

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE SAAD DAHLAB DE BLIDA 1

Faculté des Sciences

**Département
d'informatique**



Mémoire Master Ingénierie Logiciel

*Développement d'un PLUGIN d'évaluation automatique des réponses courtes
pour une plateforme de télé-enseignement*

Présenté par :

- ❖ Snoussi El Hareth
- ❖ Madani Ahmed Abderraouf

Sous la direction de :

- ❖ Mme Leila Ouahrani

Mémoire présenté en Juillet 2019 devant un jury composé de :

Lamia Oukid, Présidente du jury

Sana Aroussi, Examinatrice

Promotion 2018/2019

المخلص

يعد التعليم الإلكتروني بلا شك أحد أسرع الاتجاهات نموًا في مجتمع التعليم نظرًا لربح الوقت والتكلفة. يستخدم التعليم الإلكتروني عبر الإنترنت أنظمة إدارة التعلم لدعم وتحسين عملية التعليم/التعلم وتعزيزها. عند مقارنة بعض أنظمة إدارة التعلم المشهورة، نستنتج عدم وجود آليات فعالة لتقييم الاجابات الوصفية مثل الاجابات القصيرة. بالإضافة إلى ذلك، من الممل بالنسبة للمعلمين إجراء تقييم يدوي لعدد كبير من الطلاب. لذلك، فإننا نقترح إضافة مكون إضافي (بلوغين) للتقييم التلقائي للإجابات القصيرة يتم تكييفها مع اللغة العربية بناءً على منهج للتقييم التلقائي من خلال دمجها في منصة التعلم عن بعد مودل وتزويد الطلاب بتقييم فوري. نعتمد نهجاً مطوراً حول قياسات التشابه النحوي والدلالي. ومع ذلك، فإننا نستخدم خادم عن بُعد لتنفيذ هذا النهج والذي بدوره يستقبل اجابات الطلاب والاجابات النموذجية من خلال واجهة لبرمجة التطبيقات لتقييمهم ليتم نشر التقييم في المنصة.

Résumé

L'apprentissage en ligne est, sans contestation, une des tendances du milieu de l'éducation ayant une des plus fortes croissances en raison de l'exploitation du temps et du coût. L'apprentissage en ligne utilise les systèmes de gestion de l'apprentissage LMS (Learning Management Systems) pour soutenir et améliorer le processus enseignement-apprentissage. Lorsqu'on a comparé certains des LMS populaires, un manque de mécanismes efficaces pour évaluer les réponses descriptives telles que les réponses courtes est constaté. En outre, il est fastidieux pour les enseignants d'évaluer manuellement un grand nombre d'étudiants. Nous proposons dès lors un Plugin d'évaluation automatique des réponses courtes adapté à la langue arabe basé sur une approche d'évaluation automatique en l'intégrant dans une plateforme de télé-enseignement Moodle et fournissant aux étudiants un retour immédiat de la note. Nous adoptons une approche développée autour des mesures de similarité syntaxique et sémantique. Nous utilisons un serveur distant PythonAnywhere pour exécuter cette approche qui, à son tour, reçoit des réponses étudiantes et des réponses modèles à travers un Flask API afin de les assigner. Le retour immédiat de la note est publié instantanément dans la plateforme Moodle.

Mot clés : L'apprentissage en ligne, LMS, Plugin, PythonAnywhere, Similarité syntaxique, similarité sémantique, Plateforme, Evaluation automatique des réponses courtes, Flask API.

Abstract

E-learning is unquestionably one of the fastest growing trends in the education community due to the gain of time and cost. E-learning uses systems such as Learning Management Systems (LMS) to support and enhance the teaching-learning process. When comparing some of the popular LMSs, a lack of effective mechanisms for evaluating descriptive answers such as short answers is detected. In addition, it is tedious for teachers to manually evaluate a large number of students. We are therefore proposing a short answer automatic assessment Plugin adapted to the Arabic language based on an automatic assessment approach by integrating it into a Moodle E-Learning platform and providing students with immediate feedback. We adopt a developed approach around syntactic and semantic similarity measurements is made. However, we use a PythonAnywhere remote server to execute this approach which, in turn, receives student answers and model answers through Flask API to assign them. The immediate return of the note is instantly published in the platform.

Keywords: E-Learning, LMS, Plugin, PythonAnywhere, Syntactic similarity, Semantic similarity, Platform, Automatic assessment of short answers, Flask API.

Remerciements

Ce travail n'aurait pas eu l'impact recherché sans le précieux apport, le professionnalisme et les compétences avérées de notre promotrice, Mme L. OUAHRANI, qui n'a tari d'aucun effort pour nous conseiller, nous orienter et nous montrer la voie de la réussite.

Notre gratitude va également à l'Université de Bouira qui nous a fourni l'accès à un serveur distant pour accomplir notre travail.

Nos remerciements vont également au groupe d'étudiants master IL, qui ont accepté de faire le test d'examen en ligne pour l'évaluation qualitative de ce travail.

Cette recherche aurait été incomplète sans l'aide de Y. ATOUB, A. BENAYAD, A. ABDALLAH et K. GAROUDJA, qui nous ont assistés dans la finalisation de ce projet, qui est, en fait, la continuité d'un travail entamé il y a une année.

A vous tous, nous disons : merci !

El Hareth et Abderraouf

Liste des tables

<i>Tableau 1 :représentation des quatre réponses possibles</i>	6
<i>Tableau 2 :représentation des six réponses possibles</i>	7
<i>Tableau 3 : représentation des deux réponses possibles</i>	7
<i>Tableau 4 : représentation des douze réponses possibles</i>	8
<i>Tableau 5 :Quelques exemples sur la syntaxe de Pattern match [3]</i>	11
<i>Tableau 6 :Quelques exemples sur la technique Pattern match dans Moodle [3]</i>	12
<i>Tableau 7 : Fonctions du fichier edit_arabicshortanswer_form.php</i>	52
<i>Tableau 8 :Fonctions du fichier questiontype.php</i>	53
<i>Tableau 9 :Fonctions du fichier question.php</i>	55
<i>Tableau 10 :Fonctions du fichier renderer.php</i>	56
<i>Tableau 11 : Echantillon dataset Cairo University [28]</i>	68
<i>Tableau 12 :Echantillon dataset STS250 SemEval [22]</i>	69
<i>Tableau 13 : Echantillon dataset Cyber Criminality Arabic Short Answer Grading</i>	69
<i>Tableau 14 :Signification des valeurs de corrélation de pearson [23]</i>	70
<i>Tableau 15 :Résultats du modèle CM avec l'espace sémantique Khaleej</i>	71
<i>Tableau 16 :Résultats du modèle CM avec l'espace sémantique CNN</i>	72
<i>Tableau 17 :Résultats de la combinaison (CM,Dice) avec l'espace sémantique Khaleej</i>	72
<i>Tableau 18 :Résultats de la combinaison (CM,Dice) avec l'espace sémantique CNN</i>	73
<i>Tableau 19 :Résultats du modèle SV avec l'espace sémantique Khaleej</i>	73
<i>Tableau 20 :Résultats du modèle CM avec l'espace sémantique CNN</i>	74
<i>Tableau 21 :Résultats de la combinaison (SV,Dice) avec l'espace sémantique Khaleej</i>	74
<i>Tableau 22:Résultats de la combinaison (SV,Dice) avec l'espace sémantique CNN</i>	75
<i>Tableau 23 :Comparaison des résultats de l'algorithme CM entre Stem lourd et léger</i>	76
<i>Tableau 24 :Comparaison des résultats de l'algorithme SV entre Stem lourd et léger</i>	76
<i>Tableau 25 : Comparaison des résultats de la combinaison (CM,Dice) entre Stem lourd et léger</i>	76
<i>Tableau 26 :Comparaison des résultats de la combinaison(SV,Dice) entre Stem lourd et léger</i>	77
<i>Tableau 27 :Résultat de toutes les approches avec l'espace sémantique cybercriminalité</i>	77

Liste des figures

Figure 1 Les ères et les tendances du classement automatique des réponses courtes [1]	2
Figure 2 L'option de cas de sensibilité	13
Figure 3 Exemple de test Moodle	21
Figure 4 Rapport des activités des participants	23
Figure 5 Rapport des notes étudiant d'un test Moodle.....	23
Figure 6 Progrès vu par l'étudiant.....	24
Figure 7 Types de Questions Moodle	24
Figure 8 Question de type Shortanswer de Moodle	25
Figure 9 La forme générale de l'évaluation automatique des réponses courtes[11].....	26
Figure 10 Diagramme d'état de transitions d'un comportement particulier de la tentative de question [13].....	28
Figure 11 SCHEMA DES ETAPES PRINCIPALES [7].....	31
Figure 12 Phases de création de l'espace sémantique.....	32
Figure 13 Algorithme De Coals pour la construction de l'espace sémantique [7]	36
Figure 14 Les phases de conception du Plugin Arabic Short Answer	42
<i>Figure 15 Conception d'un moule pour n'importe quelle approche d'évaluation automatique</i>	43
Figure 16 : Fonctionnalités du Plugin Arabic Short Answer	44
Figure 17 Diagramme de cas d'utilisation « Installation de Plugin	45
Figure 18 Diagramme de cas d'utilisation « Rôle des entités de test ».....	45
Figure 19 Table qtype_arabicshortanswer_options via PhpMyAdmin.....	47
Figure 20 Table qtype_shortanswer_options via PhpMyAdmin.....	47
Figure 21 Modèle E/A de la base de donnée du moteur de question	48
Figure 22 Vu d'ensemble du Plugin « Arabic ShortAnswer ».....	49
Figure 23 Connexion avec le serveur distant PythonAnywhere à travers FLASK API.....	51
Figure 24 CONTENU DU FICHER LANG/AR/QTYPE.ARABICSHORTANSWER.PHP	57
Figure 25 Diagramme de séquence «Scénario: Installer Plugin »	59
Figure 26 La liste de l'élément « Plugins » dans la rubrique « Administration ».....	59
Figure 27 L'affichage d'une nouvelle page web pour glisser le fichier compressé du Plugin	60
Figure 28 Diagramme de séquence de scénario « scénario : Installation du Flask et Intégration de l'approche d'évaluation automatique».....	61
Figure 29 Ajouter une nouvelle application web dans l'onglet « Web »	62
Figure 30 Se rendre dans notre console Bash dans l'onglet « Consoles ».....	62
Figure 31 Les fichiers de notre application FLASK dans l'onglet « Files ».....	63
Figure 32 Le fichier.py principal de FLASK pour intégrer les algorithmes de l'approche sémantique.....	63
Figure 33 Diagramme de séquence « scénario : Faire le test »	64
Figure 34 Les lignes de code en cUrl PHP assurant l'envoi et la récupération des données .	65
Figure 35 Les lignes de code en Python Flask assurant la récupération et l'envoi des données	65

Introduction générale

Dans l'introduction générale, nous présentons le contexte de notre travail, en précisant la problématique de notre travail, ainsi que les objectifs de ce dernier, ses motivations et ses limites.

- **Introduction :**

L'utilisation d'un ordinateur à des fins d'évaluation en matière d'enseignement a considérablement augmenté au cours des dernières années. Les objectifs de l'utilisation de l'évaluation automatique dans l'enseignement comprennent la réalisation et la consolidation des avantages d'un système présentant les caractéristiques suivantes : *Premièrement*, réduire la charge de travail des enseignants en automatisant une partie de la tâche d'évaluation, *Deuxièmement*, fournir aux apprenants des informations détaillées sur leur période d'apprentissage de manière plus efficace que l'évaluation traditionnelle, et, enfin, intégrer la culture d'évaluation au travail quotidien des apprenants dans un environnement d'e-Learning¹. Nous nous intéressons dans ce travail aux questions ouvertes à réponses courtes (quelques mots à quelques phrases construites en **langage naturel**) visant l'avantage de demander à l'apprenant de construire une réponse plutôt que de choisir parmi un certain nombre d'options prédéterminées (rappel et reproduction des connaissances acquises). L'**automatisation** de l'évaluation de ces réponses n'est pas simple en raison de variations linguistiques (une réponse donnée pourrait être articulée de différentes façons), nature subjective de l'évaluation (multiples réponses possibles), manque de cohérence dans la notation humaine, etc. Le travail est orienté vers la **langue arabe** qui, bien que largement utilisée aujourd'hui, n'a pas encore bénéficié de recherche et de résultats matures dans le domaine de l'évaluation automatique.

- **Problématique :**

Le principe général de l'évaluation consiste à comparer la réponse de l'apprenant avec la réponse **modèle** (de référence) formulée par l'enseignant et d'attribuer un score. Dans le cadre du projet actuel, plusieurs approches ont été déjà développées tournant sur des applications web. Dans la continuité des objectifs, le **passage à l'échelle dans une situation réelle** doit passer par l'intégration des approches dans une plateforme de télé-enseignement afin que l'évaluation puisse se faire en consolidation avec les cours qui se déroulent sur la même plateforme.

¹ Le e-learning -ou apprentissage en ligne- est une méthode d'apprentissage qui repose sur la mise en disposition des ressources pédagogiques à travers un support électronique

- **Motivations / Objectif :**

- ✓ Développement d'un PLUGIN ² (Arabic Short Answer Question Type Plugin) bénéficiant de toutes les fonctionnalités de la plateforme Moodle tel que l'historique du test, l'historique du Feedback, l'assignement des notes du test et bien d'autres fonctionnalités déjà disponible d'une plateforme de télé-enseignement .
- ✓ Conception d'un PLUGIN présentant un moule pour n'importe quelle approche d'évaluation automatique.
- ✓ Intégration d'un PLUGIN dans le module Quiz (test) de Moodle.
- ✓ Possibilité de combiner notre nouveau type de question avec les autres types de Moodle dans un test.
- ✓ Intégrer l'évaluation automatique des réponses courtes dans la plateforme de télé-enseignement basée sur les approches sémantiques.
- ✓ Offrir à l'enseignant la possibilité de construire son test de manière naturelle et uniforme en combinant différents types de questions déjà intégrées sur la plateforme (choix multiple, essais, remplir les vides, ...) avec le nouveau type d'évaluation proposée pour la langue arabe.
- ✓ Familiariser les apprenants à un environnement de test et au feedback consolidés sur la même plateforme de cours et offrir ainsi une navigation et utilisation facile et familière du PLUGIN auprès des enseignants et étudiants.
- ✓ Evaluer les approches déjà développées par rapport à un passage à l'échelle.

- **Expérimentation :**

Nous travaillons dans le contexte de la plateforme **Moodle** et tenterons d'ajouter à cette dernière un module (Plugin) contenant les outils d'évaluation automatique des réponses courtes en langue arabe.

- **Limites de notre travail :**

- Dans ce travail, nous avons utilisé un serveur web distant PythonAnywhere qui accepte le langage Python seulement.
- Ce serveur web contient un espace disque de 500 méga octets, si l'approche générée par un espace sémantique dépasse la taille du disque, nous devons passer par le paiement en ligne.
- Ce travail traite de la langue arabe. Ainsi, toute ressource, mot d'une langue autre que l'arabe est ignoré.
- Dans la suite de notre travail nous considérons qu'une réponse est « réponse courte » si elle respecte les propriétés suivantes :

² Un plugin est un outil composé d'un ensemble de fichiers informatiques, et qui permet d'installer des nouvelles fonctionnalités en marge d'un logiciel auquel il est rattaché.

