges décalés dans le temps et l'espace: courrier électronique (e-mail), forum (de discussions), portfolio, wiki, blog et micro-blog, etc. Quand une personne envoie un mail, le récepteur peut le lire quelques minutes, heures ou jours plus tard.

Internet s'utilise un peu plus pour promouvoir le processus d'apprentissage et contribuer au développement durable des compétences. C'est grâce aux outils mentionnés auparavant que l'apprenant arrive à poursuivre facilement son apprentissage sur le net. Nous allons les décrire et citer leur importance dans le sous-chapitre suivant.

2.4) 1.Importance des outils de communication dans le parcours d'un apprenant :

A distance contrairement aux cours en présentiel, il n'y a pas de contact physique, mais les différents moyens de communication mis en place tels que messagerie (chat), forum, blog... permettent d'échanger et d'avoir des contacts avec d'autres personnes. Ce qui rend la formation plus vivante et créer des liens qui facilitent l'apprentissage et augmente le confort du E-Learning.

Allié à l'ordinateur, l'Internet constitue un support nouveau et performant pour la diffusion de l'information. L'un de ses avantages est la possibilité qu'il offre de chercher dans le Web riche en informations scolaires et parascolaires, de dialogue et de communication immédiate avec d'autres internautes, ce qui dépasse la relation individuelle entre le lecteur et le livre ainsi que la relation traditionnelle entre l'élève et le professeur.

Car si l'apprenant passe son temps sur une plateforme d'étude, cela ne l'empêche pas de consulter d'autres pages web pour se procurer les concepts du domaine ou de discuter de ses difficultés sur des forums afin que d'autres personnes puissent lui venir en aide.

Nous allons décrire les autres moyens, que ceux cités précédemment, qu'Internet met en disposition de l'apprenant pour lui faciliter sa formation.

2.4) 2. Les Forums :

Qu'est-ce qu'un forum?

Nous pouvons définir un forum comme :

« Un lieu virtuel où se construit le discours d'un groupe, où l'apprenant s'approprie de nouvelles connaissances en conversant avec d'autres.

Une agora de collaboration et de socialisation qui permet:

- de produire collectivement des objets concrets (cf. recherche)
- d'exposer des idées, d'élaborer sa pensée, de construire de nouvelles connaissances, de les confronter, de les valider et de les confirmer avec le soutien du groupe
- de faciliter l'apprentissage de connaissances complexes (synthèse, analyse, etc.)
- de favoriser une attitude réflexive sur l'apprentissage
- à un groupe de développer un engagement sociocognitif qui peut donner plus de sens à un apprentissage donné »[4].

Les différents forums pour l'apprentissage :

Il existe plusieurs forums tels que :

- Futura-sciences
- Développez
- Comment-ça-marche

Qui aident les apprenants (et les autres personnes) à trouver de la lumière dans leurs recherches. Pour décrire l'un d'eux, nous choisirons Futura-sciences, vu que les deux autres sont très connus.

D'après le Figaro(quotidien français), Futura-sciences est considéré comme un magazine online de la découverte, de la science et de l'innovation, Il propose à ses lecteurs un contenu mis à jour en permanence et richement illustré avec un suivi quotidien de l'actualité, des dossiers de fond, de nombreux forums de débats et de discussions et une multitude d'autres espaces de connaissance.

Vu la pluralité des forums sur la toile, ceux-ci essayent d'attirer le plus d'utilisateurs en leur offrant d'autres possibilités souvent très pratiques que nous verrons ci-après.

Avantages et ressources offerts pour les apprenants :

La participation au forum peut permettre aux apprenants de :

- S'obliger à produire des arguments.
- Poser des questions, émettre des hypothèses.
- Se repositionner par rapport aux arguments des autres.
- Travailler en collaboration.
- S'apercevoir qu'on peut apprendre de ses pairs.
- Connaître l'avis des autres.
- Participer à la discussion de différents groupes
- Passer par la production d'un écrit
- Ne pas être contraint par le temps
- Organiser leurs travaux de groupe

Concernant l'Algérie, le site web "espace-etudiant.net" comporte un forum interactif composé de cours, d'exercices, travaux pratiques et examens corrigés pour tous les niveaux et spécialités. Néanmoins, les étudiants (apprenants) se tournent plus vers les groupes situés sur Facebook (ex: le carrefour des étudiants algériens, les informaticiens d'Algérie...) ou bien les plus utilisés sur le net, en occurrence Développez et Comment-ça-marche citez ci-dessus.

2.4) 3.Les Blogs :

On l'appelle aussi carnet web, joueb ou carneticiel. Il nous vient du mot Weblog inventé en 1997 par Jorn Barger. C'est un site personnel composé de commentaires, impressions, réactions publiés au jour le jour et associés à des informations glanées sur la toile. Le message le plus récent apparaît automatiquement en haut de page, tandis que les anciens sont peu à peu archivés, eux aussi de façon automatique.

C'est aujourd'hui un moyen très répandu pour s'exprimer sur la toile : on trouve des blogs de toutes sortes, du journal de campagne d'un candidat à la tribune de journalistes indépendants qui trouvent dans ce support un moyen de s'exprimer de façon plus personnelle et indépendante. Il existe aussi de plus en plus de blogs pédagogiques dans lesquels professeurs et étudiants échangent sur des thèmes libres ou imposés; exemples : " L'Etudiant ", " MSDN Blogs ", " Octo " , " Adelaide "....

Les caractéristiques du blog :

1. **son caractère non institutionnel** : les blogs sont en général le fait d'individus ou de groupes d'individus intervenant à titre personnel.
Intérêt pédagogique - encourage la créativité et l'autonomie en étant un espace d'expression individuelle libre.
2. **l'utilisation d'outils dynamiques** : les blogs sont très fréquemment mis à jour et leurs archives restent consultables. La fonction "commentaire" permet aux lecteurs d'intervenir, de donner leur point de vue sur un article.
Intérêt pédagogique : on peut suivre la progression de ses acquisitions en relisant les productions antérieures.
3. **l'interconnexion** : les blogs contiennent de multiples liens vers d'autres sources d'informations, sites web ou autres blogs.

Intérêt pédagogique : incite à la recherche, encourage l'esprit d'analyse et de synthèse. Facilite l'échange et la communication. Moyen puissant de construction collective des savoirs *connaissance de la toile*.

4. **l'aspect multimédia** : il existe aujourd'hui des podcasts (audioblogs) et des vidéos blogs.

Intérêt pédagogique : permet de travailler tout types de compétences, à la fois de compréhension et de production.

2.4) 4.Apprendre avec les réseaux sociaux :

Les réseaux sociaux sont aujourd'hui des systèmes d'organisation en plein essor. Ils offrent de multiples potentialités et génèrent de fortes dynamiques formatives sur le plan individuel et collectif.

Définition :

Un réseau social désigne un ensemble de personnes réunies par un lien social. À la fin des années 1990, des réseaux sociaux sont apparus sur Internet, réunissant des personnes via des services d'échanges personnalisés, chacun pouvant décider de lire les messages de tel ou tel autre utilisateur. [Futura-Science.]

Quelques exemples de réseaux sociaux :

Facebook : LE réseau social ! Celui dont tout le monde parle, et qui contient plus de 300 millions de comptes !

LinkedIn : Un réseau social important, ciblé sur les relations entre professionnels.

MySpace : Un réseau principalement centré autour de la musique et de la découverte de nouveaux groupes, à la clientèle assez jeune.

Twitter : Un réseau un peu différent, basé sur le principe du microblogging (publication fréquente de messages brefs). Le réseau qui monte, puisqu'il a connu une croissance annuelle de 1382% !

Nous constatons que les réseaux sociaux sont de plus en plus présents dans tous les secteurs d'activité et depuis quelques années, ils ont fait leur apparition dans le secteur de l'éducation et de l'apprentissage.

Le paradigme de l'apprentissage évolue en s'enrichissant de nouvelles configurations avec l'usage généralisé des réseaux sociaux, l'apprenant est en position de diversifier ses sources d'information et de faire appel à ses pairs outre les échanges avec les enseignants s'ils sont aussi en ligne. Ceux-ci ne sont plus les seuls intervenants capables de valider les procédures d'acquisition de savoir comme c'est le cas dans la salle de classe présentielle. De nouveaux acteurs interviennent entre l'apprenant et l'enseignant, à savoir les pairs et les experts du domaine, choisis souvent par l'apprenant et sa communauté. Par ailleurs, le réseau social met les élèves ainsi que les anciens élèves en contact. Ce qui permet de l'entraide, du partage des expériences vécues au sein de la classe et une discussion de différents travaux et bien plus encore.

Il est certain que ceux-ci représentent une manière alternative d'échanger des informations. Également, ils peuvent servir à agrémenter la communication de façon originale entre les membres d'une même plateforme. Comme par exemple, si un élève souhaite recevoir une aide ou un éclaircissement concernant certains points du cours dans la plateforme, il pourrait venir la poster sur un réseau social de son choix, de cette façon, il recevrait rapidement des explications par d'autres élèves, voire même par l'enseignant. C'est par l'intermédiaire de ce message, utilisé sous forme de trace, que le tuteur (formateur) arrive à suivre l'interaction et l'évaluation des apprenants sur la plateforme.

Conclusion :

Dans ce chapitre, nous avons fait une présentation d' E-Learning de manière générale, en explicitant ses avantages et ses inconvénients; les plateformes d'apprentissage ainsi que les différents moyens de communication (forums, blogs ...) et leurs importance dans le parcours d'un apprenant.

Rappelons que l'objectif de notre travail consiste essentiellement en la modélisation de l'apprenant afin de détecter les difficultés rencontrées pendant son apprentissage. En effet pour avoir des informations fines de toutes les activités qu'il effectue, nous allons utiliser ses traces laissées pendant son apprentissage. C'est ce que sera étudié en détail dans le chapitre suivant.

CHAPITRE II : TRACE

Introduction :

Ce mot « trace », souvent utilisé dans notre langage courant, semble être une notion bien connue. Or une simple recherche de sa définition rend compte de l'ambiguïté de cette notion. Cette dernière peut avoir différentes interprétations, liées principalement au domaine d'utilisation (mathématiques, histoire, chimie, etc.).

Dans la suite de ce chapitre, nous présenterons les différents types de traces, le modèle de trace ainsi que le système à base de trace notamment l'architecture présentée par "SETTOUTI". C'est ici qu'on abordera les différents formats de représentation de trace (entre autres, fichiers Log).

Et au final, une conclusion en fera le récapitulatif.

II-1. A la recherche d'une définition :

Si on s'intéresse au domaine informatique, on remarque que les traces ont leur particularité qu'elles soient numériques, et que l'étude des traces laissées par les utilisateurs des systèmes informatiques devient de plus en plus importante. Cela est dû au fait que les traces peuvent servir, non seulement à analyser l'activité de ces utilisateurs, mais aussi à leur fournir une aide automatique facilitant leur travail.

Dans le cadre de EIAH (Environnement Informatique pour l'Apprentissage Humain), Pernin [5] définit une trace comme un indice de l'activité des acteurs d'une situation d'apprentissage; il complète, par ailleurs, sa définition en précisant qu'il s'agit d'un résultat obtenu au cours ou au terme d'une activité, d'un événement ou d'un ensemble d'évènements relatifs au déroulement de la situation d'apprentissage.

Dans le même sens, Jermann [6] décrit une trace comme une observation ou un enregistrement de l'interaction de l'apprenant avec un système en vue d'une analyse.

Dans une optique légèrement différente, Champin [7] parle d'une séquence temporelle d'états et de transitions représentant l'activité de l'utilisateur : « *la séquence temporelle des objets et opérations mobilisés par l'utilisateur lorsqu'il utilise le système est appelée trace d'utilisation* ».

Notant que ces dernières sont "cachées" à celui qui les produit, (Krämer, 2012) déclare : « On ne fabrique pas une trace, on la laisse, et ce sans intention aucune. [...] À la différence du signe que nous créons, la signification d'une trace existe au-delà de l'intention de celui qui la génère. C'est justement ce qui échappe à notre attention, à notre contrôle ou à notre vigilance qui, à partir de nos actes, prend la forme d'une trace » [8]. Ainsi il est presque impossible pour l'utilisateur d'interagir avec ces propres traces.

Définition retenue :

En se basant sur ces différentes déclarations, on peut définir une trace comme une suite temporelle située d'observés, c'est un enregistrement d'éléments d'interaction entre un utilisateur et son environnement dans le cadre d'une activité donnée.

La trace relève soit d'une communication entre humains interconnectés de diverses façons par ordinateur, soit d'un ensemble d'actions et de réactions séquentielles entre un humain et une machine. Elle est numérique puisque il s'agit de sauvegarde d'actions effectué par ordinateur. Qu'il s'agisse d'une trace involontaire ou d'une marque volontaire, elle représente un indice qui montre quelque chose à propos des objets connectés.

La trace permet d'accéder à l'objet qui l'a produit et de reconstruire les processus, auquel cas elle est dynamique, tout comme Ginsberg (1989) l'affirme, la trace est un point d'entrée au passé que l'on ne peut interroger que par elle.

Dans les EIAH comme dans les autres environnements informatiques, les traces d'interaction constituent donc une source d'information sur l'expérience de l'utilisation.

II-2. Types de traces :

Nous distinguons plusieurs types de traces et cela dépend de la façon dont les outils d'enregistrement rendent observable l'interaction humaine. En effet, les traces peuvent être constituées de sources multiples et multimodales (enregistrements informatiques y compris des captures d'écrans, audio et vidéos d'êtres humains en interaction avec une machine ou entre eux via un ordinateur).

Pour la suite, nous donnons quelques définitions :

- Un **observé** est une unité interactionnelle qui compose la trace. Il contient tous les éléments de contexte de l'action effectuée qu'il représente : l'acteur qui la réalise, le type d'action réalisée, le ou les entités impactées, etc.

- Le **modèle de trace** est l'ensemble des concepts et relations types qui décrit les éléments (acteurs, entités et actions) qui peuvent être contenus dans la trace, et comment ils s'organisent dans la trace.

- La **trace première** est la trace directement produite par le système de collecte et décrite dans le langage du concepteur. Toutes les informations sur l'expérience tracée utilisées dans notre système de visualisation proviennent de cette trace première ou bien sont déduites des informations de la trace première.

- **Trace métier**, ce terme est utilisé par (Georgeon *et al.*, 2006) pour définir une trace qui représente l'expérience d'une activité donnée et qui est décrite dans le contexte de cette activité. Elle contient donc un vocabulaire, et une sémantique compréhensible par l'utilisateur qui verra cette trace.

Selon l'utilisateur à l'origine des traces, et selon "l'espion" qui les perçoit, les traces peuvent être de deux natures. Il peut en effet s'agir de trace dont "le partenaire" est à l'origine, cela veut dire que l'utilisateur du système en question laisse des traces volontairement, on peut ainsi les appeler marques, indices, repère ...

Comme il peut également s'agir de traces laissées involontairement par l'utilisateur du système, c'est des inscriptions d'interactions de l'utilisateur avec le dispositif ou avec son

correspondant via l'ordinateur, cela se fait de manière "cachée", on les qualifie donc comme étant des **traces propres**, non contrôlées.

Nous pouvons aussi identifier des propriétés des traces en fonction de l'utilisation qui y en est faite. Il s'agit dans ce cas davantage de caractéristiques de l'utilisation des traces plutôt que de propriétés qui leurs sont propres. Nous avons repéré deux utilisations distinctes des traces. Premièrement, la participante peut « simplement » consulter les traces en cours d'activité. Il s'agit par exemple de lecture de l'historique du *chat* publié, que les analystes repèrent par des retours d'ascenseur dans cet espace. Dans ce cas, nous qualifions cette utilisation des traces de «utilisation consultative».

Deuxièmement, la participante peut effectuer des opérations sur les traces. Ces opérations peuvent être du copiage, du collage des traces. Dans ce cas, nous qualifions cette utilisation des traces de «utilisation opératoire».

Pierre Deransart, dans son modèle descriptif considère deux types de traces : la trace *effective* qui correspond à la séquence des états qui ont été *effectivement* observés (ou même une partie d'un état, c'est-à-dire un sous-ensemble des variables d'état).

La trace *virtuelle* qui est une description complète de ce que pourrait être la trace si elle était observable totalement (par exemple l'automate équivalent à la trace). La trace virtuelle n'est naturellement pas toujours disponible.

De plus, P.Deransart propose deux fonctions permettant de passer d'une trace à l'autre :

L'*extraction* permettant de construire une trace effective; et la *reconstruction* permettant de synthétiser la trace virtuelle à partir de traces effectives. Le cas limite est quand la trace effective contient *toute* l'information permettant de construire la trace virtuelle.

II-3. Modèle de trace :

Le modèle de trace est le vocabulaire de la trace qui décrit de manière abstraite les objets qui en font partie, c'est cette description formelle de la structure et de son contenu. Il ouvre donc la possibilité de définir des mécanismes d'inférence puissants et réutilisables sur les traces modélisées.

Dans un contexte purement informatique, on considère un modèle comme étant un ensemble de classes (classes de la programmation orientée objet) donnant une description aux observés de la trace.

Le modèle de trace fournit donc des informations sur les propriétés de cette dernière et sur les éléments et les relations qu'elle contient. Ces éléments sont typés et leurs types sont définis formellement dans le modèle, chaque type est caractérisé par un nom et un ensemble de propriétés souvent représenté sous la forme ATTRIBUT-VALEUR incluant au minimum une notion de temps servant à situer temporellement les éléments de ce type. De plus, le modèle est capable d'être modifié indépendamment du système observé pour s'adapter au regard d'un observateur donné.

Ainsi, On peut dire que l'association d'une trace à un modèle garantit la possibilité de l'interprétation critique d'une trace.

II-4. Système à base de trace :

Le système à base de traces représente les traitements et les processus structurés appliqués à la donnée (trace) depuis sa collecte jusqu'à son application, cela en utilisant des outils adéquats tout en respectant une politique de gestion définie en fonction de la nature des risques que l'utilisateur tracé peut rencontrer. Ce système agit sur tout types d'environnements tracés en particulier les EIAH (environnement informatique pour l'apprentissage humain).

Nous décrivons ici les différents composants du système à base de traces ainsi que les démarches à suivre pour une gestion de qualité.

4.1. Les démarches à suivre pour gérer une trace :

Tout tracer est possible, mais la surabondance de données devient très vite inexploitable.

De plus, les performances et le volume des traces restent des critères importants à prendre en compte. On recherchera donc les événements significatifs.

Le terme « traitement » des traces, au sens de la Commission nationale de l'informatique et des libertés (CNIL), recouvre les 5 phases suivantes :

4.1.1. Phase 1 : La collecte

La phase initiale de tout projet de gestion des traces débute par l'identification des besoins. Avant d'être exploitées dans un SBT, les traces sont d'abord collectées à l'aide du système de collecte.

Le système de collecte est un ensemble structuré de processus permettant de convertir les données en traces. Nous pouvons dire aussi, que la collecte est le processus qui consiste à exploiter d'une façon automatique ou manuelle un ensemble de sources de traçage (fichier ou flux d'information structurés dans un format explicite quelconque XML, texte, vidéo, audio) afin d'obtenir une trace et cela nécessite souvent plusieurs retours aux dites sources pour son amélioration. En effet, le processus de collecte des traces ne peut être bénéfique que s'il est utile, légal, loyal et gérable dans la durée.