- La réponse doit être rédigée dans un langage naturel.
- Elle doit être issue (formulée) de connaissances externes à la question elle-même.
- Sa longueur est de l'ordre d'une vingtaine de mots.
- La réponse valorise le contenu et non pas le style d'écriture.

TABLE DES MATIERES

CHAPITRE I : ETAT DE L'ART	2
INTRODUCTION :.....	2
I. UNE VUE SUR L'HISTORIQUE DES ASAG [1].....	3
II. LES APPROCHES ET LES OUTILS D'EVALUATION AUTOMATIQUE DES REPONSES COURTES DANS LES PLATEFORMES DE TELE-ENSEIGNEMENT :	4
III. LES APPROCHES ET LES OUTILS D'EVALUATION AUTOMATIQUE DES REPONSES COURTES DANS LA PLATEFORME DE TELE-ENSEIGNEMENT MOODLE :	6
1. <i>Les expressions régulières</i> [2] :.....	6
2- <i>La correspondance de modèle (Pattern match)</i> [3] :	10
3- <i>Le Preg Match</i> [4]:.....	13
IV. ÉVALUATION DES SYSTEMES ACTUELS :	15
<i>Les obstacles de la langue arabe dans le contexte de l'évaluation automatique</i>	16
V. LES TRAVAUX CONNEXES A NOTRE RECHERCHE :.....	17
CONCLUSION :	19
CHAPITRE II : PRESENTATION DU SYSTEME DE QUIZ DE LA PLATEFORME MOODLE [12]..	20
INTRODUCTION :.....	20
1- PRESENTATION GENERALE DE LA PLATEFORME.....	20
2- FONCTIONNALITES DE MOODLE :	21
3- LA MODULARITE SUR MOODLE :.....	22
4- LA FORME GENERALE DE SYSTEME DE L'EVALUATION AUTOMATIQUE DES REPONSES COURTES :	27
5- VUE D'ENSEMBLE SUR LE MOTEUR DE QUESTION MOODLE (QUESTION ENGINE) [13] :	28
CONCLUSION :	31
CHAPITRE III : APPROCHE D'EVALUATION AUTOMATIQUE DEPLOYEE	32
1. ETAPE 1 : CONSTRUCTION DE L'ESPACE SEMANTIQUE :	33
2. ETAPE 2 : MODELES DU CALCUL DE SIMILARITE SEMANTIQUE/SYNTAXIQUE ENTRE DEUX REPONSES COURTES :.....	40
3. ETAPE 3 : PASSAGE AU SCORE :	41
CHAPITRE IV : CONCEPTION, DEVELOPPEMENT ET INTEGRATION DU PLUGIN DANS LE SYSTEME DE QUIZ DE MOODLE	43
INTRODUCTION :.....	43
1. CONCEPTION DU PLUGIN D'EVALUATION AUTOMATIQUE DES REPONSES COURTES :.....	44
2. DEVELOPPEMENT DU PLUGIN D'EVALUATION AUTOMATIQUE DES REPONSES COURTES :.....	53
2.1 <i>Modèle de plugin de type question</i> :	53
2.2 <i>Disposition du fichier du Plugin</i> :.....	54
3. INSTALLATION DU PLUGIN ET INTEGRATION DES OUTILS D'EVALUATION AUTOMATIQUE DES REPONSES COURTES :.....	60
3.1. <i>Intégration et installation du Plugin</i> :.....	60
3.2 <i>Installation du Framework FLASK et Intégration de l'approche d'évaluation automatique dans le serveur PythonAnywhere</i> :	62
3.3 <i>Limites du serveur PythonAnywhere</i> :.....	67
CONCLUSION :	68
CHAPITRE V : RESULTATS EXPERIMENTAUX ET ÉVALUATION	69
1- ENSEMBLE DE DONNEES (DATA SET) :	69

2- METRIQUES D'EVALUATION :	72
3- RESULTATS ET DISCUSSION :	73
4- L'EVALUATION QUALITATIVE :	80
CONCLUSION GENERALE	82
BIBLIOGRAPHIE.....	83
ANNEXE.....	85

Chapitre I : Etat de l'art

Ce chapitre expose les outils et approches d'évaluation automatiques des réponses courtes ainsi que les techniques d'évaluation des réponses courtes déjà existantes sur les plateformes de télé-enseignement, et enfin les travaux connexes à notre travail.

Introduction :

Le e-learning ou l'apprentissage en ligne, est un système d'apprentissage basé sur un enseignement formalisé à l'aide de ressources électroniques. Bien que l'enseignement puisse se dérouler dans ou hors des salles de classe, l'utilisation des ordinateurs et d'Internet constitue la principale composante du e-learning. L'apprentissage peut également être qualifié de transfert de compétences et de connaissances par le biais d'un réseau à un grand nombre de destinataires à un moment identique ou différent.

L'apprentissage en ligne permet aux enseignants de rester en contact avec leurs élèves en dehors des heures de classe afin d'échanger des ressources et des idées, ainsi que l'utilisation des nouvelles technologies donne aux enseignants la liberté d'expérimenter leur pratique pédagogique par rapport à la grande variété de ressources, tels que des vidéos, des présentations et des quiz, qu'ils peuvent utiliser pour adapter leurs méthodes de tutorat aux styles d'apprentissage de leurs élèves.

Il existe des dizaines de modalités dans la catégorie du e-learning tel que les MOOCs (pour Massive Open Online Course, en anglais) ou bien formation en ligne ouverte à tous, aussi appelée cours en ligne ouvert et massif, est un type ouvert de formation à distance capable d'accueillir un grand nombre de participants.

Les MOOCs proposent également des formations très spécifiques, notamment en informatique, en gestion, en design, etc. Ils sont devenus d'une grande importance dans l'éducation. Les chercheurs ont essayé de concevoir et de compléter les outils automatisés d'évaluation pour les réponses libres, c'est-à-dire ceux rédigés par l'examineur d'une façon complète, et jusqu'à maintenant les recherches et les réalisations n'ont pas atteint les objectifs souhaités que partiellement en raison de l'existence de complications problématiques et importantes dans ce domaine, qui fait partie de la compréhension automatisée du domaine des langues humaines.

Dans le domaine des réponses libres et pour faire la différence entre les deux types de grandes réponses : Les réponses qui contiennent plusieurs lignes. Les courtes réponses qui sont pour la plupart composées d'une ligne, et parfois deux ou trois lignes. Et clair que les deux défis dans deux cas différents, des expressions et des idées ont des mots de plus en plus

complicés dans les longues réponses. Comme réponses courtes sont souvent centrées autour d'une idée ou deux idées dans lesquelles l'installation des mots est d'un nombre limité. Dans le domaine de la correction automatique des réponses courtes ou longues, il existe des outils qui sont basés sur le taux de similarité entre les réponses modèles et les réponses de l'étudiant. Les réponses modèles sont organisées dans un panier de réponses (Banque de questions), ce qui contient des réponses correctes et partiellement correctes puisque chaque réponse de cette banque comportant une valeur ou un score indiquant son évaluation. Dans une question à réponse courte, l'étudiant saisit un mot ou une proposition en réponse à une question. Les réponses peuvent être sensibles ou non à la casse. La réponse peut être un mot ou une expression, mais elle doit correspondre exactement à l'une des réponses acceptables. La bonne réponse devrait être aussi courte que possible afin d'éviter les erreurs.

I. Une vue sur l'historique des ASAG [1]

L'évaluation automatique des réponses courtes (ASAG) consiste à évaluer les réponses courtes en langage naturel à des questions objectives à l'aide de méthodes de calcul. La recherche active dans ce domaine a considérablement augmenté ces derniers temps avec plus de 80 articles répondant à une définition d'ASAG avec 35 systèmes dans 5 méthodes différentes identifié par Steven Burrows et al. dans leur article [1].

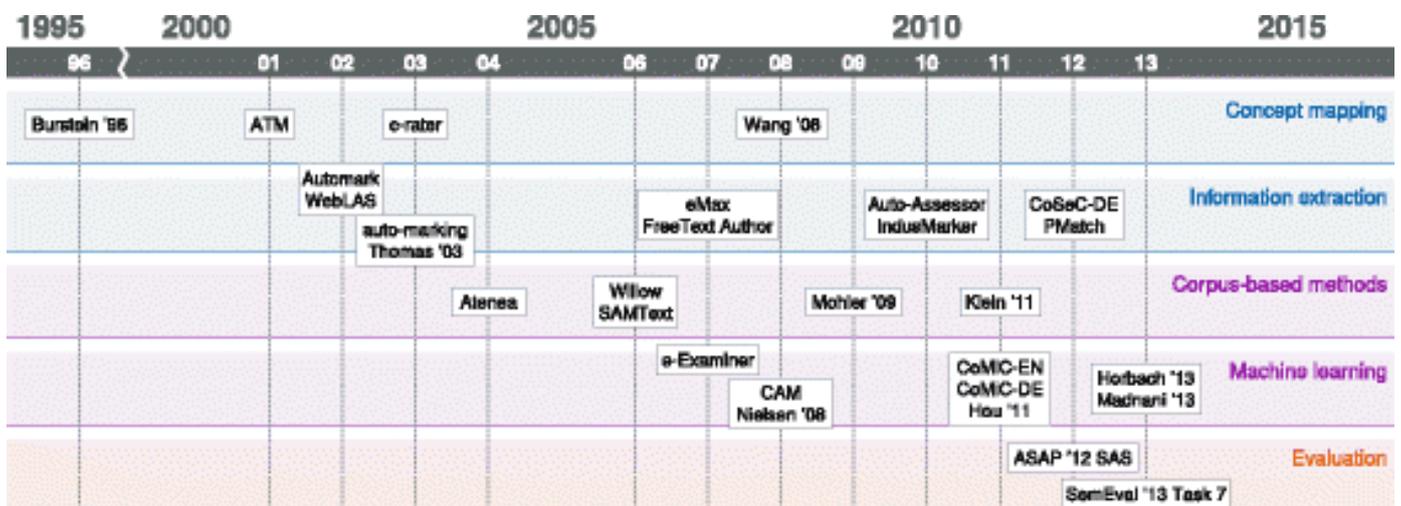


FIGURE 1 LES ERES ET LES TENDANCES DU CLASSEMENT AUTOMATIQUE DES REPONSES COURTES [1]

- **Les méthodes basées sur le mappage de concepts (Concept Mapping) :**

Le principe ici est de considérer les réponses des élèves comme constituées de plusieurs concepts et de détecter la présence ou l'absence du concept clé lors du classement. Cela se fait en parcourant les concepts clé (ceux de la réponse modèle) un par un dans la réponse étudiant et pour chaque concept trouvé, incrémenter le score.

Cette méthode s'adapte bien avec deux types de question. L'un demande une solution à un problème plus une justification. L'autre demande plusieurs explications au même problème.

- **Les méthodes basées sur l'extraction d'information (Information Extraction) :**

Le concept de cette méthode est l'extraction des données à partir des sources non structurées comme les textes libres, ensuite les mettre sous forme structurée (arbres d'analyse...). Les méthodes d'extraction d'informations peuvent être considérées comme une série d'opérations de correspondance de modèle telles que les expressions régulières ou les arbres d'analyse. Dans le cas des systèmes des réponses courtes, chaque réponse est présentée sous forme structurée et par la suite évaluer leurs dépendances.

- **Les méthodes basées-corpus (Corpus Based Methods) :**

Le principe est d'utiliser des propriétés statistiques des corpus, qui sont des ensembles de textes. Ces méthodes peuvent être utiles lors de l'interprétation des synonymes dans les réponses courtes. Alors, afin de limiter les réponses correctes qui peuvent être identifiées, utiliser seulement le vocabulaire des réponses modèles. Ensuite, pour renforcer ce vocabulaire, prendre en considération ses synonymes et effectuer des traductions en autre langue afin d'éviter les difficultés de la langue source. C'est dans ce contexte que se situe notre travail.

- **Les méthodes basées sur l'apprentissage automatique (Machine learning) :**

Les systèmes d'apprentissage automatique utilisent généralement des données étiquetées qui sont les réponses modèles notées ainsi qu'un certain nombre de mesures extraites du langage naturel (techniques de traitement et similarité) qui sont ensuite combinées afin d'avoir un score en utilisant un modèle (fonction d'estimation) de classification ou de régression. Cela peut être soutenu par une boîte à outils d'apprentissage automatique telle que Weka

II. Les approches et les outils d'évaluation automatique des réponses courtes dans les plateformes de télé-enseignement :

Une plateforme d'apprentissage appelée aussi LMS³ (Learning Management System) ou plateforme e-learning est un logiciel qui assiste la conduite des formations ouvertes et à distance. Ce type de logiciel regroupe les outils nécessaires aux trois principaux utilisateurs - formateur, apprenant, administrateur - d'un dispositif qui a pour premières finalités la consultation à distance de contenus pédagogiques, l'individualisation de l'apprentissage et le télé tutorat. Plateformes de télé-enseignements comme **Blackboard learn**⁴ qui est un système de gestion d'apprentissage centré sur l'étudiant qui améliore les tâches d'enseignement de base. Il propose une version gratuite pour l'évaluation. Il concerne la gestion et la prestation de cours personnalisables. Il fournit également une structure en ligne intuitive pour que les instructeurs organisent les cours. Les types de questions d'évaluation de **Blackboard learn** comprennent :

- Questions de format d'essai, questions sur le téléchargement de fichiers, points vides, questions correspondantes, questions de commande, phrases confuses, choix multiple, bol à questions, questions à réponse courte, questions vraie/faux, bloc aléatoire, Formulaire de question Pool.
- Avantages : interface simple et navigation simple, encouragement des interactions entre étudiants, gestion simple du contenu de l'apprentissage, système de gestion de fichiers simple, portfolios électroniques personnels, collaboration instantanée, convivialité pour les mobiles.
- Limites : Absence de notifications, possibilité d'auto-évaluation et d'évaluation par les pairs, mais l'évaluateur doit classer manuellement les réponses descriptives, y compris les questions de type manuel.

Il existe aussi la plateforme de télé-enseignement Desire2Learn⁵ qui est un LMS institutionnel. C'est un environnement d'apprentissage pour les cours en ligne et mixtes. Ceci est principalement utilisé pour la distribution de documents, notes de cours, mises à jour de notes, etc. La technologie sous-jacente est Microsoft.Net. Il fournit une interface organisée et une vue feuille de calcul en classification.

- Les types de questions d'évaluation dans Desire2Learn sont les suivants : questions image, questions à réponse longue, questions à réponse courte, questions correspondantes, choix multiple, questions vrai/faux, sélection multiple, type de commande, questions arithmétiques et chiffres significatifs.
- Avantages : interface utilisateur conviviale, accessibilité, environnement collaboratif, flexibilité, etc.
- Limitations : problèmes de notification, difficulté à télécharger et télécharger des notes à exceller, évaluation manuelle des réponses descriptives.

³ (Learning Management System) est une application utilisée pour planifier, mettre en œuvre et évaluer un processus d'apprentissage spécifique.