Voici un passage du [*Forum des compétences de la sécurité des systèmes informatiques dans son travail sur la gestion des traces générées par les systèmes d'information*] où les conditions de collecte sont présentées :

"Au moment de la conception d'une politique de gestion des traces, 4 questions sont à poser :

** Est-ce **utile** à quelqu'un, est-ce pertinent pour le besoin ?*

** Est-ce **légal** ?*

Des choses simples peuvent enfreindre la Loi, comme ouvrir un répertoire intitulé «personnel» pour un administrateur système ou messagerie.

** Est-ce **loyal** ?*

La CNIL insiste beaucoup sur ce point et les tribunaux aussi. Les usagers doivent être informés des moyens de surveillance qui ont été mis en place sur le système.

** Est-ce **gérable** dans la durée ?*

Le cycle de vie complet des traces, et non pas la simple collecte, sera-t-il assuré sans remise en cause profonde de la capacité des machines, de l'exploitation ou de l'organisation. "

Ces traces résultantes du processus de la collecte sont donc constituées d'une collection d'observés temporellement situés associée à un modèle de trace tel que nous l'avons décrit

précédemment, elles sont appelées **traces premières** du SBT car c'est les premières à être manipulables dans ce système à l'issue de la collecte.

On les appelle aussi « données brutes », c'est ces fichiers de collecte qui contiennent les enregistrements d'événements produits directement par les ressources informatiques. Ces données contiennent également tous les éléments permettant de les dater, de retrouver leur environnement, etc.

La figure ci-dessous décrit de manière générale le fonctionnement d'un système de collecte.

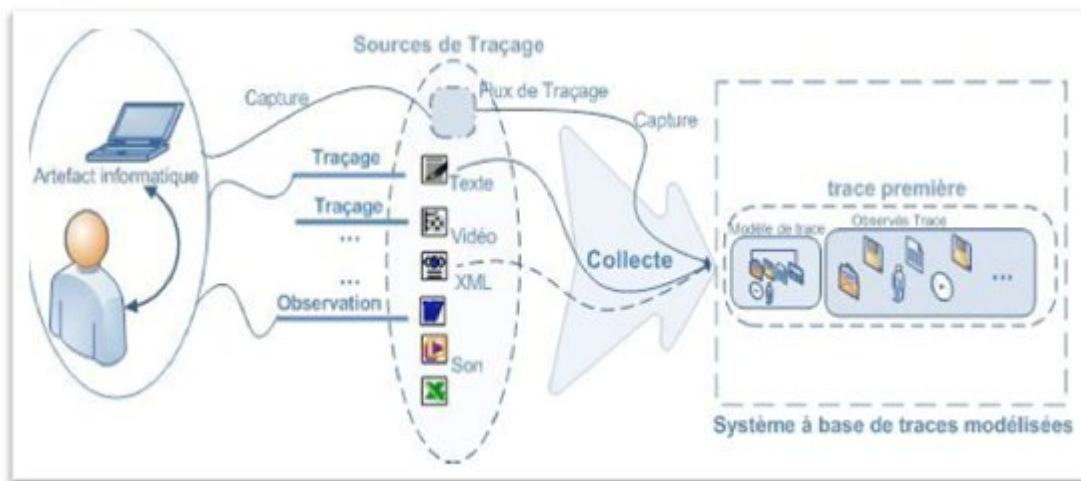


Figure 3 : Fonctionnement du système de collecte

Le SBT stocke les traces collectées dans **la base de traces**, celle-ci regroupe l'ensemble des traces du SBT et de leurs modèles respectifs. Elle permet le stockage permanent et l'accès à tout moment aux traces. Elle représente la base de connaissances du SBT.

4.1.2. Phase 2 : Traitement

La sécurité et l'intégrité des fichiers de traces doivent être assurées. Par ailleurs, les outils de traitement des traces doivent répondre aux besoins des équipes d'exploitation, notamment :

- offrir des possibilités de traces riches et pertinentes;
- configurables en fonction des objectifs de sécurité retenus qui détermineront les événements significatifs ;
- permettre un format lisible et compréhensible par l'homme ;

En effet, le système de transformation permet l'ajout, l'édition et la suppression des traces. Il regroupe les processus de conservation des traces et des droits d'administration de ces dernières.

Les traces premières d'une base de traces sont les seules traces non transformées du système, cependant, elles peuvent être transformées selon deux manières : la transformation manuelle qui est une simple mise à jour réalisée par l'utilisateur interagissant avec sa trace. Ce type de transformation consiste à changer les éléments de la trace en ajoutant, supprimant, ou modifiant ses observés.

Et la transformation automatique qui s'applique dans un SBT moyennant des modèles de transformations respectant un modèle de transformation.

[Touch, 2006] définit le modèle de transformation comme un ensemble de règles et de contraintes, cet ensemble décrit comment doit s'exécuter la transformation.

Voici les étapes de transformation de traces comme elles sont présentées dans le travail de [Nabila Bousbia, *Traces de navigation des apprenants dans un environnement de formation sur le Web/ Laboratoire de Paris 6 (LIP6)*]

- **Nettoyage** : c'est la première étape du processus de traitement. Elle consiste à éliminer le bruit : pages non trouvées (erreur 403, 404, etc.) ; programmes à écarter du traitement des traces (l'outil de collecte, les programmes de messagerie instantanée, etc.) ; pages intermédiaires de la plateforme, telles que les pages de traitement du formulaire de connexion.
- **Filtrage** : il s'agit de filtrer les pages pour ne garder que celles ayant une URI valide (respectant le format standard des URIs).
- **Transformation** : dans cette dernière étape, si le fichier *log* en entrée ne correspond pas au modèle de trace proposé, il subit des transformations, c'est à dire, transformer les traces brutes en un produit bien structuré suivant un modèle défini.

4.1.3. Phase 3 : Exploitation

Tous les acteurs ayant des accréditations pour accéder aux informations contenues dans les traces doivent connaître les limites de leur pouvoir d'investigation. A cet égard, un contrôle doit être fait sur les utilisateurs ayant accès aux fichiers de traces.

Les experts en informatique et en sécurité ne doivent pas croiser les traces sans que cela soit explicitement prévu dans les finalités spécifiées dans les déclarations à la CNIL, et ne doivent pas mettre en cause les personnes.

Seuls les responsables hiérarchiques peuvent incriminer les personnes, mais doivent veiller à sauvegarder les droits de la défense.

4.1.4. Phase 4 : Conservation

Les traces sont conservées en ligne et hors ligne. Cette conservation doit respecter un certain nombre de mesures qui permettent aux traces de garder leurs valeurs convaincantes, notamment :

- Prévenir la saturation des capacités de stockage en ligne;
- Protéger l'accès aux données en mettant en œuvre des règles d'accès et des contrôles ;

Les moyens mis en œuvre pour conserver les traces et la durée de conservation doivent être précisés.

4.1.5. Phase 5 : Destruction

L'effacement des données de traces, y compris des copies s'il y en a, doit se faire à bonne date (de façon automatisée ou par un planning), au-delà du délai de purge, aucun nouveau contentieux ne peut être ouvert.

Ces mesures permettent une gestion optimale des capacités de stockage et d'être en conformité avec la Loi.

Une fois la collecte de traces faite, avec ou sans transformation, les récoltes acquises (traces premières ou traces après transformation) doivent être visualisées, c'est justement la tâche du système de visualisation.

Le **système de visualisation** permet de visualiser les traces et facilite ainsi leur analyse et leur interprétation. Ce système doit gérer et accéder aux différentes applications associées aux sources de traçage et Il donne l'accès direct depuis la trace à la source de traçage pour interroger celle-ci plus directement. Il peut être intégré dans un système de transformation à un niveau quelconque.

4.2. Qualités d'une trace

Les qualités d'une trace sont variées, elles reposent sur différents critères :

- **contextuelle** : C'est-à-dire qu'elle doit permettre de connaître l'environnement de la collecte (le système hôte, la version des logiciels, etc).
- **fidèle** : D'éventuelles interruptions du traçage doivent être portées à la connaissance de la personne qui effectue l'analyse (critère de *disponibilité* des données).
- **fiable** : C'est-à-dire qu'elle produit les mêmes enregistrements si l'événement se reproduit dans les mêmes conditions.
- **signifiante** : L'événement doit être explicable, il doit être interprétable par l'utilisateur.
- **exploitable** : Les délais de traitement doivent être compatibles avec les besoins d'exploitation des traces. Les moyens mis en œuvre ne doivent pas être excessifs ou démesurément consommateurs de ressources.

- **contrôle d'accès** : Une trace doit faire l'objet d'une confidentialité et d'intégrité des données.

Nous présentons ici un exemple de système à base de trace.

4.3. Exemple d'un système à base de trace :

Settoui [9] définit le Système à Base de Traces comme un système informatique permettant et facilitant l'exploitation des traces. Il est composé de différents composants comme le montre la figure suivante.

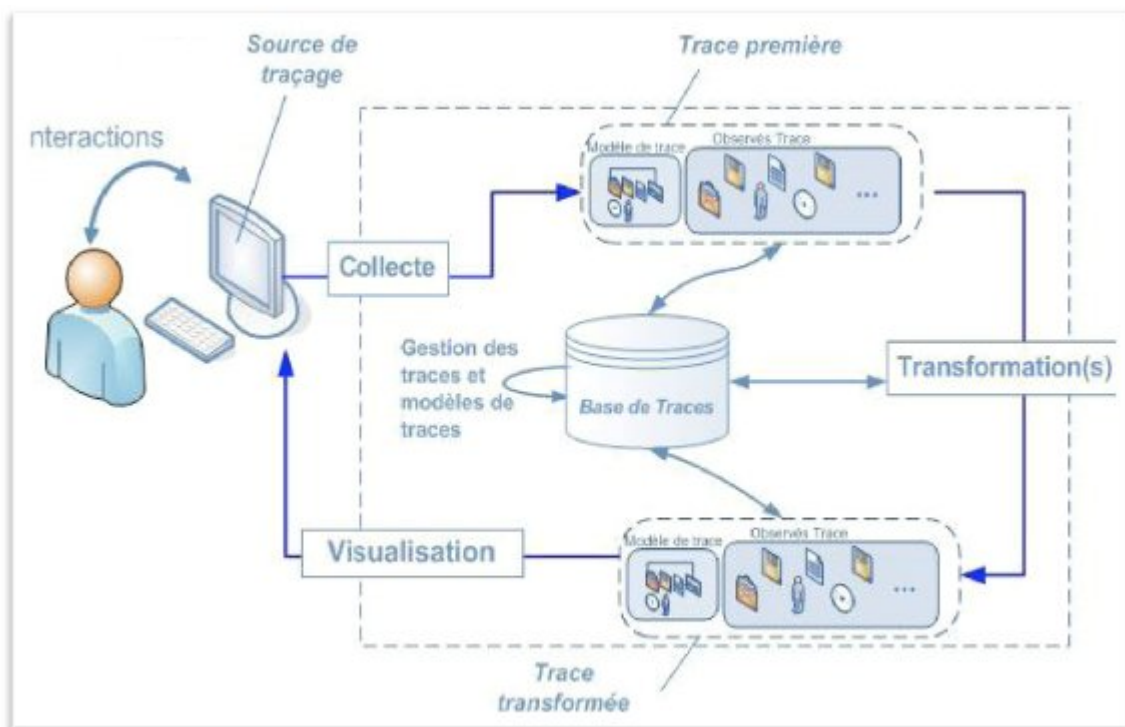


Figure 4 : système à base de traces, cas [Settoui, 2005]

II-5. Formes de traces :

Comme cité précédemment, les traces d'interaction peuvent être produites suite à l'instrumentation du système informatique, afin de pouvoir mémoriser les actions de l'utilisateur effectuant une activité particulière. Une même trace d'activité peut être informatiquement représentée différemment. Nous nous focalisons ici sur cette diversité de formes de traces et nous décrivons quelques formats.

5.1. Représentation des traces en BDD :

Nous pouvons représenter les traces d'utilisation dans une base de données. La base de traces peut alors être une base de données relationnelle ou objet qui permet de stocker et manipuler les traces facilement.

Dans le travail de Arana[10], MySQL a été choisi pour stocker les traces d'utilisation du système. Ce type de stockage est utilisé par la plupart des logiciels pour les plates-formes de la formation à distance comme MOODLE.

id	time	userid	ip	course	module	cmid	action	url	info
1	1141739813	2	127.0.0.1	1	user	0	update	view.php?id=2&course=1	
2	1141739816	2	127.0.0.1	1	course	0	view	view.php?id=1	1
3	1141743318	2	127.0.0.1	1	course	0	new	view.php?id=2	Photoshop Competency (ID 2)
4	1141744002	2	127.0.0.1	2	user	0	view	view.php?id=2&course=2	2

Figure 5 : Un exemple d'une table de traces dans une base de données

Nous pouvons également représenter une trace d'interaction sous format Log (**Log File**).

5.2. Fichiers Log :

5.2.1. Définition :

Le format le plus simple des traces d'interaction est le fichier Log qui permet de reprendre de manière chorographique l'ensemble des événements qui ont affecté le système informatique et l'ensemble des actions qui ont résulté de ces événements.

Les traces sont stockées dans des fichiers journaux (log file), toutes les opérations de l'utilisateur sur le système sont sauvegardées ligne par ligne en respectant un modèle de trace bien défini. La journalisation consiste à garder les traces sur un support sûr des événements survenus dans un système ou dans une application.

Un ou plusieurs fichiers de log au format prédéfini sont générés en cours d'exécution et conservent des messages informant sur la date et l'heure de l'évènement, la nature de l'évènement et sa gravité par un code ou une description sémantique, éventuellement d'autres informations : utilisateur, classe, etc.

La figure ci-après représente un exemple de fichier Log :

```

1 #Software: Microsoft Internet Information Services X.X-
2 #Version: X-
3 #Date: 2010-03-24 07:00:01-
4 #Fields: date time s-sitename s-computername s-ip cs-method cs-uri-stem cs-uri-query s-port cs
5 2010-03-24 07:00:01 ZZZZC941948879 RUFFLES 222.222.222.222 GET / - 80 - 220.181.7.113 HTTP/1.1
6 2010-03-24 07:00:23 ZZZZC941948879 RUFFLES 222.222.222.222 GET /2009/12/im_not_mean_im_just_ar
7 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-blank.gif - 80 - 217.
8 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /grep-options.gif - 80 - 217.2
9 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-cat.gif - 80 - 217.2
10 2010-03-24 07:00:32 ZZZZC941948879 RUFFLES 222.222.222.222 GET /terminal-pwd-cd.gif - 80 - 217
11 2010-03-24 07:00:39 ZZZZC941948879 RUFFLES 222.222.222.222 GET /robots.txt - 80 - 95.55.207.95
12 2010-03-24 07:00:39 ZZZZC941948879 RUFFLES 222.222.222.222 GET /rss-short.xml - 80 - 173.45.2
13 2010-03-24 07:00:43 ZZZZC941948879 RUFFLES 222.222.222.222 GET /2009/08/22-things-you-dont-knc
14 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /screen.css - 80 - 98.88.35.13
15 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/rss-header-red.gif - 80 -
16 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/logo.jpg - 80 - 98.88.35.1
17 2010-03-24 07:00:44 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/input-emailsend.jpg - 80 -
18 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /images/cm-ebook-banner.gif - 8
19 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/bg.jpg - 80 - 98.88.35.13
20 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/bg-top.jpg - 80 - 98.88.35
21 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /21things/checkout-login.gif -
22 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /img/topnav-contact.jpg - 80 -
23 2010-03-24 07:00:45 ZZZZC941948879 RUFFLES 222.222.222.222 GET /21things/portal-email-sub.gif

```

Figure 6 : Exemple de fichier Log

5.2.2. Utilisation :

Comme [Harrer et Al, 2003] utilisent par exemple un fichier log texte pour stocker toutes les actions des apprenants dans un environnement d'apprentissage supporté par ordinateur. Les journaux peuvent aussi être utilement utilisés par :

- * un administrateur, afin de produire des statistiques sur l'utilisation d'un système (par exemple les logs du serveur Web Apache).
- * Un développeur, afin de détecter des défaillances et de corriger les bogues qui en sont responsables. Il est plus facile de repérer la source d'une défaillance si le journal est dense en informations (fonctions appelées, valeurs des paramètres passés...).
- * Un développeur, pour éviter de polluer son code avec des `println ()`.
- * Un utilisateur, qui peut utiliser un journal afin de revenir sur un crash et refaire les opérations qui auraient été perdues (transactions).

5.2.3. Limite des fichiers log :

Il est souvent argumenté que les fichiers logs constituent un bon support d'analyse de l'expérience d'utilisation d'un système [Roussel & al, 2006]. Ce pendant, réutiliser les logs mène à la confrontation de quelques problématiques et contraintes, citons parmi elles :

- Le manque de contexte des informations qu'ils véhiculent.
- Grand volume de données à traiter par rapport au besoin.
- Manque de structure dans ces données.

C'est pourquoi toute analyse à base de logs doit s'accompagner d'un effort de restructuration et de remise en contexte des informations dans l'activité. De plus, les fichiers logs sont bien souvent des données accessibles depuis le serveur, ce qui ne permet pas de les réutiliser en local et en temps réel.

L'utilisation des fichiers log sous forme de texte pur (TXT par exemple) pour stocker les traces d'utilisations manque de flexibilité et rend les analyses et manipulations un peu difficiles. C'est pourquoi les systèmes informatiques récents utilisent les fichiers textes semi structurés au format XML ainsi que les ontologies pour palier à ce manque.

5.3. Format XML :

5.3.1. Définition :

XML, acronyme de eXtensible Markup Language (qui signifie: langage de balisage extensible), est un langage informatique qui sert à enregistrer des données textuelles. Maintenant très populaire, Il permet de faciliter l'échange d'information sur l'internet. C'est un standard fortement utilisé pour structurer des contenus. Un document XML à la forme suivante:

```
<?xml version="1.0" encoding="utf-8" ?>
<xml>
  <balise attribut="valeur" attribut="valeur.."
    ... contenu éventuel
  </balise>
</xml>
```

Figure 7 : Forme d'un document XML

Avantagé au HTML, XML s'intéresse non seulement à la présentation, mais aussi à la description du contenu. Il offre la possibilité de créer de nouvelles balises ce qui le rend beaucoup plus dynamique, ces balises peuvent avoir aussi des attributs, et cela fait penser aux langages de programmation à objets.

Ce qui motive le choix de XML, c'est son aptitude de sérialisation et son offre en termes d'outils d'interrogation, de transformation et de repérage. Nous présentons ci dessous un exemple de réutilisation de l'information pour former un fichier XML concernant les traces.

5.3.2. Utilisation et réutilisation de fichiers XML :

Structurer la trace dans un format XML s'il s'agit de texte : les lignes du texte deviennent des balises <ligne> XML dont les fils sont des balises <cellule> contenant chacune un des items de la ligne texte en question.

Restructurer la trace dans un format XML s'il s'agit d'une source de BDD (table ou requête) :

Les lignes de la source de BDD deviennent des balises <ligne> XML dont les attributs et leurs valeurs sont les noms des champs avec les valeurs de la ligne de source BDD en question.

Beaucoup d'autres formalismes peuvent s'ajouter à cette structure de base, mais celle-ci est suffisante pour la plupart des documents.

La figure suivante est un exemple de sortie XML lors de la formalisation d'une trace d'interaction :

```
<?xmlversion="1.0"encoding="windows-1252"standalone="no"?>
<!DOCTYPE log SYSTEM "logger.dtd">
<log>
<record>
<date>2004-02-04T23:24:19</date>
<millis>1075933459913</millis>
<sequence>0</sequence>
<logger>myExample</logger>
<level>INFO</level>
<class>myExample</class>
<method>main</method>
<thread>10</thread>
<message>myinfo</message>
</record>
</log>
```

Figure 8: Exemple d'une trace d'interaction sous format XML

Ci dessous, description des noms de balises apparus dans la figure :

Tag	Description
Log	Racine du document XML
Record	Enregistrement (ligne) du journal
Date	Date et heure de journalisation
Millis	Timestamp UNIX de la date et heure
Sequence	Numéro d'ordre du fichier stockant le journal si cycle
Logger	Nom du système de log ayant réaliser la journalisation
Level	Niveau de criticité du message journalisé
Class	Nom de la classe
Method	Nom de la méthode
Thread	Numéro du processus
Message	Texte du message

Tableau 1 : Description des noms des balises

5.3.3 Quelques avantages du XML :

Lisibilité : il est facile pour un humain de lire un fichier XML car le code est structuré et facile à comprendre. En principe, il est même possible de dire qu'aucune connaissance spécifique n'est nécessaire pour comprendre les données comprises à l'intérieur d'un document XML.

Disponibilité : ce langage est libre et un fichier XML peut être créé à partir d'un simple logiciel de traitement de texte (un simple bloc-note suffit).

Interopérabilité : Quelques soit le système d'exploitation ou les autres technologies, il n'y a pas de problème particulier pour lire ce langage.

Extensibilité : De nouvelles balises peuvent être ajoutée à souhait.

Conclusion

Ayant parcouru ce chapitre, la notion de systèmes à base de traces et son exploitation dans le cadre des EIAH ont été traitée. Dans un premier temps, nous avons défini les différents moyens d'apprentissage mis en disposition de l'apprenant sur Internet, et aussi le concept de trace et de systèmes à base de trace. Maintenant que nous savons exploiter les traces laissées par l'apprenant durant son apprentissage, nous allons élaborer une ontologie du domaine qui

modélisera le cours en concepts pour faciliter la comparaison entre les messages édités par l'apprenant et le contenu du cours.

Partie 2 : Ontologies et **Indexation**

Chapitre III : Ontologie

Introduction

La notion d'ontologie intéresse à la fois l'ingénierie des connaissances, la linguistique et la philosophie. Initialement, l'Ontologie était un domaine de la philosophie concernant « l'étude de l'être en tant qu'être, c'est-à-dire l'étude des propriétés générales de ce qui existe ». Or, progressivement les objets produits par cette discipline ont pris le nom d'ontologies.

Depuis une vingtaine d'années, un sens nouveau est apparu. Il s'agit de celui utilisé en informatique. Cette notion est plus proche du point de vue de l'intelligence artificielle. Dans notre étude, nous utilisons les ontologies pour représenter un domaine sous forme de base de connaissance.

Nous présentons dans ce chapitre la théorie de l'ontologie en se basant sur sa définition et celle des éléments qui la constituent; sur sa typologie; sa représentation ainsi qu'aux langages informatiques qui s'appliquent sur cette dernière.

III-1. Définition :

1.1.1 Définitions issues de la philosophie :

Dans le domaine de la philosophie, l'ontologie est considérée comme une branche de la métaphysique qui s'intéresse à la nature et l'organisation de la réalité. Cette signification est celle de « *science de ce qui existe* » dans laquelle on ne cherche pas à expliquer le monde, mais à le représenter.

Cette définition philosophique est tirée de la définition issue du dictionnaire de l'Académie française.

" *ONTOLOGIE. Emprunté du latin scientifique *Ontologie*, lui-même composé de *Onto-*, tiré du grec *ôn*, *ontos*, « étant, ce qui est », et *Logia-*, tiré du grec *logos*, « discours, traité ». Partie de la philosophie qui a pour objet l'être en tant qu'être, qui étudie les propriétés générales de l'être".*

1.1.2 Définitions issues de l'intelligence artificielle

De nombreux chercheurs ont proposé des définitions qu'il est utile d'examiner :

- Une définition générale a été donnée par Thomas R. Gruber [11] où il décrit une ontologie comme une spécification explicite d'une conceptualisation modélisant des concepts et les relations entre concepts.
- De son côté, [Bachimont, 2004] définit une ontologie comme une description rigoureuse et structurée du vocabulaire, au sens de la logique, d'un domaine spécialisé.

• Plus clair encore, [Christophe Roche] a donné cette définition : *"Une ontologie est une conceptualisation d'un domaine à laquelle sont associés un ou plusieurs vocabulaires de termes. Les concepts se structurent en un système et participent à la signification des termes. Une ontologie est définie pour un objectif donné et exprime un point de vue partagé par une communauté. Une ontologie s'exprime dans un langage (représentation) qui repose sur une théorie (sémantique) qui garantit des propriétés de l'ontologie en termes de consensus, cohérence, réutilisation et partage"* .

Nous retenons donc cette dernière définition car elle englobe et résume les définitions précédentes.

Une ontologie est vue comme un ensemble de concepts permettant de modéliser un ensemble de connaissances dans un domaine donné. Un concept peut présenter plusieurs sens thématiques. Les concepts sont liés entre eux par des relations (de composition et d'héritage).

III-2. Composants de l'ontologie :

Les ontologies sont, à l'heure actuelle, au cœur des travaux menés en ingénierie des connaissances. Elles permettent de représenter les connaissances et les manipuler automatiquement, tout en gardant leur sémantique. Comme [Charlet, 2003] *déclare, en ingénierie des connaissances, L'ontologie est constituée d'un ensemble de concepts structuré par des relations sémantiques (décrivant les propriétés des concepts) parmi lesquelles la relation « est-un » joue le rôle privilégié de relation d'héritage de propriétés entre les concepts.*

Nous présentons ici les différents composants d'une ontologie.

2.1 Concepts :

Ce sont des notions (ou objets) permettant la description d'une tâche, d'une fonction, d'une action, d'une stratégie ou d'un processus de raisonnement, etc. Ils peuvent être abstraits ou concrets, élémentaires ou composés. Habituellement, les concepts sont organisés en taxonomie. Une taxonomie est une hiérarchie de concepts (ou d'objets) reliés entre eux en fonction de critères sémantiques particuliers.

2.2 Relations :

Une relation est définie comme une notion de lien entre des entités, exprimée souvent par un terme ou par un symbole littéral ou autre. Elle organise les concepts d'un domaine de façon à représenter un type d'interaction entre eux.

Des exemples de relations binaires (entre deux) sont : *sous-concept-de*, *connecté-à*, *sorte-de*, etc.

2.3 Fonctions :

Elles constituent des cas particuliers de relations, dans laquelle un élément de la relation, (le nième) est défini en fonction des N-1 éléments précédents.

2.4 Axiomes :

Les axiomes constituent des assertions, acceptées comme vraies, à propos des abstractions du domaine traduites par l'ontologie. Ils permettent de définir la sémantique des termes (classes, relations), leurs propriétés et toutes contraintes concernant leur interprétation.

Ils sont définis à l'aide de formules bien formées de la logique du premier ordre en utilisant les prédicats de l'ontologie.

2.5 Instance :

Les instances constituent la définition extensionnelle de l'ontologie ; ces objets véhiculent les connaissances à propos du domaine du problème, elle sont utilisées pour représenter des éléments.

III-3. Typologie des ontologies

Les ontologies peuvent être classifiées selon plusieurs critères :

3.1. Typologie selon l'objet de conceptualisation :

Cette typologie dispose des types suivants :

* **3.1.1. Les ontologies de représentation des connaissances :** regroupent les concepts impliqués dans la formalisation des connaissances. Ca intègre les primitives de représentation des langages à base de classes, instances, facettes, propriétés, relations, restrictions, valeurs etc.

* **3.1.2. Les ontologies supérieures ou de Haut niveau :** visent à étudier les catégories des choses qui existent dans le monde, comme les concepts de haut niveau d'abstraction tels que les entités, les événements, les états, les processus, le temps, l'espace et les relations.

Les concepts de l'ontologie de haut niveau sont indépendants d'un domaine ou d'un problème particulier.

* **3.1.3. Les ontologies génériques** : sont appelées, également, des méta-ontologies. Elles décrivent des concepts génériques moins abstraits que ceux décrits par des ontologies supérieures. ces dernières peuvent être réutilisées à travers différents domaines, cela en adressant des connaissances visant à résoudre des problèmes génériques.

* **3.1.4. Les ontologies de domaine** : décrivent le vocabulaire lié à des domaines particuliers (comme la physique, la chimie, la médecine et la modélisation d'entreprise) ainsi que les relations s'appliquant sur ces derniers.

* **3.1.5. Les ontologies d'applications** : dépendent, selon [Guarino], à la fois d'un domaine particulier et d'une tâche spécifique, Elles ont un domaine de validité restreint et correspondent à l'exécution d'un ensemble de tâches composant l'application. Généralement, les ontologies d'application ne sont pas réutilisables et possèdent donc un intérêt plus limité. Dans notre cas d'étude, nous allons nous baser sur l'ontologie du domaine du cours enseigné. Cette ontologie sera le pont de correspondance entre les messages édités sur les forums d'apprentissage en ligne et le contenu du cours.

3.2. Typologie selon le niveau de détail

Dans cette typologie, deux cas ont été distingués :

Quand les ontologies sont très détaillées au niveau du vocabulaire utilisé, qui est plus riche, on parle de granularité fine. Ce vocabulaire doit assurer la pertinence des concepts d'une tâche spécifique, dans un domaine particulier.

Souvent, les ontologies de domaine et les ontologies d'applications représentent des ontologies à granularité fine.

La granularité large concerne le cas où les ontologies sont moins détaillées. Un exemple est celui des ontologies de haut niveau, car elles disposent de concepts génériques qui peuvent être raffinés dans d'autres types d'ontologies.

3.3. Typologie selon le niveau du formalisme

Comprend quatre catégories :

- *Ontologies informelles* exprimées dans un langage naturel.
- *Ontologies semi-Informelles* décrites à l'aide d'un langage naturel structuré et limité.;
- *Ontologies semi-formelles* spécifiées dans un langage artificiel défini formellement.
- *Ontologies formelles* basées sur un langage artificiel contenant une sémantique formelle.

III-4. Construction des ontologies :

On peut distinguer trois types de méthodes pour la construction d'ontologie : des méthodes **manuelles, automatiques et mixtes**.

Dans la première, les experts créent une nouvelle ontologie d'un domaine ou étendent une ontologie déjà existante.

Dans la méthode automatique, l'ontologie est construite par des techniques d'extraction des connaissances : les concepts et leurs relations sont extraits de bases de connaissances et ensuite vérifiés par les inférences.

Enfin, la méthode mixte, c'est l'association des deux types précédents, les techniques automatiques sont utilisées pour étendre des ontologies qui ont été construites manuellement.

Les étapes suivantes sont considérées comme la base d'un processus standard de construction. Ce processus opère selon quatre étapes fondamentales :

4.1. L'identification de l'objectif : Elle permet d'identifier, en termes généraux, l'objectif, la portée et les limitations de l'ontologie à construire.

4.2. La création de l'ontologie : C'est l'étape la plus longue et la plus difficile, elle contient elle-même trois sous-étapes :

1. L'acquisition des connaissances : Elle sert à définir les concepts dans un domaine donné et les relations entre eux, de manière à ne pas être ambiguës.

2. Le codage : Une fois les concepts et leurs relations acquises, ça permet de représenter l'ontologie dans un langage formel. La formalisation de l'ontologie peut être de différents degrés (informel, semi-informel, semi-formel, formel) [détaillés précédemment].

3. L'intégration des ontologies existantes: C'est l'étape qui permet de réutiliser les concepts déjà définis dans des ontologies existantes.

4.3. L'évaluation de l'ontologie : Suivant [Gruber] sur son point de vu, il a proposé quelques critères pour l'évaluation d'une ontologie :

- *La clarté* : les concepts de l'ontologie doivent présenter le sens voulu des termes ;
- *La cohérence* : les raisonnements construits à partir des axiomes d'une ontologie ne doivent pas aboutir à des contradictions ;

- *L'extensibilité* : l'ontologie doit être conçue de manière à ce qu'une nouvelle utilisation se fasse sans remettre en cause ce qui a été précédemment conçu ;
- *L'engagement ontologique minimal* : l'objectif est de permettre la spécialisation des spécifications d'une ontologie donnée selon des besoins réels ;

4.4. La documentation : permet de renseigner les ontologies, leurs concepts importants ainsi que leurs objectifs.

III-5. Formalisation et outils de représentation d'ontologies :

La formalisation peut être vue comme une représentation explicite et formelle. Elle est réalisée par le biais d'un langage formel (ou formalisme) qui est un ensemble de composants sémantiques, de règles structurelles et d'une notation formelle particulière (forme) destinée à organiser les relations entre les éléments constituant l'ontologie.

L'objectif de l'utilisation d'un langage de formalisation d'ontologies, est de permettre d'une part de réduire les ambiguïtés du langage naturel; et d'une autre part de rendre l'ontologie compréhensible par les machines.

«*Les formalismes offrent un support formel à la composition des concepts et à leur comparaison*» (Barry et al., 2001).

Différents formalismes tels que les logiques de description, les réseaux sémantiques et les frames (schémas) peuvent être employés pour représenter formellement une ontologie.

* *Les logiques de description* : représentent la connaissance sous forme de propositions ou affirmations sur le domaine.

* *Les réseaux sémantiques* : tiennent compte de la structure et les relations entre ces propositions tout en gardant cette approche propositionnelle.

* *Les frames* : représentent le domaine en termes de ses objets et leurs propriétés et relations. Ces paradigmes se différencient les uns des autres par leur formalisme de représentation des connaissances et leurs mécanismes d'inférence permettant de raisonner sur les représentations.

Exemple d'outils de représentation :

* **Ontolingua** : Ontolingua est un serveur d'édition d'ontologies. Il utilise des classes, des relations, des fonctions, des instances et des axiomes pour décrire une ontologie. Une relation peut contenir des propriétés nécessaires (contraintes) ou nécessaires et suffisantes qui la définissent. En plus le serveur Ontolingua offre la possibilité d'intégrer les ontologies Ontolingua, ce qui permet la construction modulaire des ontologies.

* **Protégé 2000** : est une interface modulaire permettant l'édition, la visualisation, le contrôle d'ontologie, l'extraction d'ontologies à partir de sources textuelles, et la fusion semi-automatique d'ontologies.

Le modèle de connaissances sous-jacent à protégé 2000 est issu du modèle des frames et contient des classes, des slots (propriétés) et des facets (valeurs des propriétés et contraintes), ainsi que des instances des classes et des propriétés. Il autorise la définition de méta-classes, dont les instances sont des classes, ce qui permet de créer son propre modèle de connaissances avant de bâtir une ontologie.

III-6. Domaines d'applications des ontologies :

6.1. Système d'information

L'intégration d'une ontologie dans un système d'information vise à réduire, voire éliminer, la confusion conceptuelle et terminologique, et à tendre vers une compréhension partagée pour améliorer la communication, le partage, l'interopérabilité et le degré de réutilisation possible. Cela permet de déclarer formellement un certain nombre de connaissances utilisées pour caractériser les informations gérées par le système, et de se baser sur ces caractérisations et la formalisation de leur signification pour automatiser des tâches de traitement de l'information.

L'ontologie est utilisée différemment, on cite :

* *Enseignement assisté par ordinateur.*

* *Bibliothèque numériques.*

* *Recherche d'informations.*

* *Aide à l'analyse en biologie, suivi médicale informatisé.*

* *Commerce électronique.*

6.2. Web sémantique

C'est un courant prometteur pour l'expansion des systèmes à base d'ontologies. Grâce au Web sémantique, l'ontologie a trouvé un jeu de formalismes standards à l'échelle mondiale, et s'intègre dans de plus en plus d'applications Web, cela se fait au profit des logiciels qui à travers les ontologies et les descriptions qu'elles permettent, peuvent proposer de nouvelles fonctionnalités exploitant les effets d'échelles du Web pour en améliorer les effets.

Nous avons défini la notion d'ontologie, ses formalismes de représentation et ses domaines d'application. Dans ce qui suit, nous aborderons la phase d'indexation qui nous permettra de décomposer les messages laissés dans les forums en termes et concepts que nous allons pondérer par la suite.

III-7. Indexation :

Définition : l'indexation est le processus qui consiste à décrire et à caractériser un document à l'aide de la représentation du contenu de celui-ci. Elle a pour but de construire l'ensemble des termes représentatifs du contenu d'une requête et de chaque document de la collection lors de la recherche. Ces termes, dits termes d'index, jouent un rôle très important dans la recherche d'information puisqu'ils déterminent avec quels mots on peut retrouver un document.

7.1 - Les différents types d'indexation :

Elle peut être manuelle, automatique ou semi-automatique :

- ✓ **L'indexation manuelle :** C'est une opération humaine où chaque document est analysé par un spécialiste ou un documentaliste ;
- ✓ **L'indexation automatique :** C'est la technique la plus utilisée et la plus adaptée aux corpus volumineux. Le processus d'indexation est entièrement informatisé. Aucune intervention humaine n'est requise dans le choix des termes d'index.
- ✓ **L'indexation semi-automatique :** c'est une indexation mixte combinant les deux approches précédentes. Les termes d'index sont d'abord extraits automatiquement, puis le choix final des termes est réalisé par les spécialistes ou documentalistes.

7.2-Processus d'indexation :

Nous utilisons dans notre étude l'indexation automatique. Détaillons cette dernière.

L'indexation automatique est fondée sur l'analyse des documents en vue de l'extraction des termes représentatifs de leur contenu informationnel. Elle repose sur les étapes suivantes :

A. Extraction des termes d'indexation : Il s'agit d'une analyse linguistique du texte du document. Elle comprend les étapes suivantes :

- ✓ *Analyse lexicale* : l'analyse lexicale est un processus qui permet de transformer un texte du document en un ensemble de termes.
- ✓ *Elimination des mots vides* : Les mots représentatifs sont gardés et les mots vides (mots outils de la langue comme les propositions, les articles, les pronoms..., mots trop fréquents, mots rares) sont éliminés.
- ✓ *Normalisation des termes d'index* : basée sur deux approches, la LEMMATISATION et la TRONCATURE.

Lemmatisation : La lemmatisation d'un mot consiste à transformer le mot en son lemme. Le lemme est la racine du mot telle que définie dans la langue naturelle.

Troncature : Construit le radical d'un mot en le coupant à partir d'une position donnée. De cette manière, on réduit les variations morphologiques des mots.