⁴ <https://www.blackboard.com/blackboard-learn/index.html>

⁵ <https://www.d2l.com>

III. Les approches et les outils d'évaluation automatique des réponses courtes dans la plateforme de télé-enseignement MOODLE :

Moodle est l'une des plateformes de télé-enseignement les plus utilisées au monde, la précision de la réponse attendue rend les enseignants confrontés à de nombreuses difficultés lors de l'utilisation de ces outils, tels que ceux rencontrés dans cette plateforme et dans la réponse courte : Si la réponse de l'élève contient un caractère espace supplémentaire, cette réponse devient fautive. Afin de rendre ces outils relativement efficaces, le professeur utilise un certain nombre de commandements, soit individuellement, soit dans un livret. L'élève doit observer scrupuleusement les commandements et prêter une attention particulière à ce qu'il écrit pour ne pas laisser aux lettres (ou aux lettres manquantes) le motif de l'annulation de la réponse. Parmi les commandements trouvés, par exemple, la nécessité d'utiliser un seul caractère d'espacement entre les mots sans utiliser la composition, le livret peut contenir des recommandations issues de la recommandation du test-orienté pour le succès du test électronique de la programmation de la classification par exemple en langage Java. Ces outils sont relativement efficaces lorsqu'il s'agit de vérifier une syntaxe stricte telle que celle du Coran et des hadiths, la poésie et les instructions de programmes informatiques. Pour réduire le problème des caractères supplémentaires et des fautes d'orthographe sur la plate-forme Moodle, une technique peut être utilisée qui s'agit d'employer l'astérisque (*) comme joker pour indiquer que n'importe quel caractère est accepté à cet endroit. Par exemple, en écrivant `ك*ب` pour accepter tout mot ou toute proposition commençant par `ك` et se terminant par `ب`

Si une question exige deux réponses `الشرب و الأكل`, le nombre de variantes doit être limité en écrivant ceci :

`الأكل*الشرب`

Cela acceptera les réponses « الأكل الشرب », « الأكل, الشرب », « الأكل ; الشرب », « الأكل / الشرب », « الأكل أو الشرب », « والشرب », « الأكل/الشرب ». Toutefois, cela acceptera aussi les réponses « الأكل أو الشرب », « الأكل|الشرب », et « الأكل بلا الشرب », ce qui pourrait être moins bon, mais ce n'est pas un système absolument sans risques !

1. Les expressions régulières [2] :

Certains outils sont plus flexibles et permettent l'identification de réponses à l'aide de l'expression régulière. Cette dernière est une méthode utilisée pour décrire le texte à rechercher, le modèle et la correspondance dans le texte ou les phrases, pour les comparer aussi, les utiliser en vérifiant la validité. Par exemple, si tous les espaces sont supprimés d'une phrase à l'aide des expressions standards, une seule phrase ne dépassant pas une ligne peut supprimer tous les caractères des espaces. Les expressions typiques raccourcissent le temps et les efforts, en particulier lors du traitement de textes volumineux, et de leurs objectifs pour rassembler des réponses alternatives et

enrichissantes en une seule expression. L'application de cette technologie se fait par de nombreux langages tels que PHP, JavaScript, et d'autres. À l'instar de la question à réponse courte, les expressions régulières des questions à réponse courte s'attendent à ce que le répondant réponde à une question "ouverte" par un mot ou une phrase courte. Cependant, le système des expressions régulières donne accès à un système plus puissant d'analyse des réponses de l'élève dans le but de fournir une rétroaction immédiate plus pertinente.

1.1- Réponse courte correspondant à un modèle d'expression régulière :

A. Premier exemple

Il n'est pas possible de donner des exemples complets des vastes possibilités offertes par ce système, et voici quelques possibilités. Dans un premier exemple, supposons que la question soit " ما هم أسماء أولاد نبينا محمد صلى الله عليه وسلم".

- Dans la zone Réponse 1, la "meilleure" réponse sera assignée, par exemple :

انهم القاسم, إبراهيم, عبد الله

- Dans la zone Réponse 2, cette expression régulière est comme suit :

(إنهم|إنهم القاسم, إبراهيم (و) عبد الله

N° Réponse	La réponse
Réponse 1	انهم القاسم, إبراهيم, عبد الله.
Réponse 2	انهم القاسم, إبراهيم و عبد الله.
Réponse 3	إنهم القاسم, إبراهيم, عبد الله.
Réponse 4	إنهم القاسم, إبراهيم و عبد الله.

TABLEAU 1 : REPRESENTATION DES QUATRE REPNSES POSSIBLES

B. Deuxième exemple :

Question : " ما هو الحيوان الذي يصطاد فريسته في مجموعة؟".

On aura comme réponses :

Réponse 1 : الضبع

Réponse 2 : انه(م) | الضب؟(ا)ع

Et Cela correspond à l'une de ces 6 réponses :

N° Réponse	La réponse
Réponse 1	الضبع
Réponse 2	الضباع
Réponse 3	انه الضبع
Réponse 4	انهم الضبع
Réponse 5	انه الضباع
Réponse 6	انهم الضباع

TABLEAU 2 : REPRESENTATION DES SIX REponses POSSIBLES

Note. - Le début de cette expression rationnelle "(انه(م|)" correspondra les deux cas suivants : vide ou "انه" ou "انهم" . Dans "الضبع(؟)", le point d'interrogation signifie : le caractère précédent (ou le groupe de caractères entre parenthèses) zéro ou une fois ; il est utilisé ici pour correspondre surtout à l'anglais britannique et à l'orthographe américaine.

C. Troisième exemple :

Question : " كلمة تتكون من ثلاثة أحرف وحرف البداية هو الحرف ج و حرف النهاية هو حرف ل".

Réponse 1 : "جبل"

Réponse 2 : "ج[بم]ل"

N° Réponse	La réponse
Réponse 1	"جبل"
Réponse 2	"جمل" ou "جبل"

TABLEAU 3 : REPRESENTATION DES DEUX REponses POSSIBLES

En remarquant dans la syntaxe des expressions régulières, l'inclusion de caractères entre crochets signifie que TOUS ces caractères peuvent être utilisés. Ainsi, dans l'exemple ci-dessus, l'expression régulière "[بم]" est l'équivalent exact de (م | ب). Interdiction d'inclure le caractère de canalisation comme séparateur dans les expressions [...] régulières. Par exemple, "[م | ب]" NE FONCTIONNERA PAS CORRECTEMENT.

D. Quatrième exemple

La fonctionnalité « permutation » (introduite dans la version Regexp 2012102900 pour Moodle 2.3+).

Dans la syntaxe Expressions régulières, un certain nombre de caractères spéciaux ou méta-caractères ont des fonctions spéciales mais il est possible de forcer ces caractères spéciaux à être interprétés comme des caractères normaux (ou littéraux) en les préférant avec un caractère dit caractère d'échappement, la barre oblique inversée "\". Ci-dessous une liste (partielle) de ces méta-caractères est comme présentée comme suit : . ^ \$ * () [] +? | {} \ /

1.3- Les réponses courtes acceptées :

Les réponses courtes acceptées sont des réponses dont la note est supérieure à zéro, c'est-à-dire sont totalement (note = 100%) ou partiellement (note > 0% < 100%) des réponses correctes. Dans ces réponses courtes, si en utilisant un ou plusieurs méta-caractères pour leur valeur littérale, ces derniers devront être échappés (c'est-à-dire les précéder d'une barre oblique inverse).

Exemple 1 :

" هذا الكمبيوتر يكلف \$ 1000 في الجزائر "

Si en souhaitant accepter la réponse courte au-dessus, en considérant le signe \$ comme une partie de la réponse et non pas un méta-caractère. Il doit être échappé de cette manière :

" هذا الكمبيوتر يكلف \\$ 1000 في الجزائر "

1.4- La validation des réponses courtes :

Lorsqu'une question est valide, le moteur de questions vérifie la validité d'expression régulière, en fonction des fonctionnalités expliquées ci-dessus. Si une erreur est détectée, un message d'erreur est affiché au-dessus de la ou des réponses erronées.

1.5- Inconvénients :

Cette approche reste toujours limitée car elle est basée sur la correspondance des lettres ; Si l'étudiant oublie une lettre ou subit à une faute de frappe, sa réponse se diffère de la réponse modèle de l'enseignant et devient fautive. En plus de ça, les enseignants ont du mal à retenir tous les méta-caractères qui existent dans le système des expressions régulières ce qui rend la saisie des réponses modèles compliquée. L'approche des expressions régulières ne couvre pas aussi la convergence du sens entre les deux réponses puisqu'elle se base sur une analyse purement syntaxique.

2- La correspondance de modèle (Pattern match) [3] :

C'est une autre technique d'évaluation des réponses courtes qui est répandue dans les plateformes de télé-enseignement comme Moodle, Les questions de correspondance de modèle permettent à l'élève de donner une réponse d'environ 20 mots, qui peuvent ensuite être notés automatiquement en faisant correspondre la réponse de l'élève à un certain nombre de modèles différents exprimés dans la syntaxe PMatch. Le premier motif correspondant détermine le score et le retour.

Le Pattern-match est une alternative plus sophistiquée au type de question à réponse courte et offre la capacité de corriger les fautes d'orthographe, avec et sans dictionnaire, la spécification de synonymes et de phrases alternatives, l'ordre des mots flexible et la vérification de la proximité des mots.

2.1- Fonctionnement de Pattern match :

La correspondance de modèle fonctionne sur la base du fait qu'on souhaite associer une réponse d'élève à un nombre quelconque de modèles de correspondance de réponse. Chaque motif est comparé jusqu'à ce qu'une correspondance soit trouvée et que des commentaires et des marques soient attribués. La clé de l'utilisation de la correspondance de modèle consiste à poser des questions qu'on peut espérer marquer correctement. Par conséquent, l'écriture de la racine de la question est la partie la plus importante de l'écriture de ces questions.

En revanche, les réponses courtes doivent être bien traitées avant que l'utilisateur qui est le professeur, les fait entrer dans la palette des options des réponses courtes dans la plateforme de télé-enseignement, si la réponse est supérieure à 20 mots, il faut limiter les réponses à 20 mots.

La vérification de l'orthographe d'un élève est aussi importante, vu combien de façons un mot en langue riche comme l'arabe peut être épilé, la vérification de l'orthographe dans la correspondance de modèle améliore la précision du marquage en insistant sur les mots figurant dans le dictionnaire système Moodle.

Dans le Match pattern (correspondance de modèle), les mots sont définis comme des séquences de caractères entre espaces. Le point d'exclamation (!) et le point d'interrogation (?) sont également pris pour marquer la fin d'un mot. Le point d'arrêt (.) est un cas particulier ; en tant que point final, il s'agit également d'un séparateur de mots, mais il ne l'est pas pour la virgule décimale.

Toute autre ponctuation est considérée comme faisant partie de la réponse, mais cette option permet de la supprimer. Tous les mots et synonymes sont spécifiés car ils doivent être appliqués par la correspondance de réponse. Ils ne doivent pas nécessairement être des mots entiers, mais peuvent être des tiges avec un caractère générique. Les synonymes ne peuvent être que des mots simples, c'est-à-dire que la possibilité de spécifier des expressions alternatives dans les listes de synonymes n'est pas autorisée.

Dans l'exemple ci-dessus, toute occurrence du mot المتن dans la réponse sera remplacée par * الصلب | * الصميم | المتن avant la correspondance. La correspondance de modèle prend en charge aussi la correspondance de la réponse complexe. Prenons le premier champ de réponse match_o(بغل حمار حصان جمل) est la correspondance exacte pour les quatre mots avec la fonctionnalité supplémentaire que l'option de correspondance 'o' permet aux mots d'être donnés dans n'importe quel ordre. Le champ de réponse match_ow(بغل حمار حصان جمل) requiert les quatre mêmes mots, toujours dans n'importe quel ordre, mais autorise également

d'autres mots. Le troisième champ de réponse `match_mow(بغل حمار حصان جمل)` autorise les fautes d'orthographe qui sont encore dans le dictionnaire, par exemple. `جبل` au lieu de `جمل`.

2.2- Syntaxe de correspondance de modèle (Pattern Match)

La syntaxe de correspondance peut être considérée en trois parties :

Les options correspondantes, les mots à faire correspondre, par exemple, `محمد مصطفى` avec les caractères spéciaux dans le mot. Et des combinaisons de correspondance, par ex. `match_any()`. Dans le tableau 5, quelques exemples sur la syntaxe de Pattern Match sont présentés.

Options de correspondance	symbole	Description
<code>allowExtraChars</code>	<code>c</code>	Les caractères supplémentaires peuvent être n'importe où dans le mot. Les auteurs sont supposés omettre <code>allowExtraChars</code> lorsqu'ils utilisent l'option mal orthographiée.
<code>allowAnyWordOrder</code>	<code>o</code>	Lorsque plusieurs mots doivent être appariés, ils peuvent être dans n'importe quel ordre.
<code>allowExtraWords</code>	<code>w</code>	Les mots supplémentaires autres que ceux recherchés sont acceptés.
Mal orthographié : <code>allowReplaceChar</code>	<code>mr</code>	Correspondra à un mot où un caractère est différent de ceux spécifiés dans le modèle. Le mot de modèle doit comporter 4 caractères ou plus, caractères génériques exclus, pour que le remplacement puisse entrer en jeu. Les auteurs sont censés omettre <code>allowExtraChars</code> lors de l'utilisation de cette option.
Mal orthographié : <code>allowTransposeTwoChars</code>	<code>mt</code>	Correspondra à un mot où deux caractères sont transposés. Le mot de motif doit comporter 4 caractères ou plus, caractères génériques exclus, pour que la transposition puisse entrer en vigueur. Les auteurs sont censés omettre <code>allowExtraChars</code> lors de l'utilisation de cette option.

TABLEAU 5 : QUELQUES EXEMPLES SUR LA SYNTAXE DE PATTERN MATCH [3]

Dans le tableau 6, quelques exemples sur la technique de Pattern Match sont présentés.

La réponse courte de l'étudiant	Option de matching	Pattern Match	Méthode retournée de match
أحمد مصطفى محمد	empty	أحمد مصطفى محمد	Vrai. C'est la correspondance exacte
مفترس	c	مفترسون	Vrai. Les caractères supplémentaires sont autorisés n'importe où dans le mot.
أحمد, مصطفى و محمد	w	أحمد	Vrai. Les caractères supplémentaires sont autorisés n'importe où dans la phrase.
أحمد مصطفى محمد	w	مصطفى أحمد محمد	Vrai. N'importe quel ordre de mots est autorisé.

TABLEAU 6 : QUELQUES EXEMPLES SUR LA TECHNIQUE PATTERN MATCH DANS MOODLE [3]

2.3- Inconvénients :

Avec la complexité, le manque d'outils et la diversité des règles d'orthographe et de ponctuation en langue arabe, cette approche (Pattern Match) rencontre des difficultés pour soigner toutes ces problématiques. Les signes de ponctuation, ne sont pas utilisés de façon régulière.

3- Le Preg Match [4]:

L'outil de Preg est un type de question qui utilise des expressions régulières pour vérifier les réponses courtes des étudiants. Les expressions régulières offrent de vastes possibilités et une grande flexibilité aux enseignants pour la formulation des questions et aux étudiants pour la rédaction des réponses. L'utilisation de type de question Preg comme réponse courte, sans aucune connaissance des expressions régulières, mais avec des hitting (caractères suggérés) avancés. Pour ce faire, le choix doit être apporté comme suit :

Notation => réponse courte de Moodle ensuite Moteur => Automates à états finis et à la fin Correspondance exacte => Oui.