B. La pondération des termes d'index: La pondération consiste à donner un poids d'importance **W** à chaque terme **t** d'un document **d**. Les approches de pondération se basent le plus souvent sur deux facteurs :

- un facteur de *pondération locale* qui exprime l'importance du terme **t** dans le document **d**. Ce facteur est fonction de sa fréquence d'occurrences **tf** dans le document.
- un facteur de *pondération globale* qui exprime le pouvoir de discrimination du terme **t** sur l'ensemble des documents de la collection. Ce facteur est fonction de sa fréquence documentaire inverse **idf** dans tous les documents de la collection.

Plusieurs formules existent et sont basées sur le schéma combiné suivant:

$$W = tf * idf$$

7.3 - Indexation dans la recherche d'informations :

Il existe aussi dans la recherche d'information différentes indexations telles que :

7.3.1 indexation classique : qui se base sur la cooccurrence des mots dans un texte et ne prend pas en considération les liens sémantiques qui peuvent exister entre eux.

7.3.2 indexation sémantique : (Sense Based Indexing) : l'indexation sémantique est une approche d'indexation basée sur le sens des mots [Sanderson, 94][Yarowsky, 93]. Elle s'appuie sur des algorithmes de désambiguïsation de mots (WSD) pour indexer les documents et les requêtes avec le sens des mots (mots-sens) plutôt qu'avec des mots simples. Une manière d'indexer serait par exemple, d'associer aux mots extraits, des mots du contexte qui aident à déterminer leur sens.

7.3.3 Indexation conceptuelle : l'indexation conceptuelle se base sur des concepts tirés d'ontologies et de taxonomies pour indexer les documents contrairement aux listes de mots simples [Woods, 97] [Stairmand et al., 96]. L'appariement des concepts (Concept Mapping) peut être utilisé dans un système spécifique en rapport à un domaine fermé, comme le domaine du sport [Khan, 00], le domaine légal [Stein, 97], le domaine médical (dans le système Textpresso [Müller et al., 04] et MetaMap (UMLS)), ou encore général, comme c'est le cas dans le système FERRET [Mauldin, 91], les travaux de Woods [Woods, 97] et ceux d'Aggarwal et ses collègues [Aggarwal et al., 01]. Mauldin et Woods utilisent un dictionnaire de langue, pour l'apprentissage des relations entre termes pour le premier (en utilisant un algorithme génétique) et pour construire avec les termes des documents des hiérarchies de concepts (taxonomie) pour le deuxième

7.4 - La différence entre l'indexation sémantique et conceptuelle :

Certains auteurs différencient l'indexation sémantique de l'indexation conceptuelle [SALT90]. L'indexation conceptuelle repose, pour eux, sur des hiérarchies de concepts ou ontologies de domaine, alors que l'indexation sémantique repose sur l'utilisation d'ontologies génériques telles que WordNet. L'ontologie WordNet est limitée par rapport à la sémantique qu'elle peut contenir par contre les ontologies de domaine peuvent par leur formalisation représenter des ressources impliquant un engagement sémantique plus fort. Donc l'indexation sémantique est l'indexation de granules documentaires à partir de n'importe quelle ontologie de domaine. L'indexation sémantique se fait en deux étapes. La première étape consiste à identifier les concepts ou instances de l'ontologie dans les granules. La deuxième étape

pondère les concepts pour chaque document en fonction de la structure conceptuelle dont ils sont issus.

Conclusion :

Après avoir relaté sur les ontologies et leur domaine d'application, nous avons introduit le processus d'indexation qui nous permet de construire le vecteur de concepts pondérés.

Ce dernier sera projeté à l'ontologie du domaine établie auparavant afin de détecter les concepts qui posent problème ainsi que les apprenants en difficultés.

Pour avoir une meilleure vue des démarches de notre étude, la partie suivante détaillera les étapes essentielles à la conception et à la réalisation de notre système.

Partie 3 : Conception et **Réalisation**

Chapitre IV : Conception

Introduction

Avec l'évolution des systèmes d'apprentissage et avec les avancées technologiques dans le domaine informatique, de nouvelles attentes ont émergé, et parmi ces dernières, nous trouvons la personnalisation et l'adaptation de l'apprentissage.

Dans notre étude, la démarche a pour but final de concevoir un système de déduction automatique des apprenants en obstacle, il est basé sur l'analyse de leurs comportements dans un environnement de formation sur le Web et permet ainsi de fournir aux enseignants des informations ou des indications sur leurs difficultés. Les informations servent à décrire le profil de l'apprenant intéressant l'enseignant dans sa tâche de tutorat et aident le concepteur de l'EIAH dans sa tâche de personnalisation des environnements de formation.

Dans ce chapitre, nous allons présenter l'environnement d'apprentissage fréquenté par l'apprenant. Nous aborderons la phase de traçage qui consiste à suivre les activités (messages postés dans les forums) de l'étudiant sur le web, puis, nous procéderons à la construction du corpus. Cette dernière se résume en l'enregistrement des données dans la base de données.

La suite du chapitre porte sur la modélisation du cours enseigné qui se fera à l'aide de l'ontologie du domaine. Nous arriverons après à l'indexation, celle là va nous permettre de retrouver l'équivalence entre ce que l'apprenant écrit et ce que le cours comporte comme concepts. Enfin, une conclusion clôturera le chapitre.

IV-1. Présentation du milieu d'apprentissage :

La navigation d'un apprenant dans un système d'apprentissage (plateforme) peut être définie comme étant une activité de formation dans laquelle l'apprenant consulte les cours, s'exerce et édite un message en communiquant avec les autres utilisateurs de la plateforme. Cependant, il éprouve le besoin de rechercher ou de solliciter des informations concernant un ou plusieurs concepts d'un contenu pédagogique auxquels il a été confronté dans son environnement d'apprentissage. Pour palier à ce manque d'informations et de renforcer ses lacunes, l'apprenant fait appel aux différents moyens qui l'entourent tels que des documents locaux (enregistrés dans son PC, sur CD) ou des recherches sur le Web (forums).

Les interactions de l'apprenant pendant sa session d'apprentissage peuvent être résumées dans la figure ci-dessous.

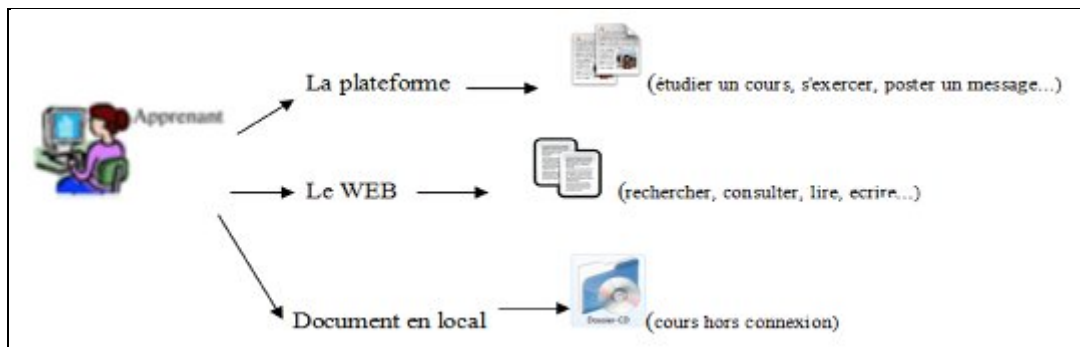


Figure 9 : Interactions de l'apprenant durant une session d'apprentissage.

Dans le cadre de notre travail, nous allons nous intéresser aux activités élaborées en dehors de la plateforme de formation, c.-à-d., sur les recherches menées sur le Web, plus précisément dans les forums en mettant en évidence l'équivalence existante entre les concepts vus en cours et ceux recherchés en dehors du cours (environnement extérieur).

Nous partons sur l'hypothèse qui dit que l'apprenant n'a recours aux forums que lorsqu'il est en difficulté sur les concepts qu'il édite d'ailleurs sur ces mêmes forums. En effet l'objectif de notre travail est de détecter les apprenants en difficulté ainsi que les concepts qui leurs posent problème, dans le but d'aider et le tuteur pour suivre ces apprenant et l'auteur du cours pour améliorer le contenu et la structure du cours sur les concepts ainsi détectés.

Afin de déterminer ces apprenants, nous devons d'abord les tracer puis récupérer et restructurer les traces brutes issues des sources de leur traçage. Pour cela nous avons opté pour les traces d'apprentissage basé sur un système d'espionnage (un keylogger). Ensuite, nous devons implémenter une ontologie du domaine concernant le cours enseigné sur la plateforme. Et enfin, on exploite nos données collectées et celles construites en comparant les concepts vus dans le cours et les informations recherchées en dehors du cours.

La figure ci-après montre l'architecture générale de notre système.

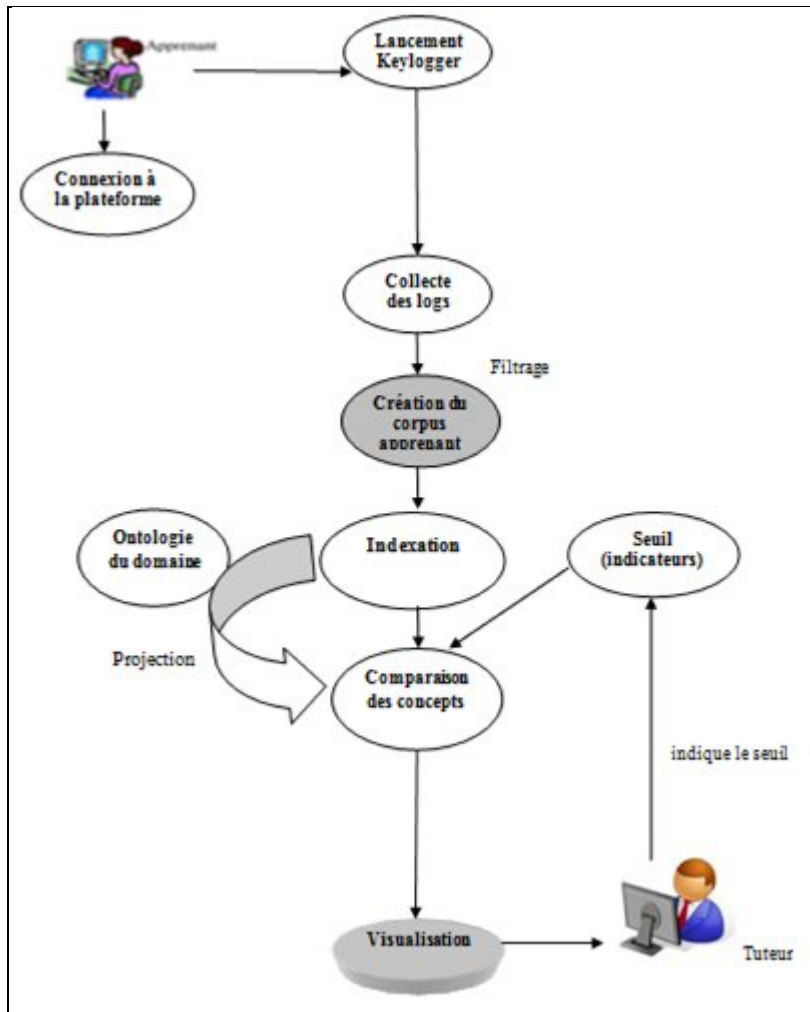


Figure 10 : Architecture générale du système.

Pour mieux comprendre le fonctionnement du système et de bien maîtriser sa complexité et d'assurer sa cohérence, une étude conceptuelle avant son élaboration est nécessaire.

Pour ce faire, nous allons procéder comme suit :

- Tracer l'apprenant et construire son profil.
- Implémenter l'ontologie du domaine résumant le cours enseigné.
- Indexation et pondération.

IV-2. Traçage de l'apprenant :

2.1. Phase collecte :

Notre objectif est d'avoir un œil sur le comportement (messages postés, sites visités) de l'apprenant en dehors de la plateforme d'étude, notamment dans les forums. Pour ce fait, nous choisissons de tracer toutes ses interactions à travers un outil d'observation installé sur sa machine, et qui sera activé au début de la session d'apprentissage par sa propre volonté (remédier au problème de confidentialité). En conséquence, les traces générées seront structurées et sauvegardées dans un fichier.

La collecte est le fait de récolter, regrouper toutes les informations (URL, données saisies, fichiers) concernant l'apprenant durant sa phase d'apprentissage. Elle se fait par le biais d'un keylogger.

Keylogger c'est quoi ?

Définition : Les keylogger sont des enregistreurs de touches et par extension des enregistreurs d'activités informatiques permettant d'enregistrer les touches utilisées par un utilisateur sur son clavier et tous les événements déclenchés.

Voici un exemple de traçage fournit par MiniKeyLog que nous utilisons dans notre étude.

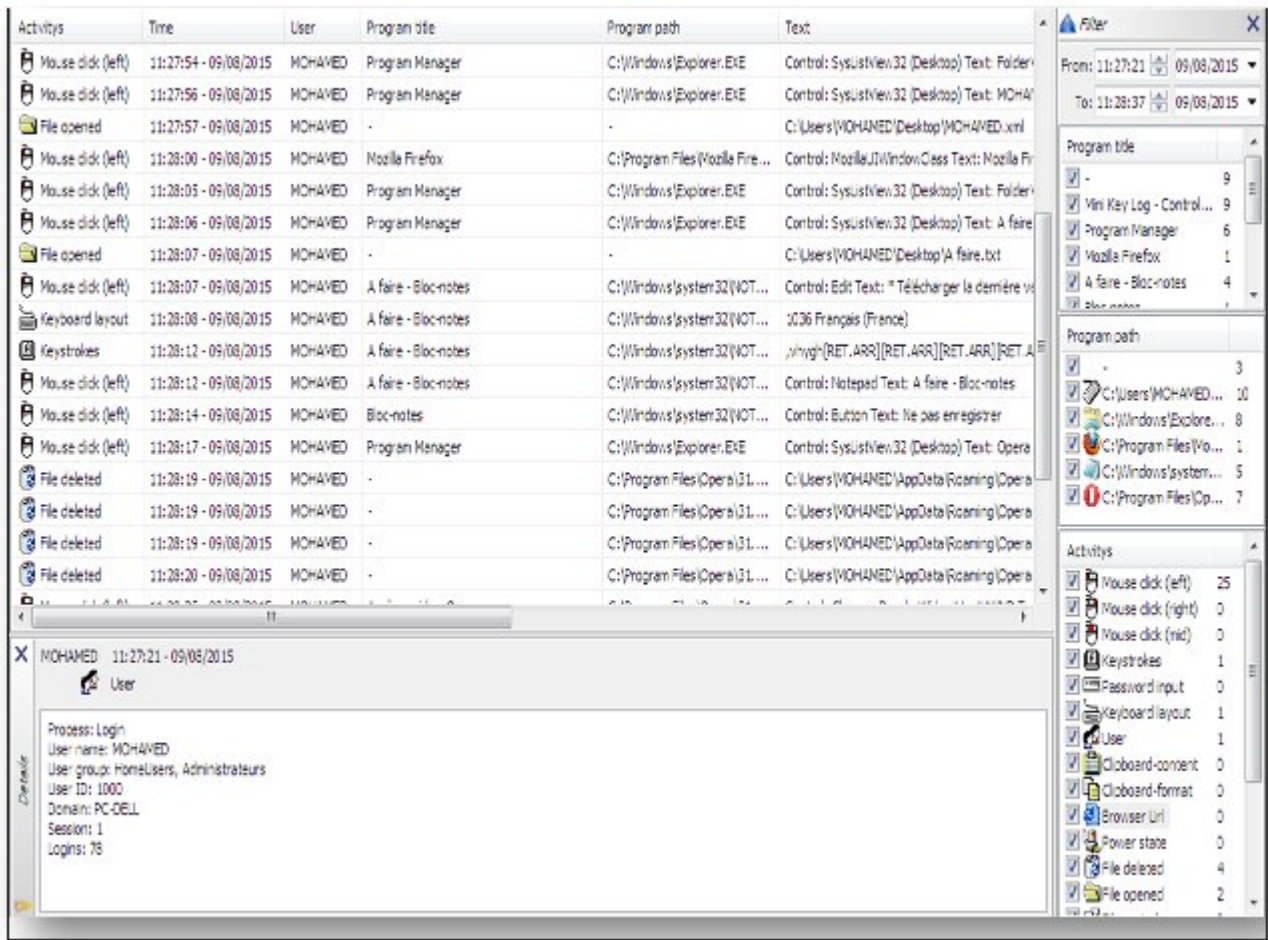


Figure 11 : Interface du MiniKeyLog lors du traçage.

2.2. Filtrage des traces brutes :

La phase de collecte produit une très grande masse de données qui peuvent être enregistrées dans des fichiers de différents formats (txt, xml, xsl, html). Dans notre travail, nous utilisons le format XML pour une meilleure maniabilité grâce à sa structure et obtenir au final un

fichier comportant la trace brute. La figure ci-dessous montre un extrait de fichier trace XML.

```

<?xml version="1.0" encoding="utf-8" ?>
- <DOCUMENT>
- <ITEM>
  <ACTIVITY>User</ACTIVITY>
  <TIME>15:57:17 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>-</TITLE>
  <PATH>-</PATH>
  <TEXT>Process: Login</TEXT>
  <TEXT>User name: MOHAMED</TEXT>
  <TEXT>User group: HomeUsers, Administrateurs</TEXT>
  <TEXT>User ID: 1000</TEXT>
  <TEXT>Domain: PC-DELL</TEXT>
  <TEXT>Session: 1</TEXT>
  <TEXT>Logins: 29</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Mouse click (left)</ACTIVITY>
  <TIME>12:56:09 - 31/08/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Mini Key Log - Control Center 2.7</TITLE>
  <PATH>C:\Users\MOHAMED\Desktop\MiniKeyLog_10.exe</PATH>
  <TEXT>Control: TntTreeView.UnicodeClass</TEXT>
  <TEXT>Text: MOHAMED</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Browser Url</ACTIVITY>
  <TIME>15:57:38 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Nouvel onglet - Windows Internet Explorer</TITLE>
  <PATH>C:\Program Files\Internet Explorer\iexplore.exe</PATH>
  <TEXT>res://www.developpez.fr\creer-base-de-donnees</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>File deleted</ACTIVITY>
  <TIME>12:57:40 - 31/08/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Program Manager</TITLE>
  <PATH>C:\Windows\Explorer.EXE</PATH>
  <TEXT>C:\Users\MOHAMED\Desktop\Nouveau fichier.accdb</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Keyboard layout</ACTIVITY>
  <TIME>15:57:57 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Nouvel onglet - Windows Internet Explorer</TITLE>
  <PATH>C:\Program Files\Internet Explorer\iexplore.exe</PATH>
  <TEXT>1036 Francais (France)</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>File renamed</ACTIVITY>
  <TIME>12:57:19 - 31/08/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Program Manager</TITLE>
  <PATH>C:\Windows\Explorer.EXE</PATH>
  <TEXT>From: source</TEXT>
  <TEXT>To: Nouveau fichier.accdb</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Keystrokes</ACTIVITY>
  <TIME>15:59:20 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Bloc-notes</TITLE>
  <PATH>C:\Windows\system32\notepad.exe</PATH>
  <TEXT>comment construire une bdd</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>File opened</ACTIVITY>
  <TIME>15:58:57 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>-</TITLE>
  <PATH>-</PATH>
  <TEXT>C:\Users\MOHAMED\Desktop\A faire.txt</TEXT>
</ITEM>
</DOCUMENT>

```

Figure 12 : Fragment de fichier trace sous format XML.