Après cela, les réponses des questions à réponse courte seront simplement copiées. Fondamentalement, ce type de question est une version étendue du type de question à réponse courte. Il étend ses fonctionnalités de plusieurs manières différentes; dans le Pattern matching, en utilisant des expressions régulières, des motifs puissants peuvent être

créés décrivant les réponses possibles des étudiants comme il (Pattern matching) peut être utilisé dans les suggestions ou indications, lorsque les étudiants sont bloqués à la question.

Sur ce, le type de question Preg n'ajoute pas de caractères suggérés à la réponse de l'étudiant (contrairement au type de question à expression régulière réponse courte), et l'affiche séparément à la place pour un certain nombre de raisons : Il appartient à l'élève de décider s'il souhaite ajouter un caractère évocateur ou suggéré à sa réponse (et d'autres éventuellement) et cela facilite légèrement la réflexion sur un indice, car lorsque la réponse est modifiée, il est trop facile d'appuyer plusieurs fois sur caractère suggéré (hint), ce qui n'est généralement pas un comportement souhaitable.

Avant d'utiliser la propriété des caractères suggérés dans le type Preg, quelques options doivent être définies comme dans la plateforme Moodle. En définissant le cas de sensibilité de toutes les expressions régulières spécifiées comme réponses. Notant qu'on peut également définir le cas de sensibilité pour les parties expressions régulières. La correspondance exacte affecte la question de la manière suivante : Si Oui, toute la réponse de l'élève, de la première à la dernière lettre, doit correspondre à votre expression habituelle. Si Non, la réponse de l'étudiant peut simplement contenir une partie qui correspond à notre expression rationnelle : par exemple, si la réponse correcte est "الأسد", "الأسد ذو الأنياب الحادة", "الأسد ذو الأنياب القوية", "الأسد ذو الأنياب الحادة و القوية" seront toutes des réponses correctes des élèves. La figure 2 met en exemple l'option de cas de sensibilité dans la plateforme Moodle.

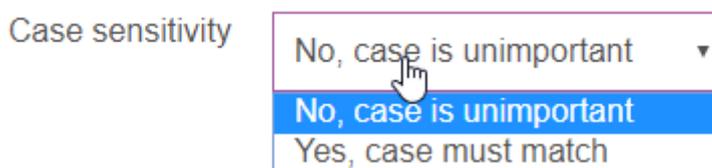


FIGURE 2 L'OPTION DE CAS DE SENSIBILITE

La propriété des caractères suggérés commence par une correspondance partielle (Partial matching). Par réponse courte partiellement correcte, c'est une chaîne commençant par des caractères corrects (correspondant à votre expression rationnelle), mais la correspondance est rompue pour certains caractères. Supposons qu'on pose la question " ما هي الألوان المشتركة في العلمين الجزائري و البلغاري؟ "

Et ont fait entrer l'expression régulière : "الأخضر, الأبيض(و) الأحمر" et un étudiant a répondu "انها الأخضر, الأبيض و الأحمر". Dans cette situation, la correspondance partielle est "الأخضر". Notant que l'expression régulière n'est pas ancrée ("Correspondance exacte" est définie sur "Non"), de sorte que la correspondance ne peut pas commencer

par le premier caractère de la réponse de l'étudiant (comme dans l'exemple ci-dessus : "انها" est ignoré). En utilisant uniquement une correspondance partielle, l'étudiant verra les parties correctes et incorrectes : **لنها الأخضر, التبييض والأحمر**

Inconvénients :

Cette approche reste ambiguë à apprendre par les enseignants surtout dans le cas où le Preg Match utilise des expressions régulières alors ça devient un double effort et une perte de temps pour l'enseignant qui doit déjà apprendre à formuler ses questions.

IV. Évaluation des systèmes actuels :

Les performances des systèmes d'évaluation automatisés semblent intéressantes dans le processus d'évaluation automatisé, et ce dans des environnements spécifiques et appropriés, caractérisés par un certain nombre d'aspects, notamment les types de questions et l'existence d'un ensemble de réponses bien établies. Dans la plupart des établissements d'enseignement, tels que les universités algériennes, l'évaluation automatique des réponses courtes est en fait pratiquement inexistante, et hostile aux complexités des systèmes d'évaluation automatisés.

Aujourd'hui les étudiants trouvent des difficultés en langues, par exemple dans les universités algériennes, est un cas fréquent dans les réponses des étudiants et même dans certaines réponses fournies par le professeur peuvent trouver de telles erreurs. Sous ces complexités, les systèmes d'évaluation automatisés existants ne peuvent pas atteindre l'efficacité escomptée pour évaluer les réponses courtes. Cela est principalement dû à la spécificité de ces outils, qui reposent sur une langue (l'anglais dans la plupart des cas), ciblent des contenus éducatifs spéciaux, reposent sur un type de question spécifique et sont disponibles dans un panier de réponses lié à ce type de question. Outre les contradictions qu'on vient de mentionner, les instruments actuels ont ignoré de nombreuses choses importantes qui existent dans le monde de l'évaluation des réponses comme :

- Ignorer les souhaits de l'enseignant pour définir ce qui est important et le plus important dans une réponse.
- Contenu variable dans la réponse courte (c.-à-d. Dans deux réponses courtes, le contenu variable est différent et les réponses sont valides).
- Plusieurs formulations possibles d'une même réponse.
- Ignorer de telles contraintes réduit l'efficacité de ces systèmes et les rend plus difficiles à adopter dans un environnement d'apprentissage réel. De plus, plusieurs problèmes sont spécifiques à l'utilisation de la langue arabe.

Les obstacles de la langue arabe dans le contexte de l'évaluation automatique

La langue arabe a beaucoup de caractéristiques, qui sont considérées comme des enjeux ou des problèmes à soulever pour l'évaluation automatique :

1- Le premier problème c'est qu'il n'y a pas de majuscules qui marquent le début d'une nouvelle phrase. La langue arabe ne supporte pas la capitalisation de noms propres tels que les noms de pays, les noms de personnes. Considérant que, dans les langues latines, ceux-ci commencent par une lettre majuscule. L'évaluation automatique arabe peut ne pas reconnaître ces entités nommées, ce qui augmente la difficulté de détecter de tels noms dans les réponses en arabe.

De plus, les signes de ponctuation, ne sont pas utilisés de façon régulière. La reconnaissance de la fin de phrase est délicate car la ponctuation n'est pas systématique et parfois les particules délimitent les phrases.

2- Le deuxième problème c'est que la plupart des mots arabes sont composés par agglutination d'éléments lexicaux de base (proclitique + base + enclitique). Par exemple, la détermination peut s'exprimer par agglutination de l'article $\text{ال}/\text{al}/$ avant le mot ($\text{المالية}/\text{al} \text{maleya}/$ (« financement »))ou par agglutination d'un pronom personnel après celui-ci ($\text{ماله}/\text{malohu}/$ (« son argent »)).

Dans toute perspective de traitement automatique, le problème est donc de décomposer le mot en ces différentes parties. Cette décomposition nécessite des connaissances de niveau supérieur en cas d'ambiguïtés (si le mot accepte plusieurs segmentations). Pour détecter la racine d'un mot, il faut connaître le schème par lequel il a été dérivé et supprimer les éléments flexionnels (antéfixes, préfixes, suffixes, post fixes) qui ont été ajoutés. La langue arabe est complexe en raison de la variation morphologique. La forme des lettres change en fonction de leur position dans le mot. De plus, le mot peut être constitué de préfixes, de lemmes et de suffixes dans des combinaisons différentes, ce qui aboutit à une morphologie très compliquée.

3- Le troisième problème c'est celui de la voyellation et qui réside dans l'absence quasi systématique de la voyellation dans les textes arabes. En effet, les signes de voyellation, lorsqu'ils sont notés, sous la forme de signes diacritiques placés au-dessus ou au-dessous des lettres, apparaissent dans certains textes (Coran, hadith) ou littéraires (poésie classique, notamment) : à vrai dire, ils sont édités en graphie voyellée. L'absence de voyelles (la non-voyellation) dans les textes arabes génère plusieurs cas d'ambiguïtés et des problèmes lors de l'analyse automatique. En effet, l'ambiguïté grammaticale augmente si le mot est non voyellé. Cela est dû au fait qu'un mot non voyellé possède plusieurs voyellations possibles, et pour chaque voyellation est associée une liste différente de catégories grammaticales.

4- Le quatrième problème

L'arabe est une langue mal adaptée pour les approches basées sur les corpus par rapport à l'anglais, car il y a un manque de données, ce qui affecte négativement la recherche sur les approches sémantiques basées sur les corpus en arabe. [5] ont passé en revue quatorze corpus arabes et les ont catégorisés par leur langue cible, objet, date du texte, lieu, domaine de texte, représentativité, mode de texte, taille. Plusieurs de ces corpus ne fournissent aucune information concernant la période couverte par les textes. De plus, pour tous les corpus, les textes constitutifs ne sont pas classés en fonction de leurs dates ou de la période à laquelle ils appartiennent ; il y a donc une limite à l'utilité du corpus et une difficulté à comparer les langues utilisées à différentes périodes, et à observer comment la langue arabe a évolué.

Pour les approches basées sur la connaissance, WordNet⁶ est utilisé dans divers domaines tels que la recherche d'information et la similarité sémantique, En raison du succès de WordNet dans les applications en anglais, plusieurs projets sont actuellement menés pour développer WordNet pour d'autres langues. WordNet arabe (AWN) a été développé en utilisant la même méthodologie qu'EuroWordNet. Il se compose de 11 270 synsets et contient 23 496 expressions arabes (mots et multi-mots). Les principales limitations de l'AWN actuel sont un manque d'informations et de concepts par rapport à WordNet en anglais, et quelques relations sémantiques entre les synsets⁷. De nombreux concepts arabes n'ont pas été inclus dans la base de données Awn⁸. Cette limitation constitue un obstacle majeur à l'utilisation d'AWN en tant que source d'approches basées sur la connaissance.

V. Les travaux connexes à notre recherche :

Le travail qu'on mène est de concevoir, développer et intégrer notre Plugin dans le système de Quiz de Moodle dans le cadre d'une approche combinée (sémantique et syntaxique) qui permet de comparer entre un couple de réponses courtes en langue arabe. Dans ce contexte, notre travail est connexe par rapport à la similarité et par rapport au développement du Plugin d'évaluation automatique des réponses courtes dans les plateformes.

- **Travaux connexes par rapport à la similarité :**

Menés par **Ouharani & al** [6], **A. Benayad et Y. Atoub** [7], **K.Garoudja et A. Abdallah** [8]. Ces auteurs ont utilisé une approche statistique basée sur le corpus pour développer un système d'évaluation à réponse courte. Ils ont testé des mesures sémantiques et syntaxiques

⁶ WordNet WordNet est une base de données lexicale. Son but est de répertorier, classifier et mettre en relation de diverses manières le contenu sémantique et lexical de la langue anglaise.

<http://www.nltk.org/howto/wordnet.html>

⁷ Un synset est un ensemble d'un ou plusieurs synonymes

⁸ Awn : Arabic WordNet <http://globalwordnet.org/resources/arabic-wordnet/>

sur le dataset (**Cairo University** Dataset et SemEval Dataset). Leurs résultats ont montré que les meilleures valeurs de corrélation obtenues (81,49) en utilisant une hybridation de l'approche sémantique [7] SV (Somme Vectoriel) avec le modèle syntaxique [8] l'indice de Dice.

Dans ce contexte, notre travail est aussi connexe aux travaux menés par Gomaa & al [9]. Les auteurs ont utilisé des mesures de similarité syntaxiques et des mesures basées sur le corpus pour développer leur système de notation à réponse courte. Ils ont testé les mesures sur le dataset (GOMAA Dataset) qu'ils ont construit eux-mêmes.

- **Travaux connexes par rapport au Plugin :**

A ce stade, notre travail est connexe au CodeRunner [10] qui est un plug-in de type question open-source gratuit pour Moodle qui peut exécuter le code de programme soumis par les étudiants en réponse à un large éventail de questions de programmation dans de nombreuses langues. Il est principalement destiné aux cours de programmation informatique bien qu'il puisse être utilisé pour noter toute question pour laquelle la réponse est textuelle. Il est normalement utilisé dans le mode de quiz adaptatif de Moodle; Les étudiants collent leur code en réponse à chaque question de programmation et voient immédiatement les résultats de leurs tests. Ils peuvent ensuite corriger leur code et le soumettre à nouveau, généralement pour une petite pénalité. CodeRunner ne fait pas de traitement NLP (Natural Language Processing) sur les réponses. Il a une architecture flexible prenant en charge différentes manières d'exécuter une tâche d'étudiant dans un environnement protégé qui est pris en charge par la version actuelle. CodeRunner utilise un serveur distinct, développé spécifiquement pour être utilisé par CodeRunner, appelé Jobe (REST API)⁹.

Notre travail est connexe au travail effectué dans [11] par un binôme de l'école normale supérieure (ENS) de Kouba. Ce travail consiste à développer un Plugin d'évaluation automatique des réponses courtes pour la langue arabe et l'intégrer dans la plateforme Moodle pour améliorer le système de notation des réponses courtes existant dans la plateforme. Les étapes de ce travail consistent à intégrer leur Plugin en dupliquant le Plugin ShortAnswer existant dans Moodle pour garder les mêmes fonctionnalités de ce dernier, ensuite ajouter des portions du code PHP pour apporter une amélioration de l'évaluation automatique des réponses courtes. Ces améliorations restent limitées puisqu'elles ne portent pas sur le traitement de la langue arabe.

⁹ API (Interface de programmation d'application) est un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels.

Conclusion :

Développer un système d'évaluation efficace et impartial pour réponses courtes en langue arabe est un problème difficile dans la mesure et l'évaluation de l'éducation, compte tenu des difficultés qu'entretiennent les enseignants avec les outils informatiques, de la diversité des réponses, le manque des outils pour la langue arabe, et de la subjectivité des classeurs. Et il y a une disponibilité limitée de tels plugins ou outils dans le classement automatique des types de questions pour des réponses courtes et descriptives.

Une vue générale et encore détaillée sur les approches et les outils d'évaluation automatique des réponses courtes dans les plateformes de télé-enseignement a été présenté dans ce chapitre, y compris les différents concepts qui ont besoin d'être compris pour mieux situer la suite de travail. Ces approches sont très limitées pour garantir une bonne évaluation automatique des réponses courtes sans acquisition des connaissances des outils informatiques. Dans ce travail, nous allons présenter un Plugin avec des approches sémantiques et syntaxiques déjà développées l'année passée [6] afin de l'intégrer dans la plateforme de télé-enseignement Moodle. L'avantage d'utiliser cette dernière repose d'un côté sur sa grande popularité et d'un autre côté sur sa documentation très riche ainsi que sa communauté open source très active qui maintient cette documentation à jour grâce aux forums multiples.

Chapitre II : Présentation du système de quiz de la plateforme Moodle [12]

Introduction :

L'apprentissage en ligne a révolutionné la scène de l'enseignement académique et de la formation en entreprise et qui est sans contestation une des tendances du milieu de l'éducation ayant une des plus fortes croissances en raison de ses services distinctifs qu'il offre. Malgré cela, il reste limité qu'à certains pays développés.

Au cours de ces développements, un certain nombre de sites Web spécialisés dans le domaine de l'apprentissage en ligne ont été créés, nous allons aborder l'une des plus importantes plateformes au monde qui est « Modular Object-Oriented Dynamic Learning Environment » plus connue sous son acronyme Moodle [12].

Cette plate-forme est destinée aux enseignants, aux professeurs associés et aux étudiants, fournissant un environnement technologique et éducative qui facilite la communication entre les différents acteurs du système éducatif.

1- Présentation générale de la plateforme

- Moodle est une plate-forme d'apprentissage en ligne (e-learning) servant à créer des communautés d'apprenants autour de contenus et d'activités pédagogiques.
- Moodle est un logiciel de « formation à distance » ou, de manière plus académique, un « environnement numérique d'apprentissage ». Il permet de diffuser des documents (textes, audio, vidéo, etc.) et de soutenir les interactions entre professeurs et étudiants, à l'aide d'une variété d'outils de communication, d'évaluation et de gestion.
- Utilisé par des dizaines de milliers d'environnement d'apprentissage du monde entier, Moodle a la confiance d'institutions et organisations grandes et petites, parmi lesquelles sont notées Shell, la London School of Economics, l'Université d'État de New York, Microsoft et l'Open University. Le nombre d'utilisateurs de

Moodle autour du monde, plus de 65 millions, tant dans le milieu académique que celui de l'entreprise, en fait l'une des plateformes d'apprentissage les plus utilisées.

- Les capacités multilingues de Moodle garantissent qu'aucune barrière linguistique n'entrave l'apprentissage en ligne. La communauté Moodle a traduit Moodle dans plus de 100 langues (Incluant l'arabe), de sorte que ses utilisateurs peuvent utiliser Moodle dans leur langue, et bénéficier de nombreuses ressources, d'assistance et de discussions de la communauté dans les diverses langues.

2- Fonctionnalités de Moodle :

Création d'un environnement d'apprentissage moderne efficace

La classe traditionnelle fusionne avec le monde en ligne et les étudiants doivent savoir comment fonctionner dans cet environnement d'apprentissage mixte. Moodle est un point d'entrée sécurisé pour les étudiants qui souhaitent commencer à naviguer en ligne et leur permet de se familiariser avec Internet. Au fur et à mesure que les étudiants suivent des cours d'apprentissage en ligne, Moodle enseigne des compétences de base en informatique, telles que la dactylographie, l'envoi de courriels, l'utilisation de PDF et la gestion de vidéos et de fichiers en ligne.

Diversifier les méthodes d'enseignement

Dans Moodle, des éléments multimédias peuvent être intégrés à des cours tels que vidéo, audio et diaporamas. Semblables à la ludification, ces différents éléments permettent de stimuler et d'engager les cerveaux en proposant diverses méthodes d'enseignement.

Simplifier le processus de notation

La notation des tâches et des tests peut être complétée directement dans Moodle, via le plugin Gradebook.

Communications ouvertes et accessibles

Un flux constant de communication est essentiel au succès dans les hautes éducation et Moodle ouvre ce canal.

Les différentes voies de communication disponibles dans Moodle comprennent les forums, la messagerie, le chat, les commentaires. et des articles de blog pour que les étudiants et les enseignants puissent communiquer permettant l'apprentissage à aller au-delà de la journée scolaire

Collaboration plus facile

Moodle permet et encourage la collaboration et le travail d'équipe. Les fonctions de communication dans Moodle énumérées ci-dessus (discussions, forums, messagerie)

(discussions, forums, messagerie) permettent aux étudiants de travailler ensemble, de partager des idées et de poser / répondre à des questions, si nécessaire.

3- La Modularité sur Moodle :

Le M dans Moodle est synonyme de modulaire. Le moyen le plus simple et le plus facile d'utiliser les fonctionnalités de Moodle consiste à l'utilisation de ses plugins. Les plugins permettent d'ajouter des fonctionnalités à Moodle, telles que de nouvelles activités, de nouveaux types de questions de questionnaire, de nouveaux rapports, des intégrations avec d'autres systèmes et bien d'autres.

- **Les Test (Quiz)**

Le module d'activité Test permet à l'enseignant de concevoir et de créer des questionnaires constitués d'une grande variété de types de questions , y compris les types de questions à choix multiple, les vrai-faux, les réponses courtes et le glisser-déposer d'images et de texte. Ces questions sont conservées dans la Banque de questions et peuvent être réutilisées dans différents questionnaires. la Figure 3 montre un exemple d'un test Moodle.

Question 1
Pas encore répondu
Noté sur 5,00
Marquer la question
Modifier la question

ما النتائج المترتبة عن تطبيق سياسة أمن المعلومات في النظام المعلوماتي ؟

Réponse:

Question 2
Pas encore répondu
Noté sur 3,00
Marquer la question
Modifier la question

...Choisir هي جريمة اختراق وسرقة معلومات يقوم بها قرصنة القيعات ...Choisir دون اذن لاغراض تخريبية أما اختياز

الاختراق فيتم فيه توظيف العديد من ذوي القيعات ...Choisir لاختياز أنظمة أمن أجهزة الكمبيوتر في المؤسسات مما يسمح بتحديد عيوبها

وتغراتها قصد تحسين دفاعاتها

FIGURE 3 EXEMPLE DE TEST MOODLE

• Caractéristiques du système Quiz de Moodle

- Les tests peuvent être paramétrés pour permettre plusieurs tentatives. Chaque tentative est automatiquement notée et l'enseignant peut choisir de donner ou non une rétroaction et de montrer ou non les réponses
- La rétroaction et l'autoévaluation sont des constituants importants de l'environnement d'apprentissage. Il est possible de donner des rétroactions à l'étudiant à chaque question, ou de donner une rétroaction globale sur le test. L'activité test permet de choisir à quel moment seront affichées les rétroactions et les notes
- L'enseignant bénéficie au bloc Administration du cours d'une grande variété de Rapports, en plus des Notes. Non seulement les rapports permettent de s'attarder à la tentative pour chaque question d'un étudiant donné, mais ils permettent également d'analyser la validité d'une question selon l'ensemble des réponses des étudiants.

- Les paramètres du test permettent différents modes d'affichage. L'enseignant peut ainsi afficher les questions de façon aléatoire pour chaque étudiant, de même que les choix de réponses.
- Il est possible de sélectionner manuellement ou aléatoirement des questions à partir des catégories de la banque de questions pour créer un test.
- L'enseignant peut choisir le comportement des questions dans le test. Le test pourrait être un test classique qui transmet la rétroaction à l'étudiant lorsqu'il a terminé le test; ou un test qui affiche les notes et les rétroactions pendant le test et, possiblement, donne une autre chance de répondre à la question (moyennant une pénalité) après avoir obtenu la rétroaction.

● Système de Notation En ligne (Gradebook)

Le carnet de notes de Moodle recueille les notes de toutes les activités notables d'un cours, ce qui permet à l'enseignant de les gérer au même endroit et donne également aux étudiants un aperçu de leurs notes. Les enseignants peuvent facilement éditer, recalculer et visualiser les notes comme le montre la Figure 4.

Les activités Moodle, telles que les Quiz, peuvent être automatiquement notés - un gain de temps considérable -, tandis que les essais et autres travaux soumis nécessitant une notation manuelle bénéficient d'échelles, de rubriques et de résultats personnalisables, le tout sous le contrôle d'un enseignant de cours régulier.

Nom ▲ Prénom		Adresse de courriel	Cyber ☰		
			Test Cyber 1 ↕	Test Cyber 2 ↕	Σ Total du cours ↕
	Bouchra Amirouche	 bouchera.amr@gmail.com	23,80 	7,90 	31,70
	Chaimaa Bouchoucha	 chaimabouchoucha@gmail.com	47,70 	12,70 	60,40
	Zhor Difallah	 zhor.difallah@gmail.com	49,45 	13,75 	63,20
	imene Elghers	 Elghersimenea6@gmail.com	- 	- 	-
	Oussama Hamel	 oussamahamel09@gmail.com	37,85 	10,55 	48,40
	Karim Hemina	 hemina.karim@etu.univ-blida.dz	- 	- 	-
	Yousra Ismail	 yousisi19@gmail.com	49,00 	11,60 	60,60
	Selena Lamari	 lamari.selena.si.a3@gmail.com	44,55 	12,80 	57,35

FIGURE 4 RAPPORT DES ACTIVITES DES PARTICIPANTS

Toutes les notes de chaque élève d'un cours peuvent être trouvées dans le carnet de notes du cours voire « Figure 5 ». Cela montre toutes les tentatives de quiz des étudiants, avec la note globale et la note pour chaque question. Il y a des liens pour examiner tous les détails de la tentative d'un élève, exactement comme l'étudiant le verrait.

Prénom / Nom	Adresse de courriel	État	Commencé le	Terminé	Temps utilisé	Note/85,00	Q. 1 /5,00	Q. 2 /5,00	Q. 3 /5,00
Oussama Hamel Relecture de cette tentative	oussamahamel09@gmail.com	Terminé	16 June 2019 22:24	16 June 2019 23:42	1 heure 17 min	37,85	✓ 3,15	✓ 1,90	✓ 2,05
Selena Lamari Relecture de cette tentative	lamari.selena.si.a3@gmail.com	Terminé	17 June 2019 00:11	17 June 2019 02:06	1 heure 55 min	44,55	✓ 2,70	✓ 2,70	✓ 2,80
Yousra Ismail Relecture de cette tentative	yousisi19@gmail.com	Terminé	17 June 2019 19:32	17 June 2019 21:26	1 heure 53 min	49,00	✓ 3,55	✓ 2,85	✓ 2,60

FIGURE 5 RAPPORT DES NOTES ETUDIANT D'UN TEST MOODLE

- **Suivi du Progrès**

Les cours sont des emplacements sur Moodle où les enseignants peuvent utiliser du matériel d'apprentissage pour leurs étudiants.

Un cours est composé de plusieurs activités comme des leçons, des sondages ou des Test. Et l'un des points fort de Moodle c'est la capacité de suivre le progrès des étudiants par rapport à ces activités.

La figure 6 montre le progrès d'un cours vu par l'étudiant.

 Announcements from your tutor	<input checked="" type="checkbox"/>
 Prior Knowledge assessment	<input checked="" type="checkbox"/>
 Factual recall test	<input checked="" type="checkbox"/>
 Course chat	<input type="checkbox"/>

FIGURE 6 PROGRES VU PAR L'ETUDIANT

- **Types de Questions**

Il est possible d'ajouter une variété de types de questions dans les activités test et leçon où chaque type de question est suivi d'une description de ce type. Voir « Figure 7 »

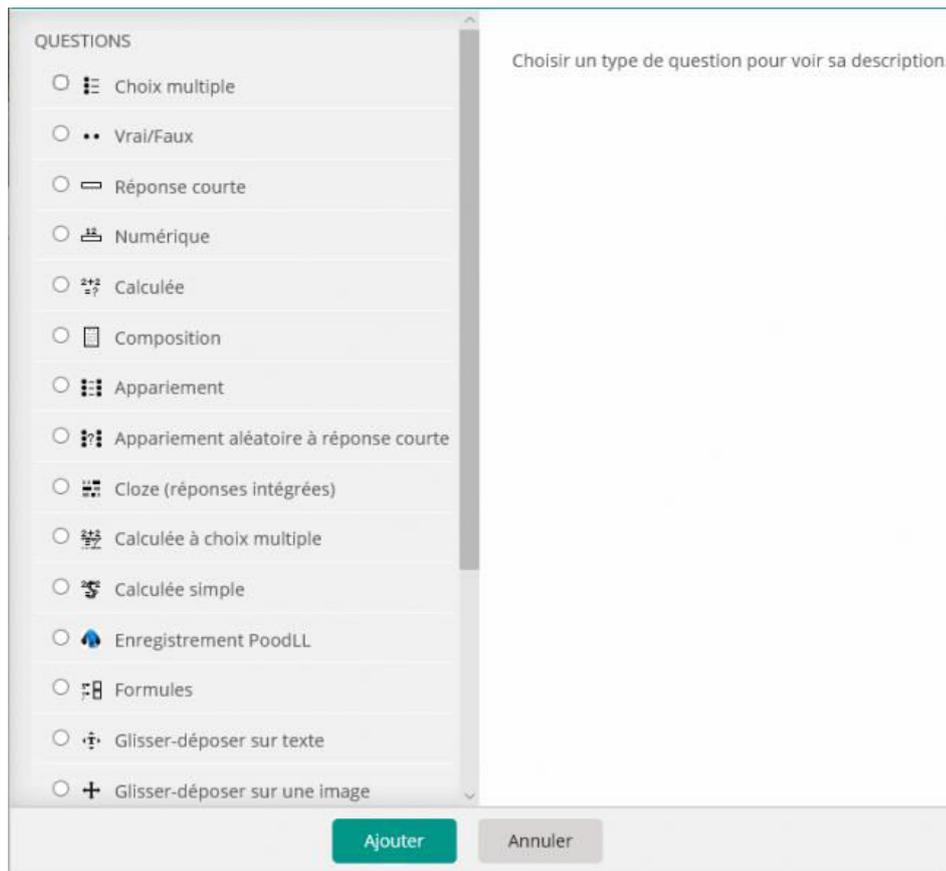


FIGURE 7 TYPES DE QUESTIONS MOODLE

- **Question à réponse courte (Shortanswer) de Moodle**

L'un des types de questions les plus populaires sur Moodle c'est les questions à réponse courte aussi appelées Shortanswer questions.

Dans une question à réponse courte, l'étudiant saisit un mot ou une proposition en réponse à une question. Les réponses peuvent être sensibles ou non à la casse. La réponse peut être un mot ou une expression, mais elle doit correspondre exactement à l'une des réponses acceptables. La bonne réponse devrait être aussi courte que possible afin d'éviter les erreurs. Les activités Test et Leçon proposent le type de question à réponse courte comme le montre la Figure 8.

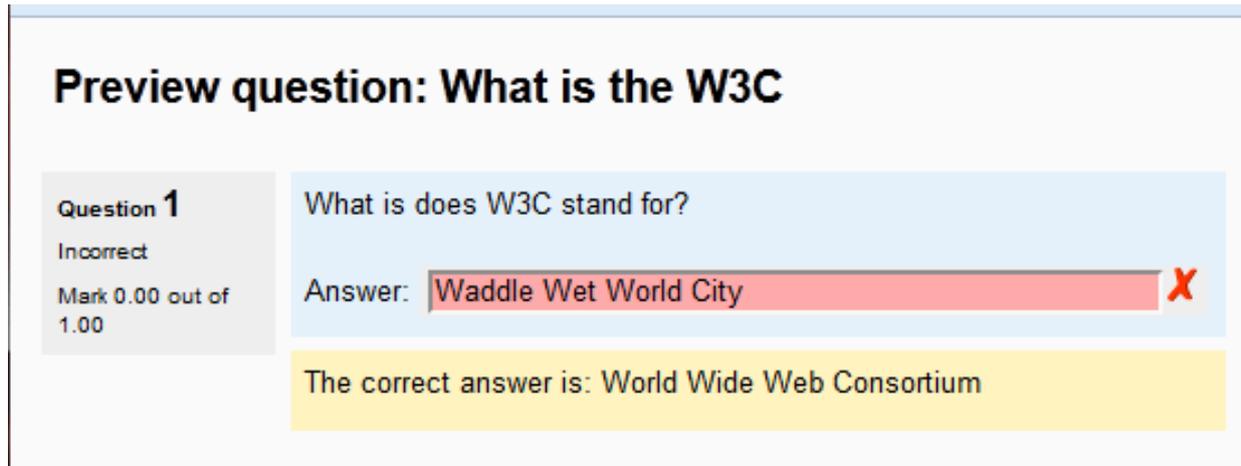


FIGURE 8 QUESTION DE TYPE SHORTANSWER DE MOODLE

Les réponses modèles peuvent être rédigées en utilisant l'astérisque (*) appelé Joker, pour indiquer que n'importe quel caractère est accepté à cet endroit. Par exemple, il faut écrire `pla*ement` pour accepter tout mot ou toute proposition commençant par `pla` et se terminant par `ement`. Si une question est posée exigeant deux réponses (carburant et oxygène), il faut limiter le nombre de variantes en écrivant `carburant*oxygène`. Cela acceptera les réponses « carburant oxygène », « carburant, oxygène », « carburant; oxygène », « carburant et oxygène », « carburant & oxygène », « carburant/oxygène », « carburant and oxygène », « carburant&oxygène ». Toutefois, cela acceptera aussi les réponses « carburant ou oxygène », « carburant sans oxygène » et « carburant|oxygène ».