Après l’obtention de notre fichier trace XML, nous procédons à la phase du filtrage des activités, qui consiste à restreindre le domaine de recherche et d’éliminer les activités dont nous n’avons pas besoin, et de garder les suivantes :

- Les sites visités par l’apprenant (ACTIVITY : Browser Url)
- La saisie au clavier dans les forums (ACTIVITY : Keystrokes)

Cela dans le but de raffiner l'analyse et faciliter la prise en main des informations récoltées.

Pour ce faire, nous optons pour le traitement suivant effectué sur les fichiers XML, nous présentons ceci dans cet algorithme :

Choisir un fichier trace xml et faire appel au module **Filtrage** conçu dans notre système.

Charger le fichier XML : Parser le fichier XML, c'est le représenter sous forme d'arbre structuré dans la machine, cela en utilisant des outils dédiés : Jdom / SaxBuilder [[annexe](#)].

La balise mère du fichier XML (dans notre cas c'est la balise <DOCUMENT>) est assignée à la racine de l'arbre.

Notre système de filtrage parcourt l'arbre comme montré dans l'algorithme suivant:

TANT QUE (<ITEM> existe)

Pointer sur <ITEM> ;

//Nous testons son propre fils <ACTIVITY>

SI (<ACTIVITY> Browser Url </ACTIVITY>)

/* Nous testons si le site est un forum intéressant (remédier au problème de confidentialité) */

SI (<TEXT> comporte un nom du domaine d'un site intéressant)

Prendre en compte le nœud <TEXT> ;

//celui-ci comporte le lien du forum visité

SINON Ignorer ; /* Remédier au problème de confidentialité en ignorant les

sites qui ne sont pas des forums intéressants*/

SINON SI (<ACTIVITY> Keystrokes </ACTIVITY>)

/* Nous testons l'environnement dans lequel le texte à été tapé (dans le but de remédier au problème de confidentialité) */

SI (<PATH> comporte un nom du domaine d'un site intéressant)

Prendre en compte le nœud <TEXT> ;

/* Celui-ci comporte le message édité par l'apprenant dans le forum cité dans <PATH> */

SINON Ignorer ; /* Remédier au problème de confidentialité en ignorant les

messages postés en dehors des forums intéressants*/

SINON Ignorer ; /* Remédier au problème de confidentialité en ignorant les activités

dont nous n'avons pas besoin*/

<ITEM> suivant ;

Fin TANT QUE

Figure 13 : Algorithme de filtrage.

Remarque : Nous avons élaboré au préalable une liste des sites (forums) les plus intéressants (les plus visités par les étudiants) pour tester le <PATH>.

Nous laissons la possibilité au tuteur/auteur d'ajouter ou de supprimer un site de la liste jugé en fonction de son importance.

Une fois le filtrage terminé, le corpus est prêt à être construit.

IV-3. Construction du corpus apprenant :

La construction du corpus se fait à travers l'enregistrement des activités, tirées du fichier trace XML filtré, dans la base de données.

Le schéma ci-après montre les étapes de construction du corpus.

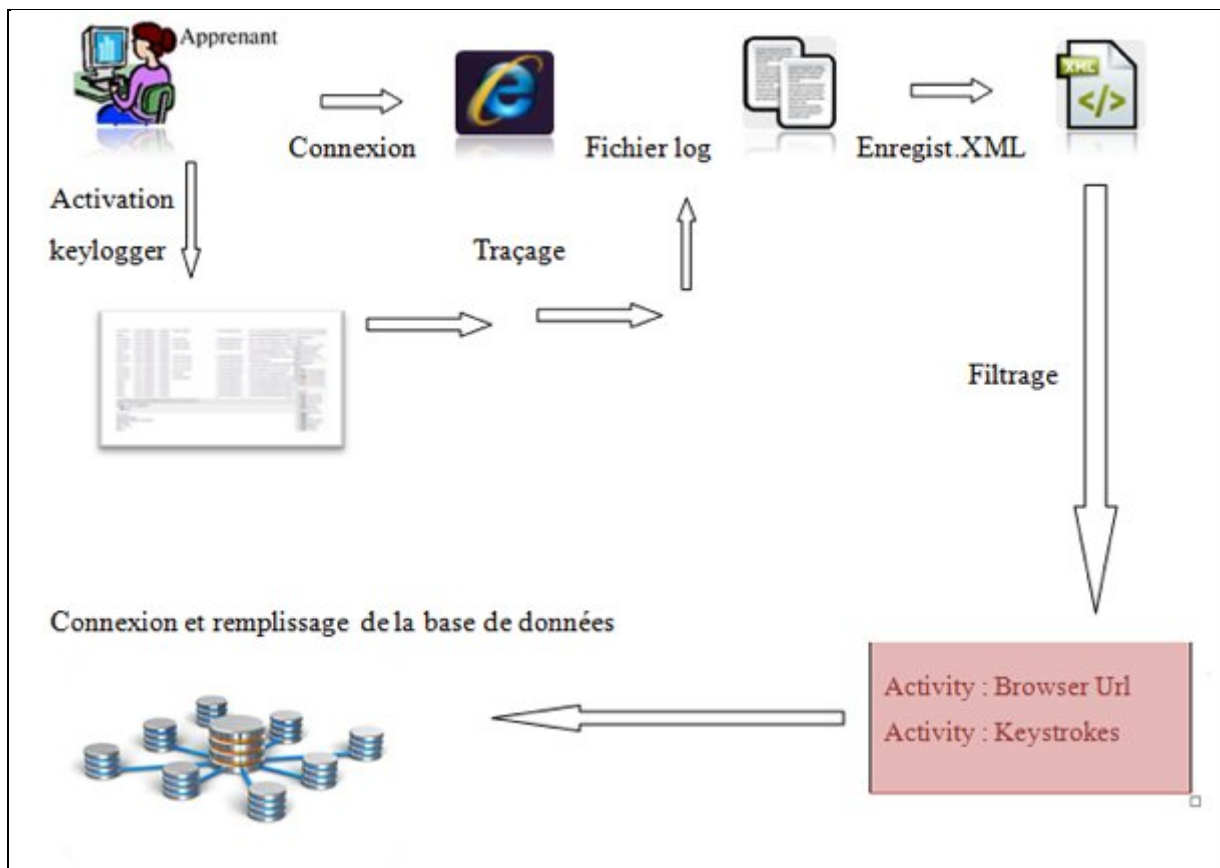


Figure 14 : Etapes de construction du corpus

Chaque fichier trace XML porte comme nom le pseudo de l'apprenant dans sa plateforme d'apprentissage, ce nom est unique et peut donc être un bon identifiant lors de la construction du corpus.

La construction du corpus qui consiste en effet aux messages édités par l'apprenant sur les forums d'apprentissage en ligne, se fait suivant un modèle entité association adéquat qui répond à ces contraintes :

- L'enseignant charge un fichier trace XML propre à l'étudiant, le nom du fichier sera l'identifiant de l'apprenant.
- Le fichier comporte les activités de l'apprenant
- L'apprenant peut visiter des forums
- L'apprenant peut poster des messages dans des forums
- L'enseignant peut charger un ou plusieurs fichiers
- L'enseignant peut superviser l'apprenant sur des différentes périodes, en ayant accès à différentes dates de consultation, c.-à-d., qu'il peut à tout moment l'évaluer.

Nous proposons le modèle Entité/Association ci-dessous :

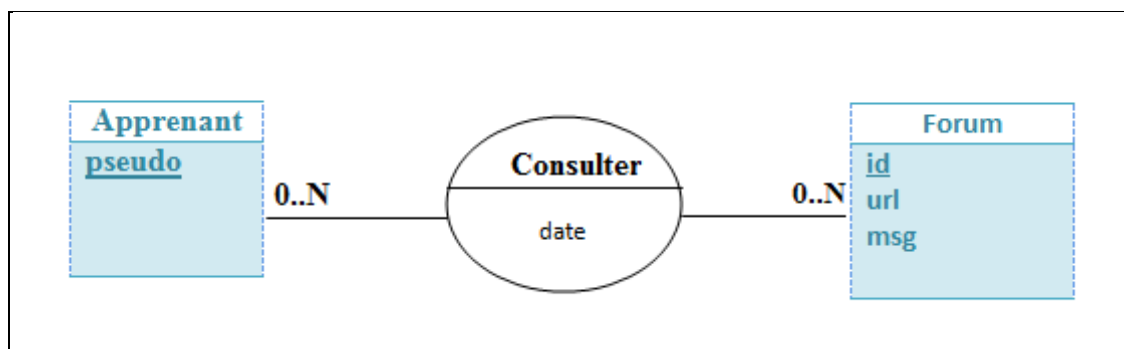


Figure 15: Modèle Entité-Association du corpus apprenant.

Nous enregistrerons les messages postés par chaque apprenant dans la table **Forum**, en prenant en compte que ceux postés dans des forums intéressants (une liste des plus utilisés est énumérée au préalable « commentçamarche, developpez... »), nous laissons aussi le choix au tuteur de rajouter ou de supprimer un forum).

Nous gardons trace des sites visités dans la table **Forum**.

La propriété **msg** peut être nulle, car, l'apprenant peut visiter un forum sans qu'il ne poste aucun message dedans.

Nous passons maintenant au modèle relationnel qui va nous faciliter l'organisation et l'implémentation de nos données dans la base de données.

Nous obtenons le schéma relationnel suivant :

Apprenant (pseudo)

Forum (id, url, msg)

Consulter (pseudo, id, date)

Après avoir obtenu le corpus des messages ainsi édités, nous allons procéder à leur indexation. Pour réaliser cette étape, nous allons utiliser l'ontologie du domaine du cours afin de détecter les concepts du domaine.

Une ontologie du domaine est mise en place, celle-ci représente le cours enseigné en terme de concepts reliés par des relations sémantiques. Cette ressource sémantique va nous être utile lors de la comparaison entre les activités de l'apprenant et le cours enseigné.

IV-4. Mise en place de l'ontologie du domaine :

L'étude assistée par machine d'un domaine donné requiert l'extraction de ses informations et la représentation de celles-ci sous forme de connaissances compréhensibles par la machine. Des techniques d'implémentation intégrant des connaissances sur le domaine, comme les ontologies, permettent la modélisation de cet ensemble et incluent des définitions lisibles en machine des concepts de base de ce domaine et de leurs relations.

Pour concrétiser cette approche, nous avons opté pour la modélisation du domaine du cours enseigné à l'aide d'ontologie. Celui-ci se résume aux concepts liés à l'algorithme. Nous devons apporter une précision quant à l'élaboration de l'ontologie du domaine, car elle doit être hiérarchisée au sein d'une structure et liée par des relations sémantiques.

Nous nous sommes arrêtés au nombre de quatre concepts :

- Les boucles
- Les structures de données
- Les procédures
- Les conditions

Le but est de comparer les termes (mots clés) constituant le corpus apprenant à ceux de l'ontologie du domaine élaborés auparavant. Cette phase consiste en la projection.

Pour effectuer cette dernière, nous avons opté pour le format de représentation SKOS, qui permet de représenter des termes et des concepts. Chacun de ces concepts regroupe un certain nombre de termes, organisés en terme général « preferred_label » et en terme alternatif « alter_label ». Cette modélisation accepte aussi différentes langues, définies par des attributs dans les balises (Exemple : <altlab lang="fr">tant que</altlab> ; <altlab lang="en">while</altlab>) ou "fr" désigne le français et "en" l'anglais.

Ainsi, nous verrons dans notre ontologie la structure suivante :

```

1  <Concepts name="algorithmique">
2
3  <Concept ord="1" name="structure de donnees">
4      <preferlab lang="fr">structure de donnees</preferlab>
5      <preferlab lang="en">data structure</preferlab>
6      <altlab lang="fr">vecteur</altlab>
7      <altlab lang="en">file</altlab>
8      <altlab lang="fr">liste chaînée</altlab>
9      <altlab lang="fr">matrice</altlab>
10     <altlab lang="fr">pile</altlab>
11     <altlab lang="fr">tableau</altlab>
12 </Concept>
13
14
15 <Concept ord="2" name="les procedures">
16     <preferlab lang="en">function</preferlab>
17     <preferlab lang="fr">fonction</preferlab>
18     <preferlab lang="fr">procedure</preferlab>
19     <altlab lang="fr">methode</altlab>
20     <altlab lang="fr">paramètres</altlab>
21     <altlab lang="fr">Fonctions personnalisées</altlab>
22     <altlab lang="fr">Variables publiques</altlab>
23     <altlab lang="fr">Variables privées</altlab>
24
25 </Concept>
26
27 <Concept ord="3" name="les conditions">
28     <preferlab lang="fr">conditions</preferlab>
29     <altlab lang="fr">Structures conditionnelles</altlab>
30     <altlab lang="en">La condition if... else</altlab>

```

Figure 16 : Ontologie en format SKOS

Nous procédons maintenant à l'analyse et l'exploitation de nos données, nous cherchons les concepts qui figurent et dans le cours enseigné et dans les activités de l'apprenant.

IV-5. Indexation :

Afin d'évaluer l'équivalence entre ce que l'apprenant fait en dehors de la plateforme et le contenu du cours, nous cherchons l'apparition des concepts de notre ontologie dans les messages édités par l'apprenant. Il est à noter que le cours est scénarisé par cette ontologie du domaine comme expliqué précédemment. Les activités de l'apprenant en dehors de l'environnement d'apprentissage sont quant à elles, représentées dans le corpus apprenant (ensemble de tables dans la base de données), par des textes, suite de chaîne de caractères.

Dans notre étude, la recherche se base essentiellement sur l'indexation, une approche intéressante serait de représenter ces indexes par un vecteur.

Nous remodelons notre ontologie de telle manière à ce qu'elle soit maniable. En pratique, nous optons pour le modèle d'espace vectoriel pour la construction de l'index du cours, qui consiste en la matrice dont les colonnes représentent les concepts. On trouve leurs intitulés respectifs dans la ligne 1, tandis que les lignes qui suivent comportent les termes correspondants aux concepts de la ligne 1 (labels).

Le tableau suivant est la structure du modèle utilisé :

Concept 1	Concept i	Concept n
Terme1	Terme1	Terme1
Terme j	Terme j	Terme j
Terme m	Terme m	Terme m

Tableau 2 : Modèle de représentation des concepts.

Indexation des activités de l'apprenant :

Notre processus d'indexation pour les activités comporte les étapes suivantes :

a. Identification des termes représentatifs : Cette étape nous permet d'extraire les termes qui structurent l'unité textuelle d'un texte. Une unité textuelle peut être représentée par une phrase, un paragraphe ou un nœud texte dans un document XML... etc

Le processus d'identification des termes se déroule comme suit :

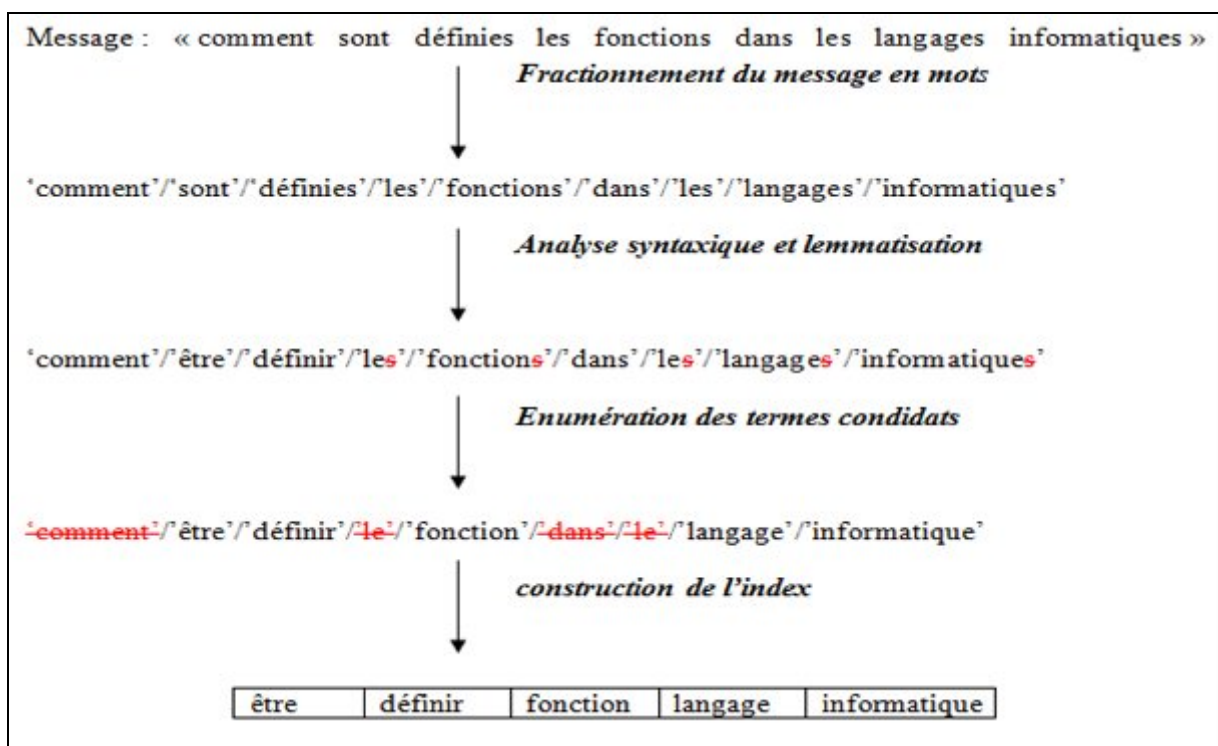
Analyse syntaxique et morphosyntaxique : Analyser le texte afin d'extraire les mots qu'il contient, on divise le texte mot par mot. Ces mots sont ensuite étiquetés syntaxiquement [annexe] et lemmatisés [annexe].

Enumération des termes candidats : Repérer les séquences de mots susceptibles d'être des termes candidats, nous ignorons donc dans le texte les mots considérés non intéressants, on les appelle « mots vides ». (Exemple: 'conjonctions de coordination', 'quoi', 'comment' ...).

b. Construction des index :

Dans cette approche, le modèle vectoriel sémantique est utilisé pour la représentation des indexes du texte (activités de l'apprenant), un nœud de type texte est représenté par un vecteur des termes $V_t = \langle t_1, \dots, t_i, \dots, t_m \rangle$ où t est le terme candidat et m le nombre de ces termes.

Voici un exemple d'indexation d'un message :*



Exemple d'indexation d'un message

IV-6. Projection Activité/Ontologie :

Il s'agit d'une assignation des concepts aux termes. Un terme est considéré comme le label d'un concept. L'idée est alors de rechercher dans notre ressource sémantique (Ontologie) le concept qui correspond au terme considéré.

Nous procédons à l'extraction des concepts correspondants, nous lançons une fouille des termes (produits par l'indexation) dans la matrice qui représente notre ressource sémantique, on appelle cela la 'Projection'. En retour, nous aurons un **vecteur de concepts pondérés**, celui-ci contient l'ensemble des concepts correspondants aux termes du vecteur V_t recherchés par l'apprenant durant sa session d'apprentissage.

La figure suivante est un exemple résumant la démarche utilisée.