Voici quelques réponses et notes pour la question « Que consomme une fusée ? » :

1. `oxygène*carburant` - note : 100 %
2. `*carburant*` - note : 50 %
3. `*oxygène*` - note : 50 %
4. `*air*` - note : 40 %
5. `*` - note : 0 %

L'ordre des réponses modèles est important. Les réponses sont évaluées de la première à la dernière. Lorsqu'une réponse correspondante est trouvée, le processus s'arrête. Si aucune correspondance n'est trouvée, la note est de 0 et la rétroaction générale est utilisée. Il est bon de mettre simplement un joker comme toute dernière réponse pour que le processus d'évaluation sache quoi faire si aucune correspondance n'est trouvée.

4- La forme générale de système de l'évaluation automatique des réponses courtes :

La figure (9) au-dessous montre la forme générale du système à travers lequel l'évaluation automatique de la réponse courte est effectuée :

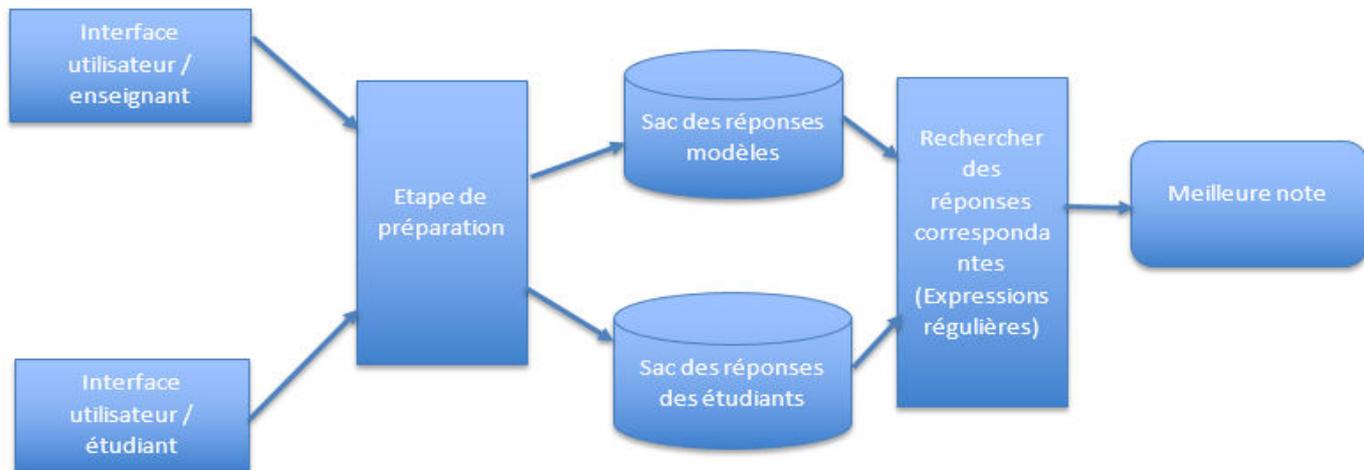


FIGURE 9 LA FORME GENERALE DE L'EVALUATION AUTOMATIQUE DES REPNSES COURTES [11]

5- Vue d'ensemble sur le moteur de question Moodle (Question engine) [13] :

Dans le moteur de questions pour l'évaluation en ligne. Clairement, ayant des questions, qui viennent de la banque de questions. Elles vont être pris comme donnés. La cible de cette partie c'est lorsqu'un étudiant tente de répondre à certaines questions (par exemple, lorsqu'il tente de répondre à un quiz). Cela est géré en pensant à la tentative de chaque élève de traiter chaque question séparément, puis d'agréger ces tentatives de question distinctes dans l'ensemble de celles constituant la tentative de l'élève pour une activité particulière.

Pour chaque tentative de question, ayant besoin de savoir dans quel état elle se trouve actuellement (en cours, complétée correctement, ...) si elle a été notée et, dans l'affirmative, quelle note a été attribuée. La Figure 10 montre tous les états d'une tentative de question.

Étant donné l'état actuel d'une question, elle doit être affichée en HTML. La question devra peut-être être affichée de différentes manières. Par exemple, peut-être que les étudiants ne sont pas encore autorisés à voir les informations sur les notes attribuées, mais si un enseignant les examine, ils devraient voir les notes. Les enseignants peuvent également avoir besoin de voir d'autres bits d'interface, tels que des liens d'édition. Il peut y avoir une ou plusieurs questions affichées incorporées dans une page HTML particulière, par exemple une page d'une tentative de quiz.

Enfin, la question doit passer d'un état à un autre. Il y a deux sortes de chemin qui peuvent arriver. Tout d'abord, l'élève peut avoir interagi avec certaines questions affichées dans le cadre d'une page HTML, puis soumis cette page. En d'autres termes, le moteur de question doit traiter directement les données soumises dans le cadre d'une requête HTTP POST. Deuxièmement, l'élève peut avoir effectué une autre action (par exemple, cliquer sur les boutons Démarrer la tentative ou Soumettre tout et terminer dans le quiz). Par conséquent, une autre partie du code Moodle peut vouloir faire changer l'état de certaines questions. Non seulement examiner l'état actuel de chaque question, un enregistrement de tout l'historique de ce qui s'est passé sera indispensable. Par exemple, un enseignant peut souhaiter revoir tout ce que l'élève a fait. Ou, lorsqu'une question interactive est classée, une lecture d'intégralité de la séquence est obligatoire de ce que l'élève a fait.

Pour être flexible, le moteur de question doit prendre en charge tout plug-in de type question dont notre PLUGIN Arabic ShortAnswer. Il doit également être flexible dans la manière dont les questions se comportent. Par exemple, certaines questions ne peuvent être notées manuellement par un enseignant. Les questions classées automatiquement peuvent être configurées pour se comporter de différentes manières, par exemple de manière interactive avec des retours immédiats et de multiples tentatives au cours d'une tentative de question ou avec des commentaires différés, lorsque les questions ne sont notées que lorsque l'étudiant les soumet et les termine.

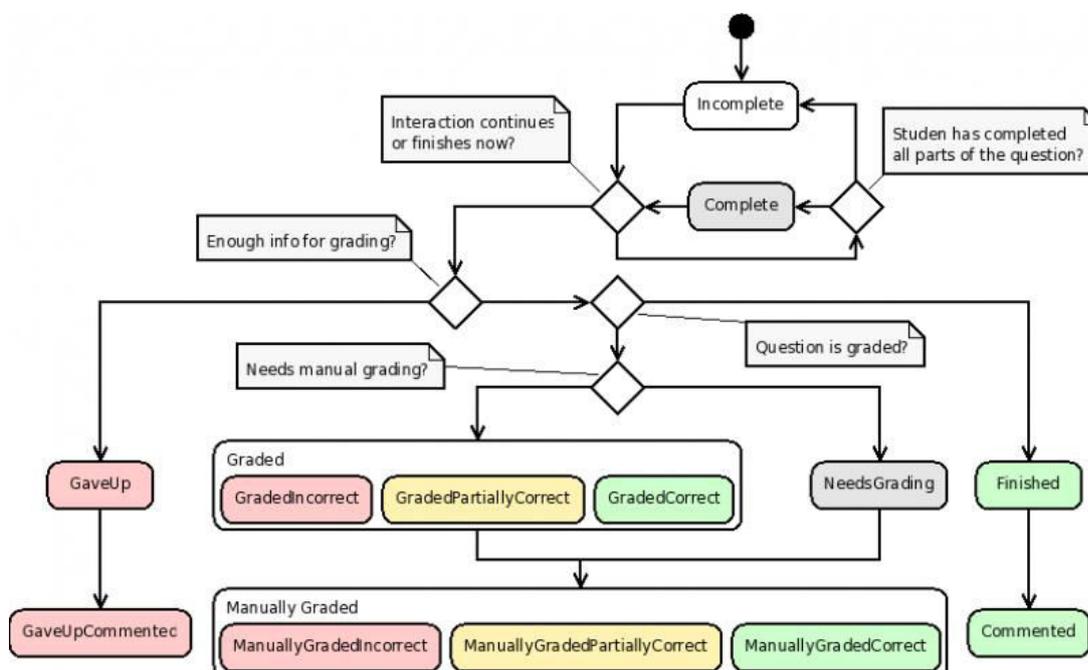


FIGURE 10 DIAGRAMME D'ÉTAT DE TRANSITIONS D'UN COMPORTEMENT PARTICULIER DE LA TENTATIVE DE QUESTION [14]

5.1- L'état d'une question :

Comme indiqué plus haut, l'un des problèmes clés est l'état actuel de la question tentée et, en fait, l'historique complet des États traversés. Ceci est la responsabilité de la classe « `question_attempt` ». On n'est pas seulement intéressé par l'état actuel de la question, mais par l'historique complet des états qu'une question a traversés. Par conséquent, une `question_attempt` consiste principalement en une liste de « `question_attempt_steps` ».

5.2- Les tentatives de question (Question attempts) :

Lorsqu'un étudiant commence à répondre à une question, cela indique au système Moodle qu'il y a une tentative de question. Les pas de cette dernière « `question_attempt_steps` » comportent les champs suivants :

- **Etat (state)**

L'état de la question est à cette étape. Ce sera un `question_state` comme `question_state::$correct`.

- **Fraction**

Le score obtenu par l'étudiant pour cette question jusqu'à présent. Cela commencera par nul et changera plus tard. Voir la note sur les scores ci-dessous.

- **Temps créé (timecreated)**

Horodatage du moment où l'action qui a créé cet état s'est produite.

- **Identifiant d'utilisateur**

L'utilisateur dont l'action a créé cet état. Lors de la tentative, il s'agira de l'identifiant de l'utilisateur qui a tenté de répondre à la question, mais supposons qu'un enseignant note manuellement la question finie plus tard, ce sera alors l'identifiant de l'enseignant.

- **Note maximale (maxmark)**

Combien de marques vaut cette question. Tous les scores stockés sous forme de fractions doivent être multipliés par ceci avant d'être affichés.

- **Fraction minimale (minfraction)**

Certaines questions peuvent renvoyer des notes négatives. Par exemple, une question à choix multiples avec 5 choix peut renvoyer +1 pour le bien et -0,25 pour le faux, de sorte que le score moyen pour deviner est de 0. Ou, si un marquage basé sur la certitude est utilisé et que l'élève est confiant mais faux, ils peuvent recevoir une note négative. La déclaration correcte est que les scores des fractions sont sur une échelle de **minfraction** à 1.

- **Signalé (flagged)**

Les questions peuvent être signalées. Il s'agit d'un booléen marche / arrêt qui peut être basculé par l'étudiant. C'est comme une simple forme de bookmarking. C'est stocké ici.

(L'historique des modifications apportées à l'état signalé n'est pas considéré comme important. C'est pourquoi l'activation du drapeau ne crée pas de nouvelle étape.)

Et finalement, certaines informations sont enregistrées pour faciliter l'exploitation efficace des rapports.

- **Résumé de question (questionsummary)**

Ceci est un résumé en texte clair de la question posée à l'étudiant. (Ce champ résume la question spécifique posée à l'étudiant.)

- **Résumé d'une réponse (responsesummary)**

Ceci est un résumé en texte clair de la réponse donnée par l'étudiant.

- **Bonne réponse (rightanswer)**

Ceci est un résumé en texte clair de la bonne réponse.

5.3- La note sur les scores :

Le moteur de questions stocke les scores de deux manières différentes. Normalement, chaque question d'une tentative de test, par exemple, vaudra un certain nombre de points. Par exemple, la question 1 peut valoir 5 points, la question 2, 3 points, etc.

Cependant, ces marques doivent parfois changer. Par exemple, peut-être que la question 2 est maintenant considérée comme peu judicieuse et souhaitant donc modifier la notation du quiz pour qu'elle vaille maintenant zéro. Mais ensuite, on pourra peut-être reconsidérer et vouloir redonner à 3 points.

Dans le code, les choses appelées **fraction** sont toujours sur l'échelle 0 à 1; les choses appelées **mark** sont toujours à l'échelle de **fraction * maxmark**. Les marques sont ce qui est affiché aux utilisateurs dans l'interface utilisateur.

Conclusion :

Les outils de développement d'évaluation automatique des réponses courtes liés au système Moodle sont limités de manière qu'ils se basent seulement sur le modèle de correspondance de tel sorte que l'étudiant doit fournir des réponses qui ressemblent aux réponses modèles. L'étudiant doit avoir impérativement un minimum de requis sur les points faibles du plugin réponses courtes (Par exemple, un espace supplémentaire entraîne une réponse fausse). L'étudiant est aussi restreint par des instructions à suivre durant la rédaction de la réponse courte et enfin, les enseignants ont besoin d'une formation d'apprentissage sur l'utilisation des expressions régulières et Preg match.

En raison de ces enjeux, le développement de nouveau plugin pour réponses courtes est essentiel de tel sorte que l'amélioration doit porter sur l'utilisation et la navigation facile du Plugin auprès des enseignants sans avoir besoin des requis, la liberté des étudiants durant la rédaction des réponses courtes, des outils plus performants et la flexibilité avec des différentes approches d'évaluation automatiques des réponses courtes écrites en langage Python.

Chapitre III : Approche d'évaluation automatique déployée

Introduction :

Il existe plusieurs approches qui traitent le sujet d'évaluation automatique des réponses courtes, dont les approches statistiques. Ces dernières sont basées sur des corpus et elles ne nécessitent pas la compréhension du vocabulaire ou de la grammaire de la langue d'un texte. Parmi ces approches, une [7] a été générée par un espace sémantique et développée l'année passée dans un cadre de la recherche technologique et écrite en python. Comme illustré dans la fig.11, l'approche est composée de trois étapes et nous allons la retenir pour l'implémenter dans le Plugin.