Nous reprenons le vecteur des termes V_t produit auparavant.

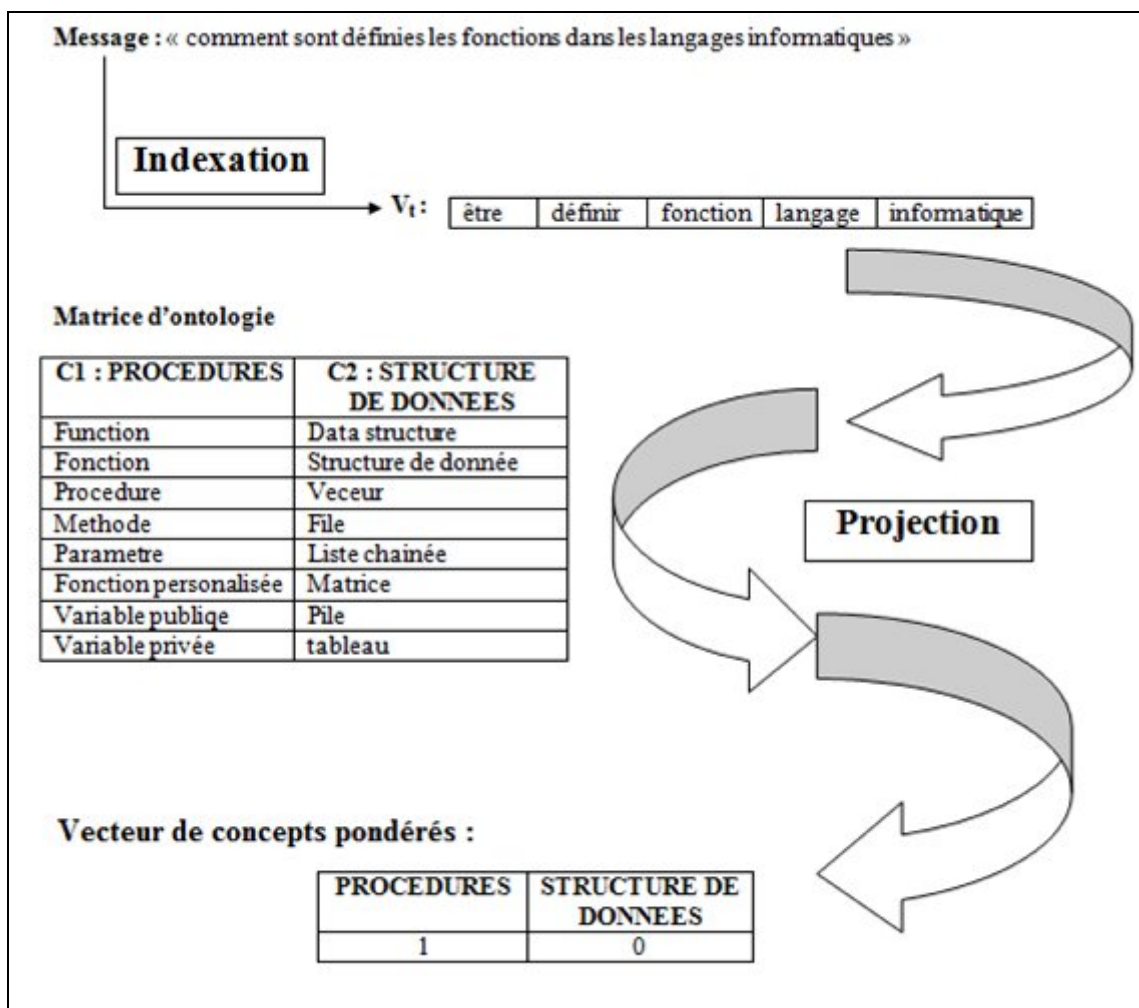


Figure 17 : Exemple d'indexation et de projection

Le tuteur/auteur intervient ensuite en fixant un seuil. Ce dernier va être comparé au vecteur de concepts pondérés pour préciser les concepts qui posent problèmes et les apprenants en difficulté.

Notre système offre deux choix d'évaluation :

Evaluation individuelle :

Si le nombre de concepts pondérés est supérieur au seuil, l'apprenant est sensé être en difficulté, c-à-d, plusieurs recherches sont effectuées, donc il a du mal comprendre le cours. Des détails concernant ses lacunes seront ensuite retournés pour le tuteur sous forme de visualisations dans l'interface graphique préparée.

Evaluation collective :

Si un concept est recherché par plusieurs apprenants, cela signifie que le cours enseigné dans la plateforme n'est pas bien structuré et adapté aux profils des apprenants. Ce résultat va conduire le concepteur de la plateforme (propriétaire du cours) à une modification et amélioration du contenu pédagogique sur les concepts ainsi détectés dans le processus précédent.

Le diagramme de séquence modelé suivant montre bien la partie exploitation des résultats par le tuteur/auteur

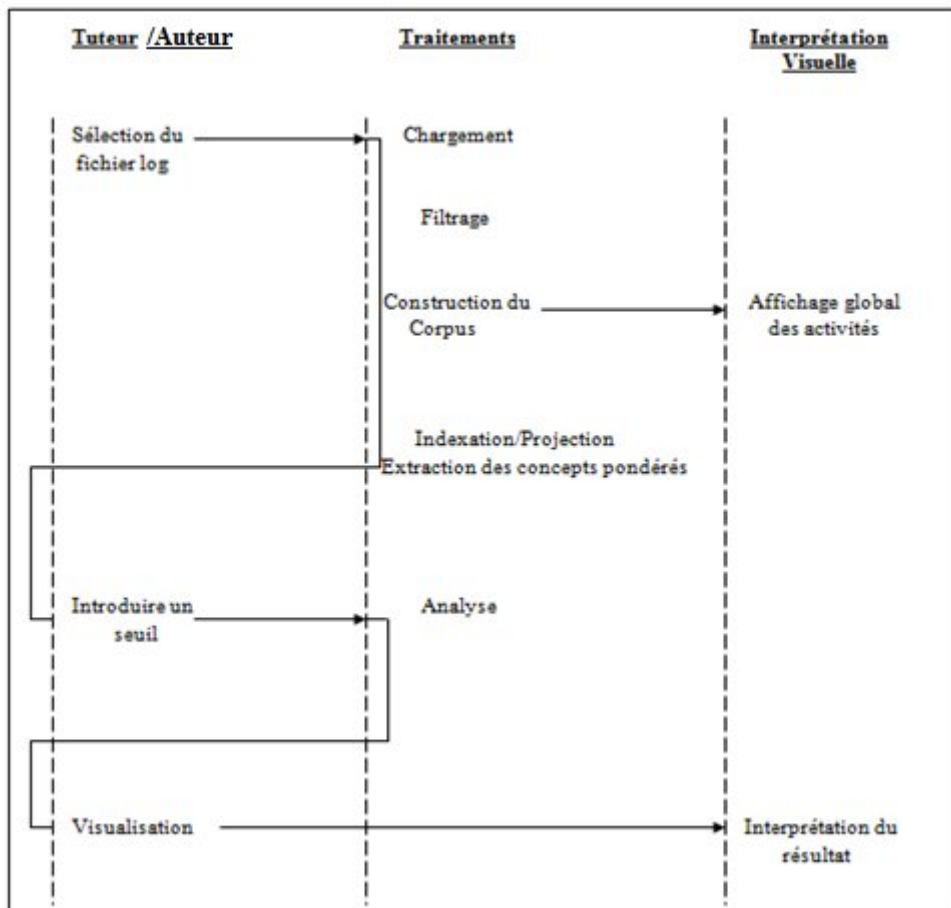


Figure 18 : Diagramme de séquence de notre système.

Conclusion :

Dans ce chapitre, nous avons décrit les étapes essentielles qui permettent la concrétisation de notre recherche, en commençant par le filtrage du fichier trace XML. Nous avons laissé uniquement les messages postés dans les forums ainsi que les sites visités. Puis nous avons mis en place une ontologie du domaine à laquelle nous avons projeté notre vecteur de concepts pondérés construit durant la phase d'indexation. Cela nous a permis de détecter les concepts qui posent problème et les apprenants en difficulté. Cette partie présente aussi notre étude conceptuelle et annonce l'implémentation de notre solution qui se fera dans le chapitre suivant.

Chapitre V : Réalisation

Introduction

La réalisation joue un rôle essentiel dans le développement des systèmes informatiques. Elle résulte des spécifications décrites lors de l'étude conceptuelle afin de produire un outil capable d'effectuer les traitements apportés.

Après avoir décrit dans le chapitre précédent les différentes étapes de l'analyse et conception de notre étude, nous allons présenter dans celui-ci l'environnement de développement et les outils choisis pour la réalisation de notre application ; Nous afficherons ensuite quelques fonctionnalités de l'application ainsi que les résultats retournés dans des captures d'écrans de l'application conçue. Nous terminerons enfin par une conclusion.

V-1. Description du matériel utilisé :

La conception de notre application a été faite sur un micro-ordinateur dont les caractéristiques sont les suivantes :

Edition Windows: Windows 7 Edition Intégrale.

Mémoire installée (RAM) : 6,00 Go/Utilisable : 2.40Go

Type du système : Système d'exploitation 32 bits.

Processeur : Intel ® Core™ i5-4200U CPU @1.60GHz 2.30GHz

V-2. Logiciels utilisés :

MiniKeyLog :

C'est un enregistreur de touches et par extension enregistreur d'activités informatiques permettant d'enregistrer les touches utilisées par un utilisateur sur son clavier et tous les événements déclenchés. Il s'agit d'un logiciel espion installé sur la machine. Il permet de retrouver toutes les activités faites durant une session d'espionnage sous forme de traces.

NetBeans :

NetBeans est un environnement de développement intégré (EDI), placé en *open source* par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License) et GPLv2.

En plus de Java, NetBeans permet également de supporter différents autres langages, comme C, C++, JavaScript, XML, PHP et HTML ainsi que bien d'autres. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web).

Conçu en Java, NetBeans est disponible sous Windows, Linux, Solaris, Mac OS X ou sous une version indépendante des systèmes d'exploitation (requirant une machine virtuelle Java). Un environnement Java Development Kit JDK est requis pour les développements en Java.

NetBeans constitue par ailleurs une plate forme qui permet le développement d'applications spécifiques (bibliothèque Swing (Java)). L'IDE Netbeans s'enrichit à l'aide de greffons.

V-3. Présentation du langage de programmation utilisé :

Pour développer notre application, nous avons utilisé le langage de programmation JAVA (Jdk-8u20).

Java est un langage de programmation à usage général, évolué et orienté objet. Ses caractéristiques ainsi que la richesse de son écosystème et de sa communauté lui ont permis d'être très largement utilisé pour le développement d'applications de types très disparates.

V-4. SGBD :

Pour la création de notre base de données, nous avons choisi de travailler sur le logiciel PostgreSQL (9.3).

PostgreSQL est un système de gestion de base de données (SGBD) et objet (SGBDRO). C'est un outil libre, disponible selon les termes d'une licence de type BSD (*Berkeleysoftwaredistributionlicense*).

IL peut stocker plus de types de données que les types traditionnels entiers, caractères, etc. L'utilisateur peut créer des types, des fonctions, utiliser l'héritage de type, etc. PostgreSQL est largement reconnu pour son comportement stable, proche de *Oracle*.

V.4.1. Connexion à la base de données :

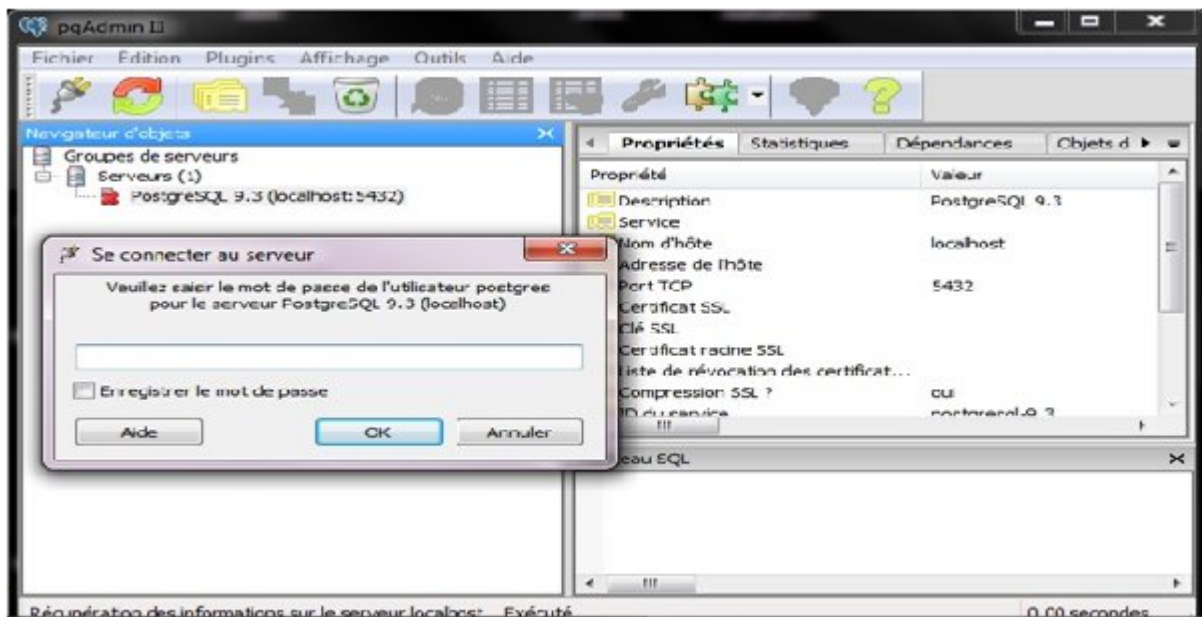


Figure 19 : Capture d'écran PostgreSQL lors de la connexion à la base de données

V.4.2. Création d'une table :

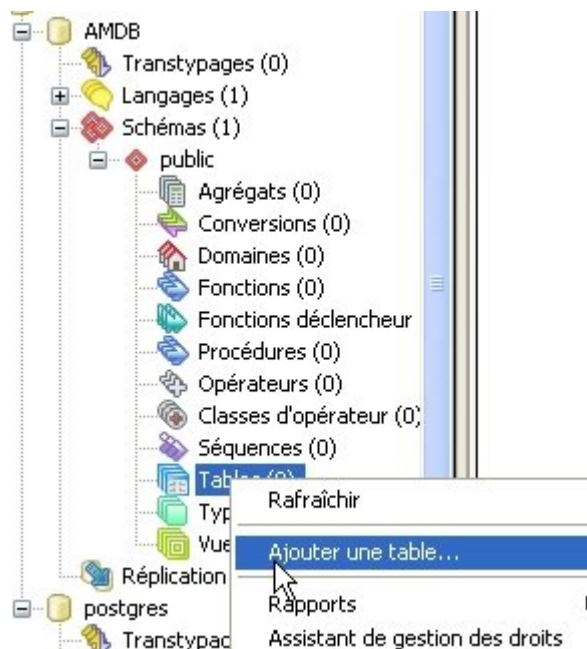


Figure 20 : Capture d'écran PostgreSQL lors de la création de tables

V.4.3 L'ajout d'une table :

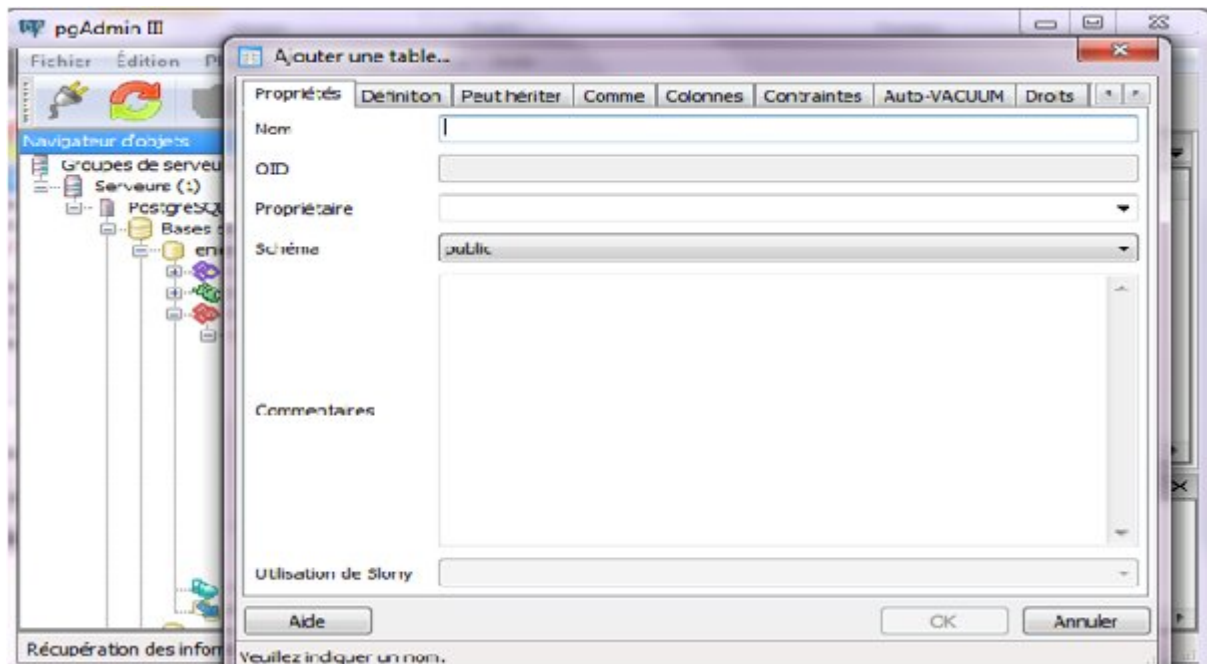


Figure 21 : Capture d'écran PostgreSQL lors de l'ajout d'une table

V-5. Connexion à la base de données :

Pour établir une connexion entre le SGBD et la BDD, on a besoin d'un driver JDBC.

JDBC (*Java DataBase Connectivity*) est une interface de programmation créée par Sun Microsystems pour les programmes utilisant la plateforme Java. Elle permet aux applications Java d'accéder par le biais d'une interface commune à des sources de données pour lesquelles il existe des pilotes JDBC.

Quatre types existent :

- *Type 1* : Pilotes agissant comme passerelle en permettant l'accès à une base de données grâce à une autre technologie (JDBC-ODBC via ODBC).
- *Type 2* : Intègre les pilotes natifs et les pilotes Java. En fait, la partie Java traduit les instructions en natif afin d'être comprises et interprétées par les pilotes natifs.
- *Type 3* : Pilotes convertissant les appels JDBC en un protocole indépendant de la base de données. Un serveur convertit ensuite ceux-ci dans langage souhaité par le SGBD.

- *Type 4* : Pilotes convertissant les appels JDBC directement en un protocole réseau exploité par la base de données. Ces pilotes encapsulent directement l'interface cliente de la base de données et sont fournis par les éditeurs de base de données.

La figure ci-dessous montre le rôle du pilote lors de la connexion à la base de données.

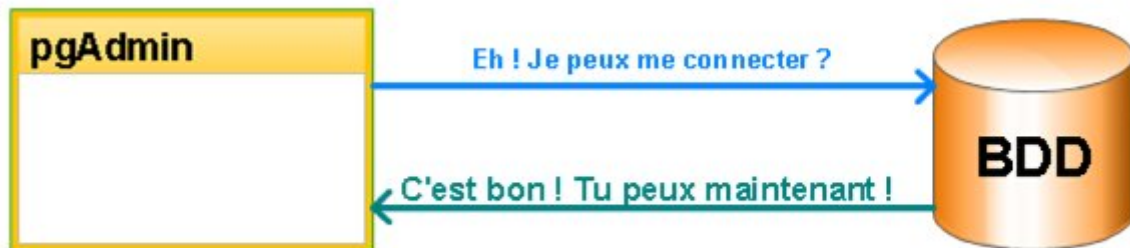


Figure 22 : Communication entre le SGBD et la base de données

6. Simulation d'un cas d'utilisation :

Nous présentons ici une simulation de notre système sur un cas d'analyse. Nous présentons dans cette partie quelques fonctionnalités du système conçu illustrées par des captures d'écran.

Dans un premier temps, un apprenant ouvre une session d'apprentissage, le pseudo de celui-ci est '[MohamedAnis](#)'. Il lance le logiciel d'espionnage MiniKeyLog installé sur sa machine pour permettre le traçage, et commence ensuite à prendre son cours sur la plateforme d'apprentissage.

Le thème du cours est « ALGORITHME ». L'étudiant entame le chapitre concernant « les boucles » et trouve une difficulté dans celui-ci. Il n'arrive pas à assimiler le comportement de l'instruction « repeat » dans un programme. Par conséquent, l'élève fait une recherche sur le WEB en ayant recours aux forums, il rentre poster le message « [comment utiliser l'instruction repeat](#) » dans le forum «[commentcamarche](#)» dont l'url est «<http://www.commentcamarche.net/forum/affich-30880749-probleme-d-utilisation-de-boucle-repeat#q=repeat&cur=2&url=%2F>».

MiniKeyLog enregistre tout ce que l'apprenant fait au moment du traçage.

L'élève quitte enfin la session d'apprentissage et arrête le logiciel d'espionnage, toutes ses activités réalisées pendant la session d'espionnage (clics de souris, sites visités, fichiers manipulés, touches du clavier ...etc) sont enregistrées dans un fichier XML nommé 'MohamedAnis21-09-2015.xml' (suivant une politique de nommage qui dit que le fichier prend comme nom le pseudo de l'apprenant concaténé avec la date).

La figure suivante montre le fichier xml généré.

```
-<DOCUMENT>
- <ITEM>
  <ACTIVITY>User</ACTIVITY>
  <TIME>10:17 - 21/09/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE></TITLE>
  <PATH></PATH>
  <TEXT>Process: Login</TEXT>
  <TEXT>User name: MOHAMED</TEXT>
  <TEXT>User group: HomeUsers, Administrateurs</TEXT>
  <TEXT>User ID: 1000</TEXT>
  <TEXT>Domain: PC-DELL</TEXT>
  <TEXT>Session: 1</TEXT>
  <TEXT>Logins: 29</TEXT>
  <TEXT></TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Browser Click</ACTIVITY>
  <TIME>10:13:23 - 21/09/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Emission - Not afraid - Youtube</TITLE>
  <PATH>C:\Program Files\Internet Explorer\iexplore.exe</PATH>
  <TEXT>https://www.youtube.com/watch?v=j5-yKMD64s</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>File opened</ACTIVITY>
  <TIME>10:58:57 - 21/09/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE></TITLE>
  <PATH></PATH>
  <TEXT>C:\Users\MOHAMED\Desktop\A faire.txt</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Keyboard layout</ACTIVITY>
  <TIME>10:23:07 - 21/09/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>Nouvel onglet - Windows Internet Explorer</TITLE>
  <PATH>C:\Program Files\Internet Explorer\iexplore.exe</PATH>
  <TEXT>1036 Français (France)</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Keystrokes</ACTIVITY>
  <TIME>10:23:09 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>
  probleme-d-utilisation-de-boucle-repeat - Windows Internet Explorer
  </TITLE>
  <PATH>
  http://www.commentamarche.net/forum/affich-3080749-probleme-d-utilisation-de-boucle-repeat?q
  </PATH>
  <TEXT>comment utiliser l'instruction repeat</TEXT>
</ITEM>
- <ITEM>
  <ACTIVITY>Keystrokes</ACTIVITY>
  <TIME>15:59:20 - 09/07/2015</TIME>
  <USER>MOHAMED</USER>
  <TITLE>A faire - Bloc-notes</TITLE>
  <PATH>C:\Windows\system32\NOTEPAD.EXE</PATH>
  <TEXT>note a faire plus tard</TEXT>
</ITEM>
</DOCUMENT>
```

Figure 23 : Fichier trace xml généré lors de la simulation

6.1 Intervention de l'enseignant :

Pour avoir un œil sur l'apprenant et suivre sa progression dans le cours, le tuteur lance l'application conçue et commence son évaluation.

6.1.a Authentification :



Figure 24 : Capture d'écran de l'application, authentification

Comme son nom l'indique, cette fenetre permet l'accès au système à l'aide d'un mot de passe. Si ce dernier est erroné, il sera signalé en rouge. L'utilisateur peut quant à lui modifier le mot de passe en cliquant sur "changer de mot de passe".

Changement du mot de passe :

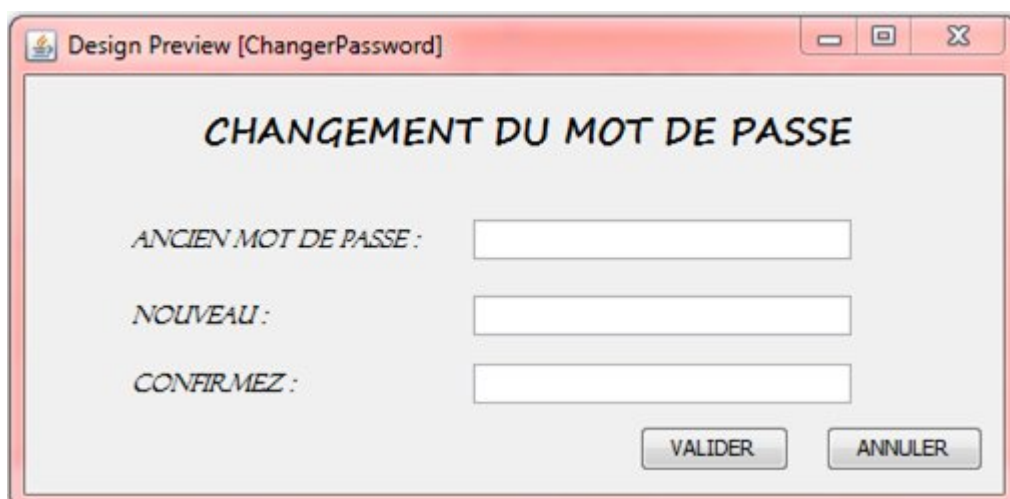


Figure 25 : Capture d'écran de l'application, changement du mot de passe

Cette fenêtre permet de renouveler le mot de passe en saisissant l'ancien. Si ce dernier n'est pas conforme, l'erreur sera signalée et le système ne nous permettra pas de le changer.

Le tuteur procède ensuite au chargement du fichier trace xml nommé 'MohamedAnis.xml' se trouvant dans le dossier 'Traces des apprenants' sur le bureau.

6.1.b Chargement du fichier :

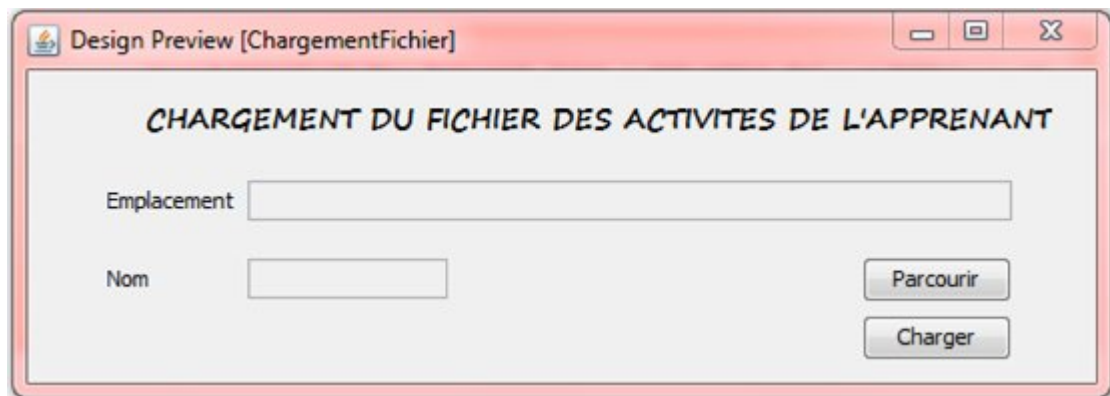


Figure 26 : Capture d'écran de l'application, chargement du fichier

L'action du clic sur le bouton "Parcourir" ouvrira la fenêtre de recherche qui permet de sélectionner le fichier à charger.

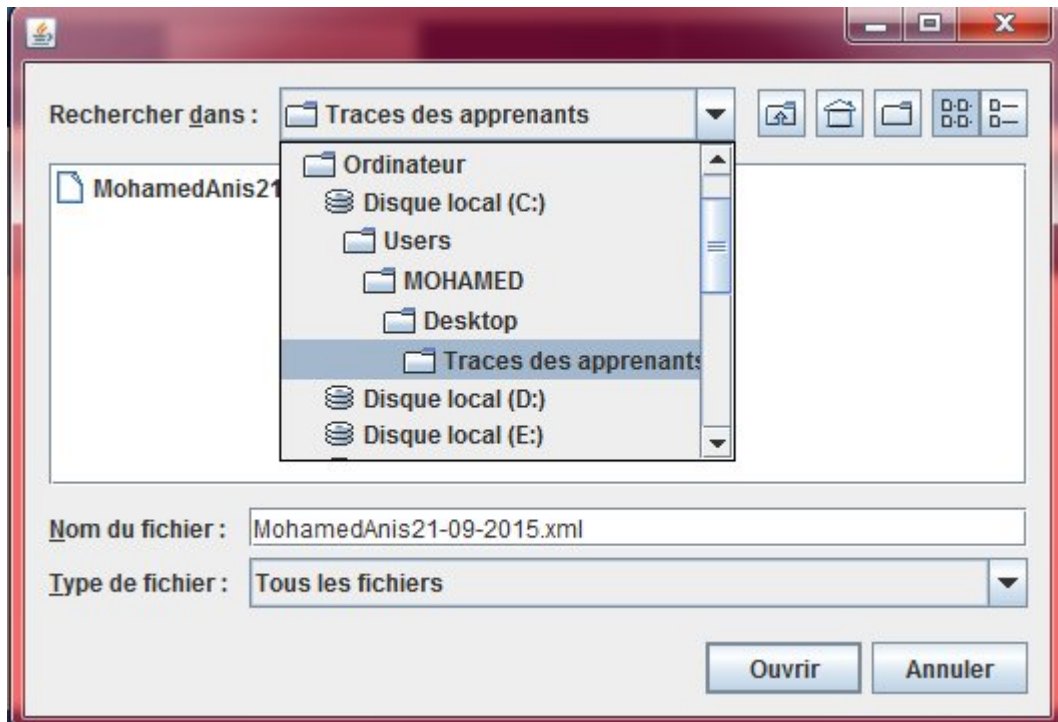


Figure 27 : Capture d'écran de l'application, sélection d'un fichier

Remarque : Un test se fait lors du choix du fichier. Le système accepte uniquement les fichiers dont le format est en "XML".

Clique 'Ouvrir'

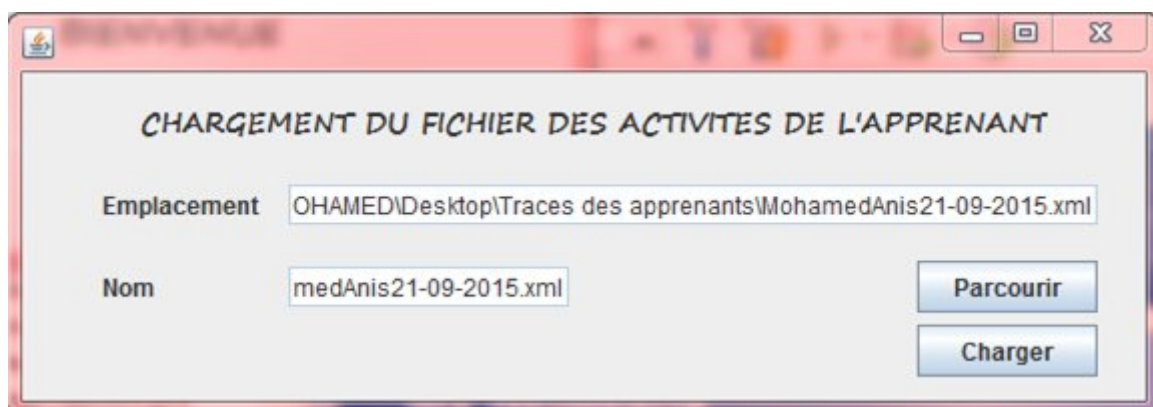


Figure 28 : Capture d'écran de l'application, choix du fichier

En cliquant sur le bouton ‘Charger’ le fichier passe par le **module de filtrage** dans le but d’alléger le contenu de la trace et de remédier au problème de confidentialité. Les activités hors ‘Keystrokes’ et ‘Browser url’ (qui représentent respectivement les touches au clavier et les sites web visités) sont ignorées.

Le module de filtrage ainsi que son algorithme sont bien détaillés dans le chapitre conception.

Après ces tests, la construction du corpus apprenant se lance et une fenêtre de dialogue apparaît au final pour indiquer au tuteur que le fichier sélectionné a été chargé.

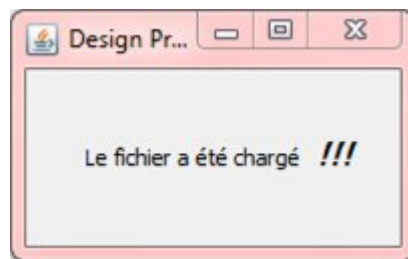


Figure 29 : Capture d’écran de l’application, message de succès

Une nouvelle étape du processus est lancée ensuite. L’analyse et L’affichage des activités enregistrées durant la session d’apprentissage.

6.1.c Affichage global des activités

Cette fenêtre affiche toutes les activités de l’apprenant à partir du corpus construit lors du chargement des fichiers.

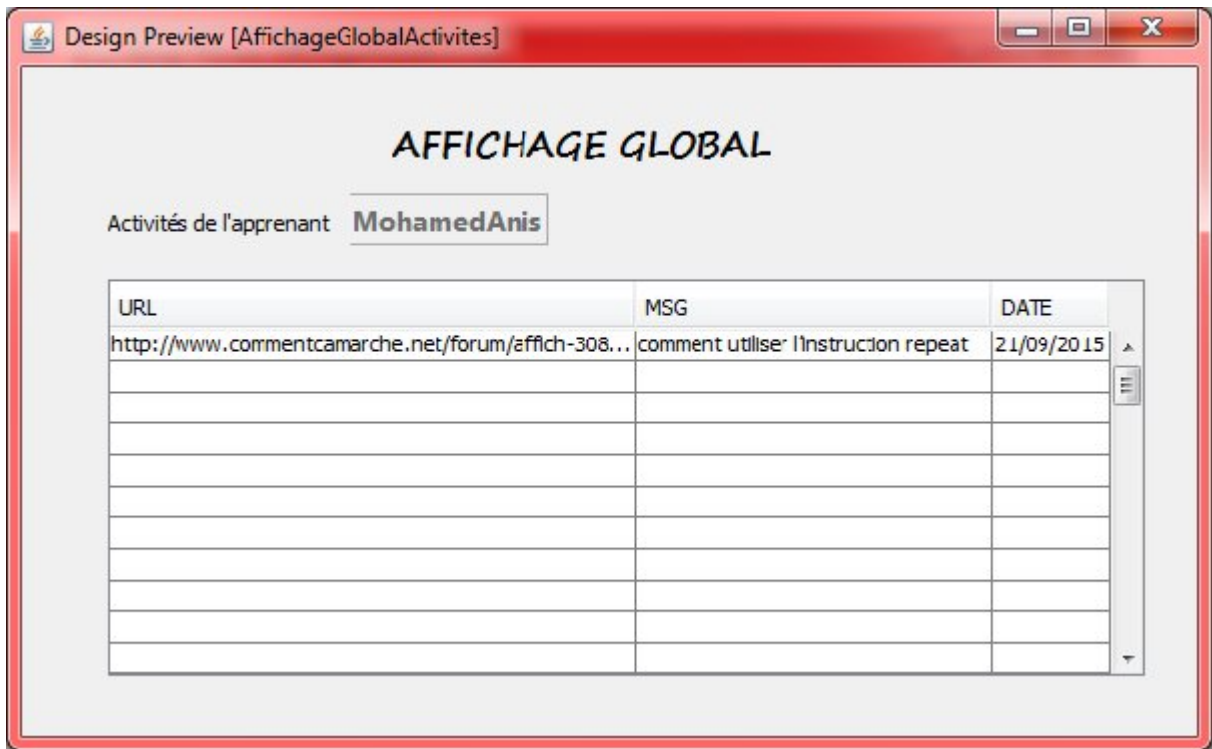


Figure 30 : Capture d'écran de l'application, affichage global des activités

6.1.d Introduction du seuil :

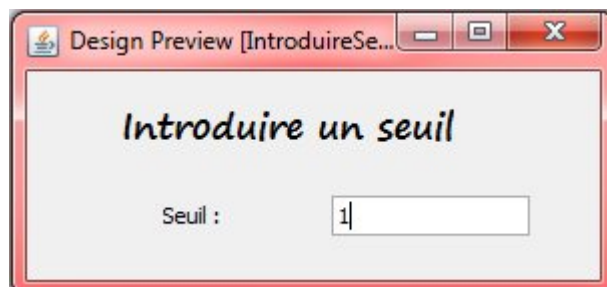


Figure 31 : Capture d'écran de l'application, introduction d'un seuil

En introduisant le seuil, le module **d'indexation et de pondération** est lancé.

Le calcul se fait ensuite pour la détection des difficultés de l'apprenant.

6.2 Affichage des résultats :

Cette étape permet d'indiquer, selon le seuil fixé, les difficultés de l'apprenant ainsi que concepts recherchés.