Dans la première étape, il s'agit de créer l'espace sémantique après acquisition et traitement de corpus. Dans les étapes deux et trois, le calcul de similarité est effectué et une note est générée. Dans la suite nous développons ces trois étapes de manière plus détaillé :

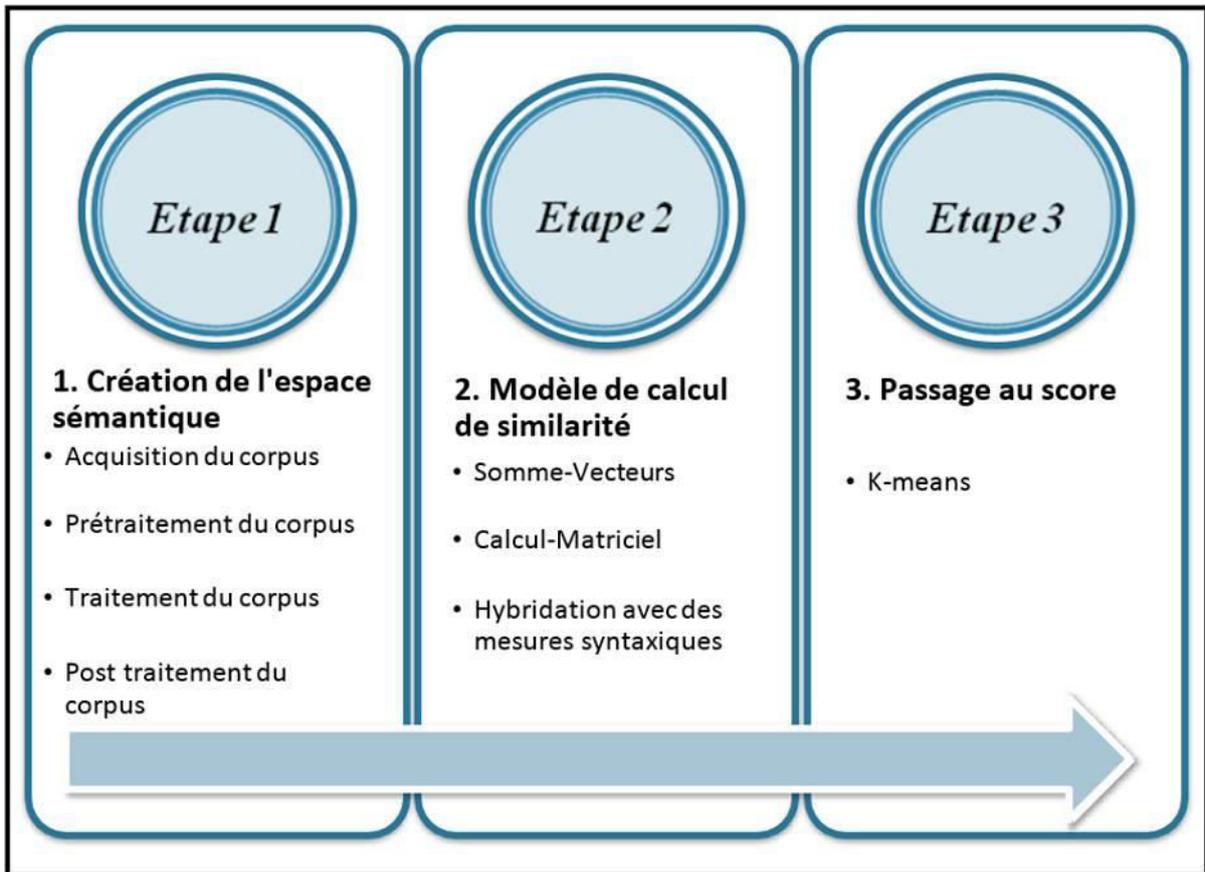


FIGURE 11 SCHEMA DES ETAPES PRINCIPALES [7]

1. ETAPE 1 : Construction de l'espace sémantique :

Cette partie présente en détail la première étape de la « Figure 11 ». Dans la « Figure 12 » les différentes phases de création de l'espace sémantique sont présentées. Tout d'abord, l'acquisition et l'analyse du corpus afin de choisir celui qui convient. Ensuite, la phase du prétraitement qui consiste à préparer le corpus pour tout usage. La phase du traitement du corpus comprend l'application de l'une des approches de similarité dans le but de construction de l'espace sémantique. En fin, le post traitement du corpus qui prend en considération l'importance des mots dans le corpus.

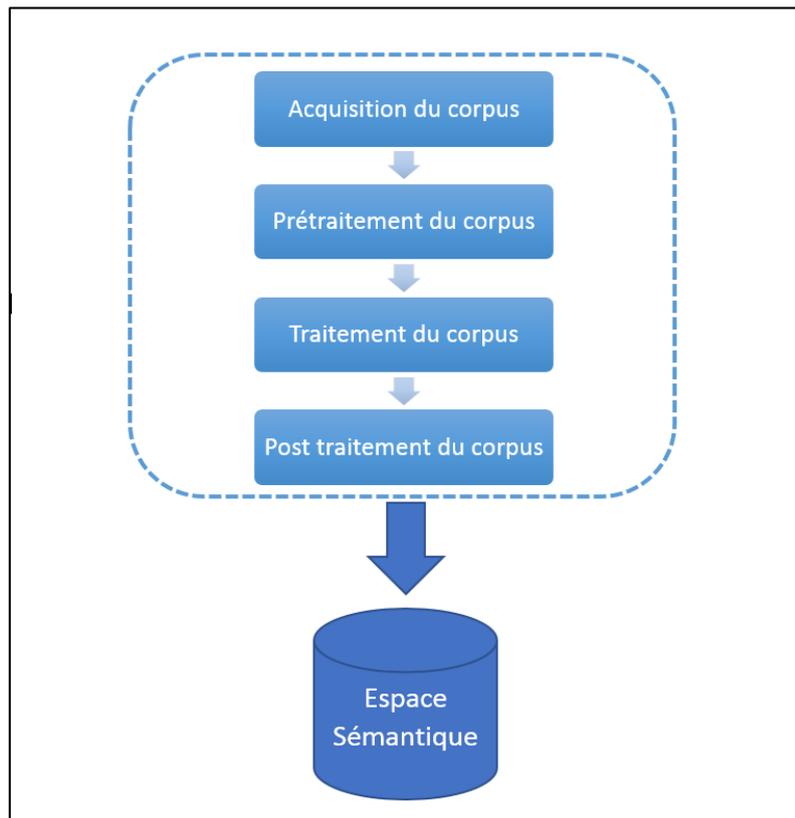


FIGURE 12 PHASES DE CREATION DE L'ESPACE SEMANTIQUE

- **PHASE 1 : Acquisition du corpus**

Sur le net, il existe une variété de corpus comme corpus Khaleej [15], Osac [16]... dont chacun d'eux est dédié pour une certaine finalité. Cette phase comprend le traitement et l'analyse des corpus.

- **PHASE 2 : Prétraitement du corpus**

Pour cette tâche, le travail doit être fait sur les corpus utilisés. Des techniques Traitement Automatique Langage connus par TAL ont été adoptées. Cette tâche se compose de deux autres sous-tâches :

PHASE 2.1 : Nettoyage de corpus :

Ceci consiste à :

- Convertir le corpus utilisé du format « html » au format « txt ».
- Réduire les successions des espaces en un espace simple.

- Supprimer les lettres non arabe (A, B...Z, a, b,...z).
- Supprimer les diacritiques.
- Supprimer les numéros.
- Normaliser les mots en remplaçant :
 - « ا, إ, ؤ, آ par ا »
 - « ي, يء, ء, عى par ئ »
 - « ة par ه »
 - « ي par ي ».
- Tokeniser chaque document et alors chaque corpus en considérant un espace simple comme séparateur.

PHASE 2.2 : L'approche par Stemming adoptée :

Un mécanisme basé sur le stemming afin d'en analyser l'impact sur l'évaluation automatique des questions à réponses courtes en langue arabe. En effet, il est très difficile de mettre en œuvre les mécanismes d'évaluation automatique pour la langue arabe, en raison de sa nature complexe, étant très flexionnelle et ambiguë en l'absence de signes diacritiques. Il n'y a eu que peu de tentatives de recherche sur ce sujet, et jusqu'à présent, aucun d'entre eux n'a été en mesure de fournir un système d'évaluation automatique entièrement fonctionnel. Les techniques de stemming ont été exploitées en combinaison avec les mesures de similarités développées. Un algorithme de stemming peut être défini comme la procédure de réduction de tous les mots qui partagent la même racine à une forme commune [17].

Pour toutes les approches de similarités développées, deux cas ont été considérés :

- Une technique de stemming lourde (Heavy Stemming) est appliquée aux réponses à comparer. Le stemming lourd, également appelé « RootStemming » (Stemming à la racine), consiste à supprimer les préfixes et les suffixes bien connus pour extraire la racine réelle d'un mot et à identifier le motif en correspondance avec le mot restant.
- Une technique de stemming légère (Light Stemming) est appliquée aux réponses à comparer. Le stemming léger est un processus moins complexe, où le stemming est arrêté sur la suppression des préfixes et des suffixes, sans tenter d'identifier la racine réelle du mot.

Le stemming consiste en général à réaliser les actions suivantes pour chaque couple de réponses à comparer :

- Suppression des nombres des deux réponses.
- Suppression des signes diacritiques des deux réponses.
- Suppression de la ponctuation et lettres spéciaux.
- Suppression des mots vides (stopwords). Une liste de mots vides est disponible dans la base de données (هي هما , اذا , هو , ان , في , و , ...).
- Enlever le (ال : AL), et ses Dérivés, (كال , وال , تال , و بال , لبال , فال , ال , لل , بال فبال , ...)
- Suppression du préfixe si la longueur du mot est supérieure à 3.
 - Suppression du suffixe, si la longueur du mot est supérieure à 3. Une liste de préfixes, suffixes est disponible et utilisée par le programme du stemmer. Cette liste est différente selon que le stemming est lourd ou léger.

• PHASE 3 : Traitement du corpus

Dans cette démarche, l'approche de similarité statistique « Corpus-Based » est adoptée à cause du manque de ressources arabe comme les dictionnaires et les lexicons dont l'approche topologique « knowledge-based » impose. Le modèle BOW¹⁰ est adopté qui ne prend pas en considération l'ordre des mots. L'approche se base sur le concept disant que les mots qui sont sémantiquement liés se trouvent dans le même contexte. Pour cela cette approche repose sur la notion du voisinage (mots voisins). Cependant la présence d'un corpus qui sera manipulé à l'aide du modèle BOW est indispensable. En outre, une fenêtre avec une taille prédéfinie est importante pour concrétiser la notion du voisinage.

La « Figure 13 » montre le processus effectué afin de générer les trois espaces sémantiques à l'aide des trois corpus BBC+CNN [18], CNN [18] et Al Khaleej [15] . Les détails de ce processus sont ci-dessous :

a) Les corpus sont acquis et prétraités selon l'étape du « Prétraitement du corpus ».

b) **Construction de la matrice des cooccurrences** : Ou matrice de fréquences/poids. Cette étape signifie la transformation des données textuelles au sein du corpus en données numériques sous forme d'une matrice. Chaque case représente la somme des poids de l'apparence du $i^{ème}$ terme (mot) avec le $j^{ème}$ terme. Cette somme est calculée par rapport à l'emplacement des termes du corpus autour du $i^{ème}$ terme (effectuer le même calcul pour chaque occurrence du $i^{ème}$ terme).

c) **Construction de la matrice des corrélations** :

¹⁰ Le modèle de sac de mots est un moyen d'extraire des caractéristiques du texte pour les utiliser dans des algorithmes de l'apprentissage automatique

a. Après avoir construit la matrice des cooccurrences, une stratégie de normalisation est appliquée. Donc une nouvelle matrice est construite en appliquant la formule ci-dessous prise de l'algorithme COALS [7] pour chaque case/élément de la matrice des cooccurrences.

$$\left\{ \begin{array}{l} w'_{a,b} = \frac{T w_{a,b} - \sum_j w_{a,j} \cdot \sum_i w_{i,b}}{(\sum_j w_{a,j} \cdot (T - \sum_j w_{a,j}) \cdot \sum_i w_{i,b} \cdot (T - \sum_i w_{i,b}))^{1/2}} \\ T = \sum_i \sum_j w_{i,j} \end{array} \right.$$

D'où :

- a et b sont les deux termes de la matrice de cooccurrences (ligne et colonne).
- $w_{a,b}$ est l'élément de la matrice de cooccurrences du terme a et b.
- i est l'indice des lignes tandis que j est celui des colonnes.
- $\sum_j w_{a,j}$ est la somme des colonnes de la ligne du terme a.
- $\sum_i w_{b,i}$ est la somme des lignes de la colonne du terme b.
- $T = \sum_i \sum_j w_{i,j}$ est la somme de tous les éléments de la matrice de cooccurrences.

Pour un large corpus, les valeurs de corrélations sont petites, alors, il est rare que la valeur de corrélation dépasse le 0.01. De plus, la majorité des corrélations sont négatives [7].

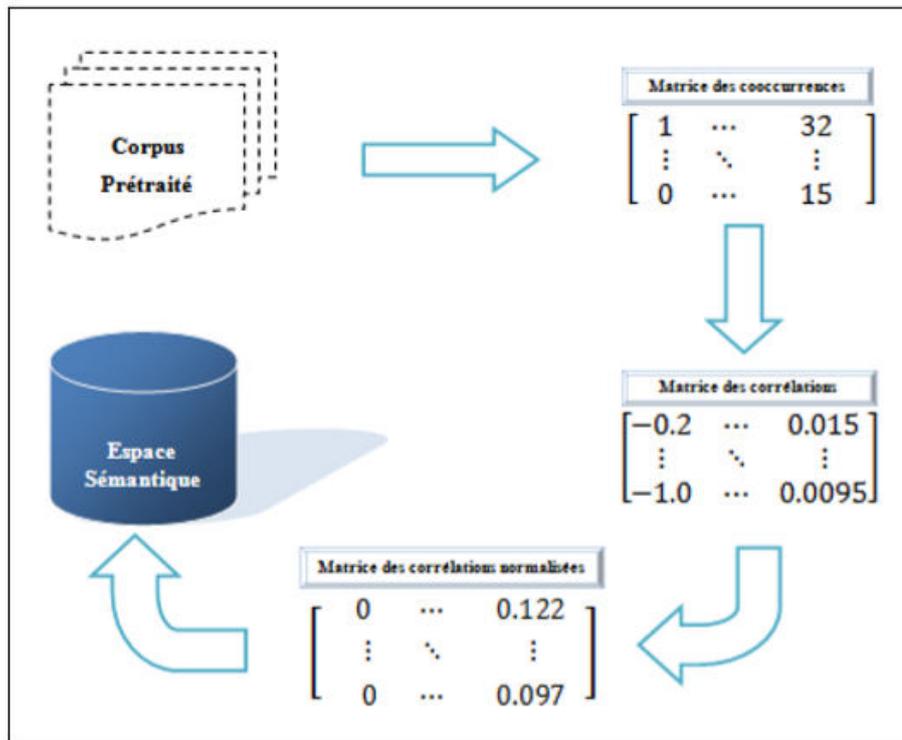


FIGURE 13 ALGORITHME DE COALS POUR LA CONSTRUCTION DE L'ESPACE SEMANTIQUE [7]

d) Construction de la matrice de corrélation normalisée :

Les corrélations négatives transportent très peu d'information, alors, une autre normalisation est effectuée. Les valeurs négatives sont normalisées à 0 tandis que les valeurs positives prennent leurs racines carrées afin d'amplifier l'importance des nombreuses petites valeurs par rapport aux grandes valeurs [19].