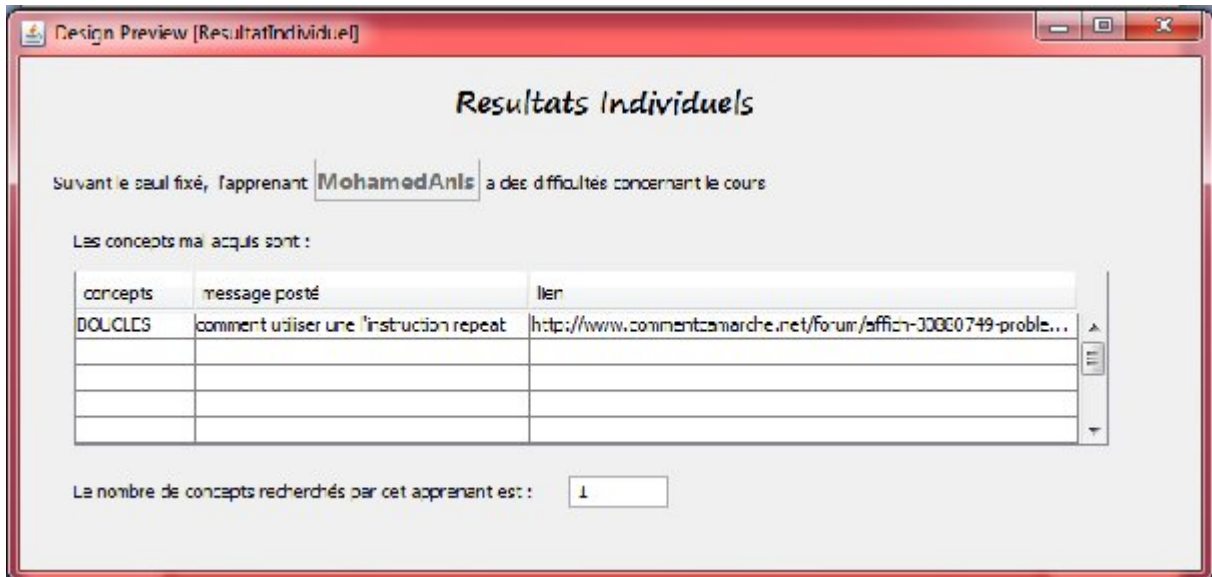


Figure 32 : Capture d'écran de l'application, affichage des résultats individuels

Le tuteur peut aussi observer les concepts mal acquis par tous les apprenants, cela lui permet d'évaluer la section entière ainsi que le cours.

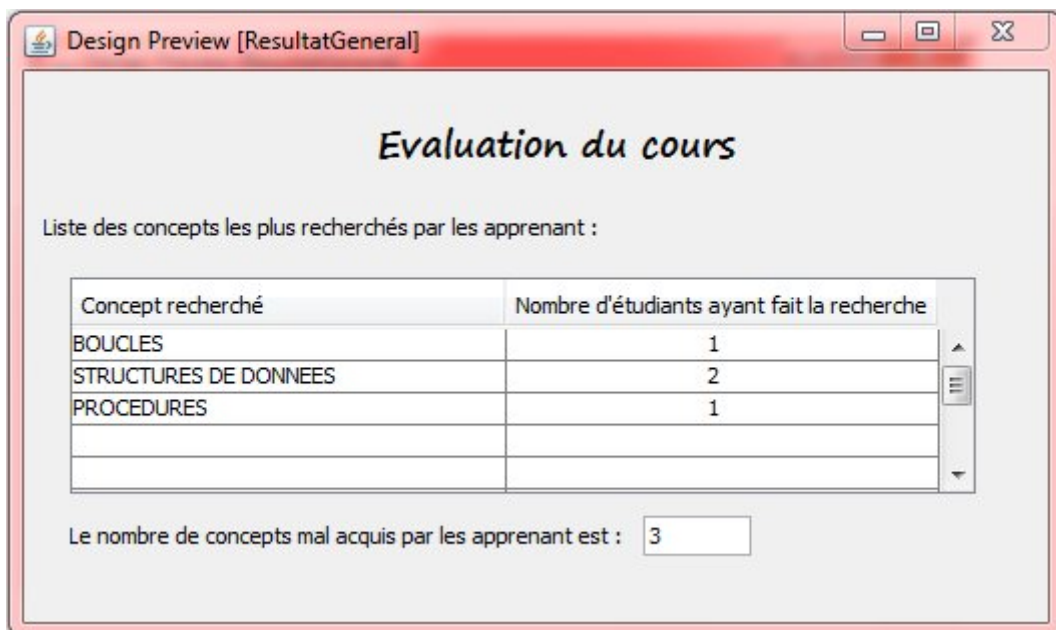


Figure 33 : Capture d'écran de l'application, évaluation du cours

Quelques statistiques sont retournées :

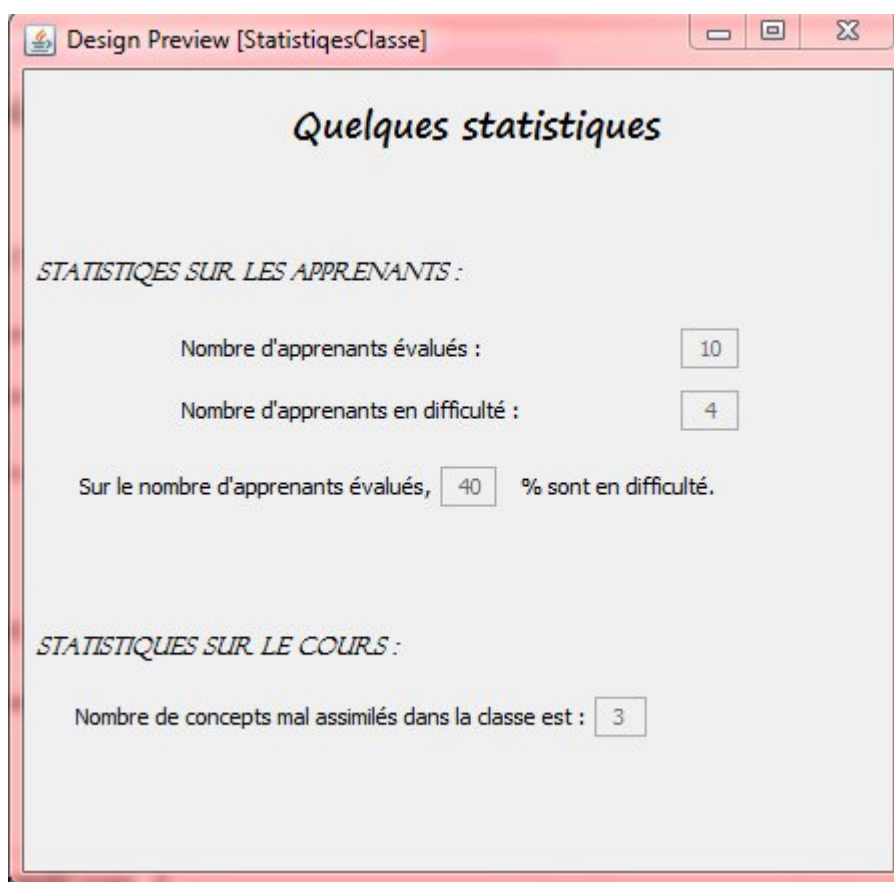


Figure 34 : Capture d'écran de l'application, statistiques générales

Cette fenêtre affiche les étudiants ayant été évalués (que ça soit en difficulté ou non).

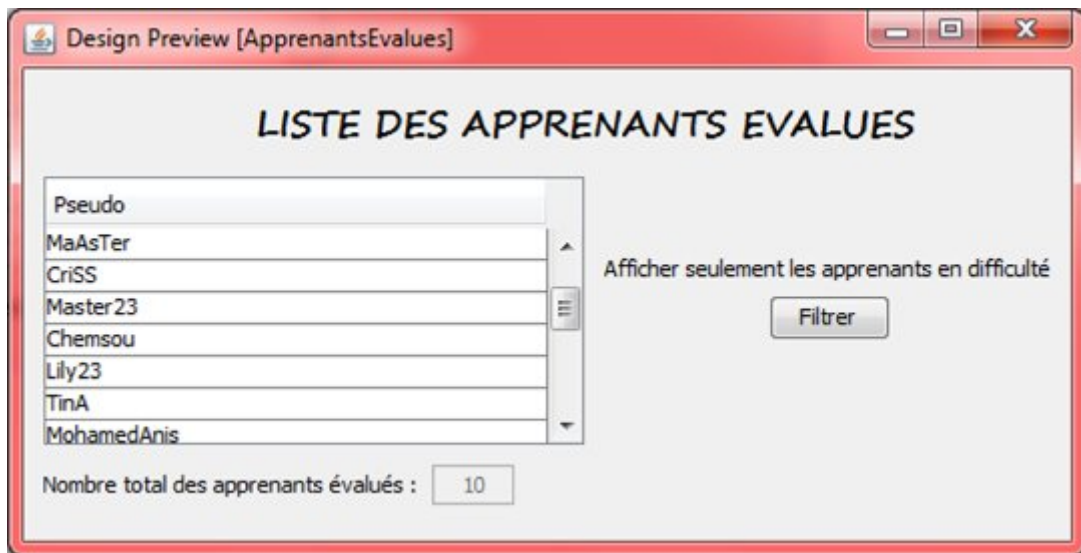


Figure 35 : Capture d'écran de l'application, affichage des apprenants évalués

Le clic sur le bouton 'Filtrer' conduit à afficher uniquement ceux qui sont en difficulté.

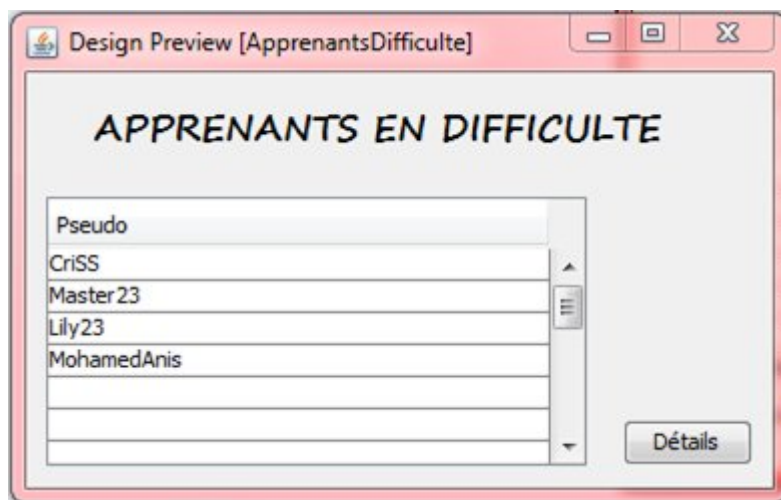


Figure 36 : Capture d'écran de l'application, liste des apprenants en difficulté

Lors de la sélection d'un élément parmi ces derniers, on peut consulter ses détails individuels concernant ses difficultés. (Comme reporté précédemment pour l'apprenant 'MohamedAnis' sur la **Figure 32**)

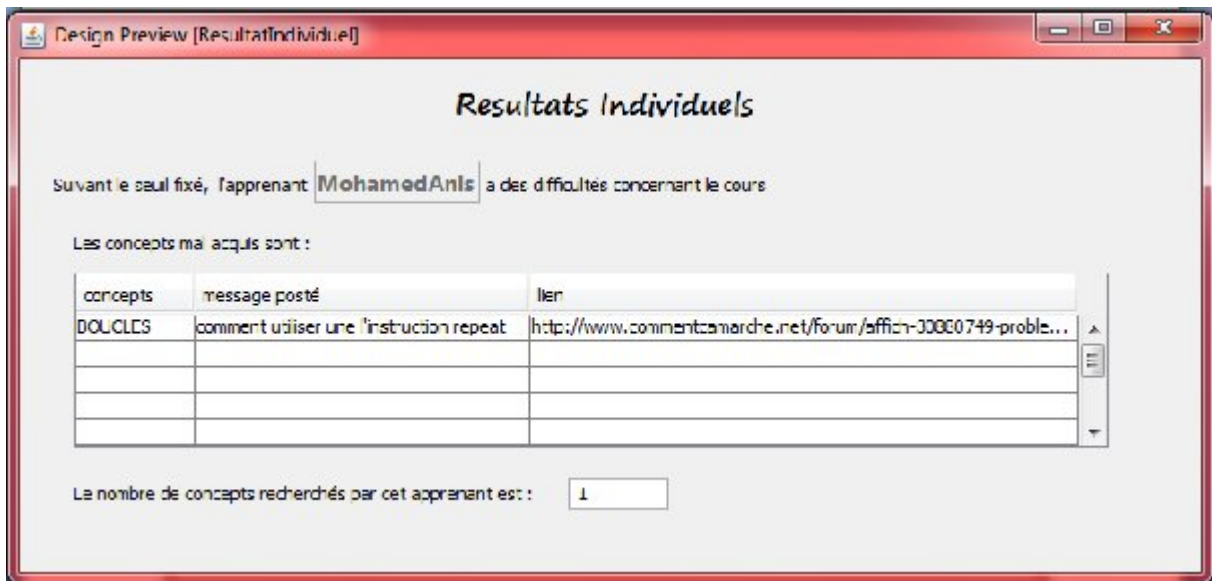


Figure 32 : Capture d'écran de l'application, affichage des résultats individuels

Conclusion :

Dans ce chapitre, nous avons indiqué les différentes étapes, illustrées par des fenêtres, de la réalisation de notre système. Le but est d'arriver à détecter les apprenants en difficulté grâce aux concepts recherchés, notamment lorsque ils postent des messages sur des forums.

Le résultat final est retourné sous forme de statistiques qui permettront à l'utilisateur d'avoir un aperçu global sur les performances des apprenants ainsi que sur le cours édité sur la plateforme.

Conclusion Générale:

L'objectif de notre travail est de créer un système de détection des apprenants en difficultés grâce à leurs messages postés dans de différents forums. Pour aboutir à un travail cohérent, nous avons abordé de différentes notions telles que : le traçage informatique, l'E-Learning, les ontologies, l'indexation ainsi que la recherche d'informations.

Le rôle du traçage dans notre approche est d'avoir un œil (espionner) sur les activités de l'apprenant durant son apprentissage. Nous avons donc utilisé un outil d'espionnage (MiniKeyLog) qui nous a permis d'observer ces activités en tenant en compte la confidentialité de l'apprenant. Toutes les informations récoltées sont ensuite enregistrées dans un fichier XML que l'utilisateur de l'application chargera dans la base de données.

Indexation et la pondération consiste en l'extraction des concepts à partir des messages postés par l'apprenant. Nous avons commencé par extraire les termes, puis trouver leurs concepts correspondants en les projetant dans l'ontologie du domaine représentant le cours déjà construite et finir par produire le vecteur de concepts pondérés. Notre étude n'est pas de construire une ontologie mais celle-ci nous aide pour effectuer l'étape ainsi décrite.

Après le passage par les différentes étapes de développement, l'application a abouti à un logiciel fonctionnel qui répond globalement aux critères imposés dans ce domaine. Le résultat retourné est une vue globale ou individuelle des apprenants en difficultés ainsi de leurs concepts mal assimilés. L'utilisateur pourrait ainsi apporter des modifications sur le cours enseigné et aider les apprenants en difficulté.

Le présent travail nous a permis d'acquérir des connaissances dans le domaine de la programmation web, et de conforter nos connaissances en conception logicielle. Au final, nous espérons que notre recherche va répondre aux souhaits des utilisateurs. D'autres options viendront s'ajouter plus tard pour la performance et une meilleure utilisation de notre système.

Annexe

La lemmatisation :

Les mots (lemmes) d'une langue utilisent plusieurs formes en fonction de leur genre (masculin ou féminin), leur nombre (un ou plusieurs), leur personne (moi, toi, eux...), leur mode (indicatif, impératif...) donnant ainsi naissance à plusieurs formes pour un même lemme.

La lemmatisation d'une forme d'un mot consiste à en prendre sa forme canonique. Celle-ci est définie comme suit :

- pour un verbe : ce verbe à l'infinitif,
- pour les autres mots : le mot au masculin singulier.

Exemples

L'adjectif petit existe sous quatre formes : petit, petite, petits et petites. La forme canonique de tous ces mots est « petit ».

Il existe beaucoup plus de formes du verbe avoir : ai, as, a, avons, ais, avons eu, ayez eu, eussions eu, aurions eu, etc. La forme canonique de eussions eu est avoir.

Usage en informatique

En informatique, il est difficile pour un programme de savoir que eussions eu et avoir sont deux facettes d'un même terme. La lemmatisation est donc une opération préliminaire pour la reconnaissance des mots d'une phrase.

L'étiquetage morpho-syntaxique :

Aussi appelé étiquetage grammatical, est le processus qui consiste à associer aux mots d'un texte les informations grammaticales correspondantes comme le genre et le nombre, etc. à l'aide d'un outil informatique

Exemple :

Texte original : « Nous sommes allées en Bretagne contempler de magnifiques allées couvertes du Néolithique. »

Texte étiqueté :

« Nous/PRO:PER sommes/VER:pres allées/VER:pper en/PRP/en Bretagne/NAM contempler/VER:infi de/PRP magnifiques/ADJ allées/NOM couvertes/VER:pper du/PRP:det Néolithique/NAM ./SENT »

Logiciels :

Les étiqueteurs grammaticaux sont très nombreux surtout pour la langue anglaise. Quelques étiqueteurs sont accessibles avec un modèle pour le français prêt à l'emploi comme le TreeTagger, LIA Tagg du Laboratoire Informatique d'Avignon, Cordial Analyseur de Synapse Développement ou le Stanford Tagger de l'Université Stanford. De nombreux autres logiciels peuvent fonctionner pour le français mais doivent être entraînés sur un corpus français pré-étiqueté.

Jdom et Sax :

SAX est l'acronyme de Simple API for XML. Ce type de parseur utilise des événements pour piloter le traitement d'un fichier XML. Un objet (nommé handler en anglais) doit implémenter des méthodes particulières définies dans une interface de l'API (SAX) pour fournir les traitements à réaliser, ensuite, le parseur appelle ces méthodes selon les événements.

JDOM utilise des collections SAX pour parser les fichiers XML. Acronyme de Document Object Model. C'est une spécification du W3C pour proposer une API qui permet de modéliser, de parcourir et de manipuler un document XML.

Le principal rôle de DOM est de fournir une représentation mémoire d'un document XML sous la forme d'un arbre d'objets et d'en permettre la manipulation (parcours, recherche et mise à jour). À partir de cette représentation (le modèle), DOM propose de parcourir le document mais aussi de pouvoir le modifier. Ce dernier aspect est l'un des aspects les plus intéressants de DOM.

Il est défini pour être indépendant du langage dans lequel il sera implémenté. DOM n'est donc qu'une spécification qui, pour être utilisée, doit être implémentée par un éditeur tiers.

JDOM utilise DOM pour manipuler les éléments d'un Document Object Model spécifique (créé grâce à un constructeur basé sur SAX, SaxBuilder). Il permet de construire des documents, de naviguer dans leur structure, d'ajouter, de modifier, ou de supprimer leur contenu.

Références bibliographiques :

- [1] E-doceo : définition d'une plateforme d'apprentissage.
- [2] Définition donnée lors de la Communication de la Commission du 30 septembre 2005 au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions.
- [3] Digital libraries : definition, issues and challenges. IFLANET UDT occasional paper. N° 8. Mars 1998.
- [4] Henri & al. (2001) : apprentissage collaboratif à distance, presse de l'université du Québec.
- [5] Pemin (2005) : définition d'une trace.
- [6] Jermann (2001) : description des traces.
- [7] Champin & al, 2004 : l'histoire des traces.
- [8] Krämer, 2012 : complication des traces
- [9] Settouti (2005) : modèle à base de trace.
- [10] Arana & al, (2003) : MySQL et base de données.
- [11] Thomas R. Gruber : Définition sur les ontologies.