Cette étape représente la dernière étape de **création de l'espace sémantique** ou proprement dit : les vecteurs du contexte de chaque mot du corpus en entrée. Il est à noter que l'espace sémantique est généré qu'une seule fois. Il est donc stocké pour une utilisation ultérieure

- **PHASE 4 : Post traitement du corpus**

En effet, la similarité sémantique est fondée sur l'idée disant que les termes similaires se trouvent fréquemment dans le même contexte. Néanmoins, dans un contexte, les termes n'ont pas tous la même importance. Par conséquent, il est plus judicieux de pondérer l'impact des termes avant l'application d'une mesure de similarité. La fréquence d'apparition d'un terme dans un document est un bon indicateur de l'importance de ce terme.

Une fonction de pondération attribue à chaque terme de chaque document une valeur. Cette valeur (ou poids) est calculée en tenant compte de deux grands critères [20] :

✓ La force (capacité) locale du terme dans le document (c'est-à-dire mesurer l'importance du terme dans le document dans lequel il paraît).

✓ La force globale (c'est-à-dire mesurer l'importance du terme dans tout le corpus). Plus un terme est présent (fréquent) dans un document, plus sa force locale est importante et plus ce terme est présent dans le corpus, plus sa force globale est élevée.

1. TF (Term-Frequency) : Ou la fréquence du terme dans un document.

Cette pondération repose sur le calcul de la fréquence du terme dans le document (le nombre de fois que le terme apparaît dans le document). Plus un terme est fréquent dans un document plus il est important dans la description de ce document.

2. IDF (Inverse-Document-Frequency) : Ou la fréquence inverse du document.

C'est une mesure de l'importance du terme dans l'ensemble du corpus. La fréquence du document (DF) est le nombre de documents du corpus dont un terme apparaît tandis que la fréquence inverse du document (IDF) est le logarithme de l'inverse de la proportion de documents du corpus qui contiennent le terme. Cette norme vise à donner un poids plus important aux termes les moins fréquents, considérés comme plus discriminants (par exemple les stop-words ont généralement une fréquence élevée alors qu'ils ne sont pas vraiment importants par rapport aux autres mots moins fréquents qu'eux).

3. TFIDF :

Cette méthode représente l'importance du terme dans le document relativement dans le corpus de document. Elle s'obtient en multipliant les deux mesures TF et IDF :

$$\text{TFIDF} = \text{TF} * \text{IDF}$$

Après la construction de l'espace sémantique, ce dernier doit être sauvegardé dans un fichier « .txt ». Chaque ligne de ce fichier représente un vecteur d'un mot. Les éléments d'un vecteur du mot sont séparés par des espaces simples. Chaque élément est sous format chaîne de caractères « String ».

2. ETAPE 2 : Modèles du calcul de similarité sémantique/syntaxique entre deux réponses courtes :

Après la construction de l'espace sémantique, nous entamons la notion du calcul de similarité sémantique. Dans ce travail, deux niveaux du calcul de similarité sont considérés :

- a. Similarité entre phrases : ce concept est réalisé par un modèle nommé « somme vecteurs » [7].
- b. Similarité mot-à-mot : concrétisé ce concept est concrétisé par un modèle de calcul matriciel.
- c. Une hybridation avec des mesures syntaxiques.

- **Le modèle somme-vecteurs (SV) :**

Les étapes du fonctionnement de la SV sont détaillées comme suit :

- a) Une étape de prétraitement (normalisation, stemming,...) est effectuée sur les deux réponses (RM et RE). Par la suite, chaque réponse est transformée en vecteur en appliquant une tokenisation. Deux vecteurs de mots résultent de cette étape.
- b) Maintenant, les mots de chaque réponse sont remplacés par leurs vecteurs de contexte. Cela est fait en récupérant le vecteur de contexte du mot à partir de l'espace sémantique construit et stocké précédemment. Un vecteur de vecteurs résulte à ce stade pour chaque réponse.
- c) Ensuite, des valeurs de pondération de chaque mot vont être récupérées. Ici deux cas se distinguent : avec une pondération et sans aucune pondération. Dans le cas de pondération, chaque vecteur de contexte du mot i est multiplié par la valeur de pondération qui représente l'importance du mot i .
- d) Chaque vecteur de vecteurs de contexte passe par une étape de sommation. Cela est obtenue en sommant les vecteurs de contexte de chaque réponse mot par mot ou proprement dit case par case.
- e) En arrivant au stade du calcul de la valeur de similarité entre les deux réponses. La mesure **Cosine** sur les vecteurs $V(RM)$ et $V(RE)$ est appliquée. Une valeur entre 0 et 1 est obtenue et qui représente la valeur de similarité entre les deux réponses.

- **Le modèle calcul-matriciel (CM) :**

Contrairement au modèle précédent, une matrice est construite dont les termes d'une réponse représentent les lignes de la matrice tandis que les termes de l'autre réponse représentent les colonnes. Par la suite, chaque élément de la matrice signifie la

valeur de similarité entre le vecteur de contexte du $i^{\text{ème}}$ terme (réponse 1) et celui du $j^{\text{ème}}$ terme (réponse 2).

- **L'indice de Dice :**

La méthode syntaxique l'indice de Dice qui mesure la similarité entre deux phrase p_1 et p_2 en se basant sur le nombre de termes communs.

- **Hybridation :**

Dans cette phase-là, la notion d'hybridation entre les mesures pour avoir de meilleurs résultats. Pour ce faire, une combinaison externe est adoptée c'est-à-dire en combinant une mesure de similarité sémantique avec une mesure syntaxique et la moyenne des deux valeurs de similarité (calculé par SV et Indice de Dice) ou (par CM et Indice de Dice) entre les deux réponses. Les résultats seront présentés dans le chapitre cinq.

3. ETAPE 3 : Passage au score :

Après avoir mesuré la valeur de similarité entre les deux réponses, un passage au score de la valeur de similarité est effectué vers une note ou score selon le barème donné. Pour ce faire, la technique suivante est adoptée :

Classifieur k-means [21] (Ou k-moyenne)

C'est une technique de classification non supervisée. Elle consiste à avoir un ensemble de données comme entrée (input) et elle donne en sortie ces données classifiées dans K classes. C'est un algorithme très populaire et facile dans sa mise en œuvre. L'inconvénient de cet algorithme réside dans l'exigence de la détermination du nombre de classe comme entrée.

Dans le chapitre 5, les couples de réponses dataset utilisés dans l'évaluation de cette approche sont notés sur 5 points. K est fixé à 11 pour avoir 11 classes de note (0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5) afin de prendre en compte la subjectivité du processus d'évaluation.

Conclusion :

Au cours de ce chapitre, nous venons de présenter les principes de système élaboré [7] dans le travail de l'année passée. Ce dernier est un ensemble de plusieurs modules intégrés pour effectuer une évaluation automatique de réponses courtes en langue arabe. Ce processus nécessite un prétraitement du corpus dont il sera utilisé pour créer le VSM « espace

sémantique vectoriel » ainsi que pour la génération des pondérations dans le but de tester son impact. Le processus suivant s'en charge du calcul de similarité qui sera convertis en score par la suite. Pour se faire, les auteurs [7] de ce travail sont inspirés d'autres travaux pour élaborer les modèles proposés (SV, CM, Hybridation).

Chapitre IV : Conception, développement et intégration du Plugin dans le système de Quiz de Moodle

Introduction :

Le principe général de l'évaluation consiste à comparer la réponse de l'apprenant avec la réponse **modèle** (de référence) formulée par l'enseignant et d'attribuer un score. Dans le cadre du projet actuel, plusieurs approches écrites en python ont été déjà développées l'année passée [7] [8] autour des espaces sémantiques. Dans la continuité des objectifs, le passage à l'échelle dans une situation réelle doit passer par l'intégration des approches dans une plateforme de télé-enseignement afin que l'évaluation puisse se faire en consolidation avec les cours qui se déroulent sur la même plateforme.

Nous travaillons dans le contexte de la plateforme **Moodle** et tenterons d'ajouter au système de **Quiz** de cette plateforme un module (Plugin) contenant les outils d'évaluation automatique des réponses courtes en langue arabe. Ce système ou module de **Quiz** peut sélectionner automatiquement des questions aléatoires et / ou spécifiques dans différentes catégories de questions et permet d'utiliser un système de feedback (Les commentaires) sur les performances et l'auto-évaluation qui sont des éléments importants d'un environnement d'apprentissage. Ce feedback peut afficher des commentaires et des scores à différents moments du test, à l'aide des options de révision des paramètres du Quiz or notre plugin doit s'adapter à ces fonctionnalités. Une grande variété de rapports de Quiz (en plus des notes) peut être utilisés par l'enseignant. Les rapports de Quiz (Test) peuvent non seulement se concentrer sur la tentative d'un seul élève de répondre à chaque question, mais ils peuvent également effectuer une analyse élémentaire de la validité d'une question basée sur les réponses agrégées de l'élève. Tous ces avantages vont jouer un rôle très important dans le cycle de conception et développement de notre Plugin de telle sorte que ce dernier doit répondre à tous les fonctionnalités de système de Quiz de Moodle et bien évidemment apparaitre dans la liste des plugins de type de question propres à cette plateforme Moodle.

Pour apprendre à programmer des types de questions dont notre plugin, la documentation de la plateforme Moodle encourage vivement à lire le code de certains des types de questions existants inclus dans la base de code Moodle (Par exemple : ShortAnswer, True/false questions...), et fournit un lien GitHub [22] redirigeant vers un simple modèle de type de question (Template vide) contenant des fichiers avec leurs fonctions vides. Ce modèle permet à implémenter n'importe quelle nouvelle approche qui est différente de celle (Expressions régulières...) implémentée dans le système actuel de Quiz.

1. Conception du Plugin d'évaluation automatique des réponses courtes :

Arrivant à la partie la plus importante dans ce chapitre, la conception de notre Plugin Arabic Short Answer qui se résume en quatre phases dans la figure 14 :

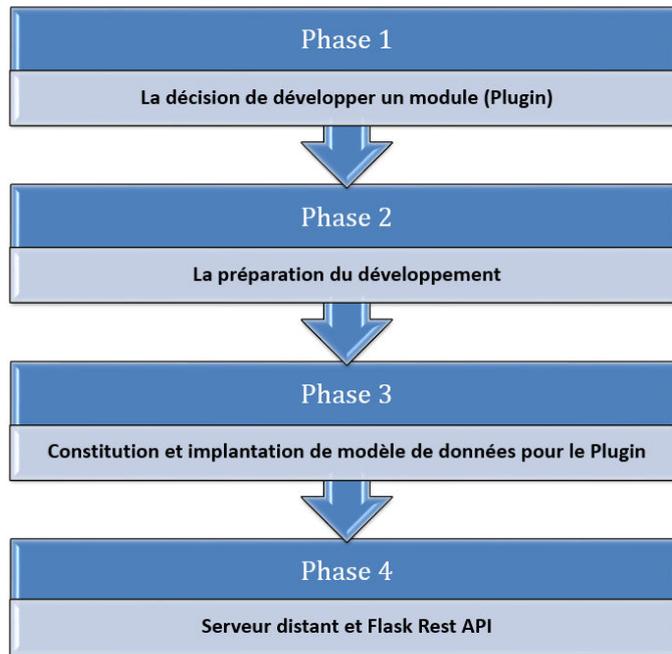


FIGURE 14 LES PHASES DE CONCEPTION DU PLUGIN ARABIC SHORT ANSWER

1.1- Phase 1 : La décision de développer un module (Plugin)

Un Plugin ou Plug-in, aussi nommé module d'extension, module externe, greffon, plugiciel, ainsi que add-in ou add-on, est un paquet qui complète un logiciel hôte pour lui apporter de nouvelles fonctionnalités. Dans notre cas, Il présente de ce fait un certain nombre de fonctionnalités particulières que les autres extensions d'évaluation automatique de Moodle ne possèdent pas surtout celle de « ShortAnswer » :

- Extensible comme un moule de manière qu'il possède certaines entrées et sorties où nous pouvons lui assigner n'importe quelle approche d'évaluation.
- Navigation facile pour les étudiants permettant une liberté de formulation des réponses courtes.
- Utilisation très aisée pour les enseignants sans avoir des requis sur la formulation des réponses modèles.
- Modélisation et utilisation des ressources de traitement automatique des langues.

- Evaluation automatique performante des réponses courtes et proche de l'évaluation manuelle.

Cette figure (15) au-dessous montre le moule qu'on veut concevoir :

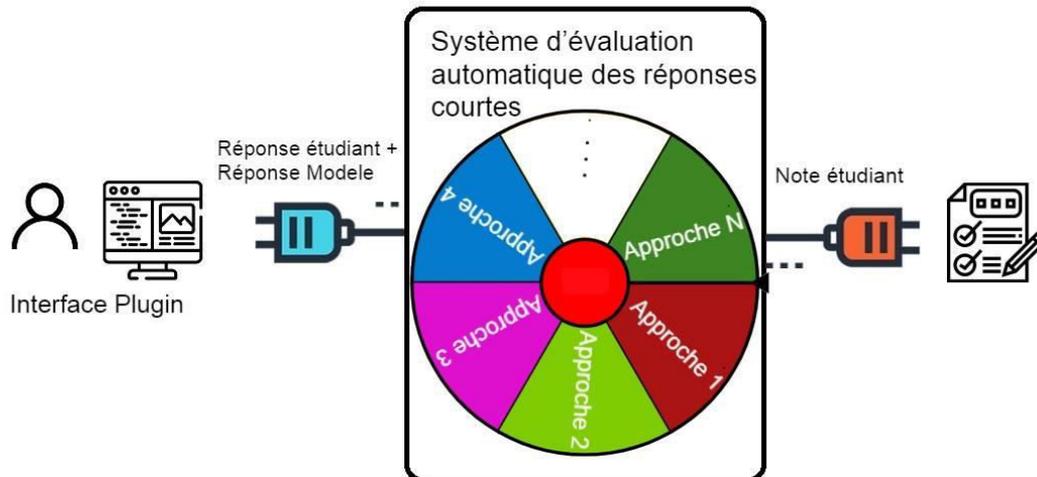


FIGURE 15 CONCEPTION D'UN MOULE POUR N'IMPORTE QUELLE APPROCHE D'EVALUATION AUTOMATIQUE

Ce plugin « Arabic ShortAnswer » nécessite un travail de recherche, l'analyse et la compréhension de la conception du Question Engine (Moteur de question) de la plateforme Moodle, des compétences en langage MySQL pour les bases de données, en programmation orientée objet en PHP, HTML et CSS pour les formulaires et des notions de base en programmation en langage PYTHON.

1.2 Phase 2 : La préparation du développement

La deuxième phase commence par une étude du "modèle propre" du Plugin, c'est-à-dire, de sa modélisation fonctionnelle indépendante du contexte. Les méthodes sont multiples, mais le minimum à apporter est plus ou moins :

- Les méthodes fonctionnelles (également qualifiées de méthodes structurées) trouvent leur origine dans les langages procéduraux. Elles mettent en évidence les fonctions à assurer et proposent une approche hiérarchique descendante et modulaire. L'analyse fonctionnelle est une démarche qui consiste à rechercher et à caractériser les fonctions offertes par un produit pour satisfaire les besoins de son utilisateur. La figure suivante explique cette démarche.

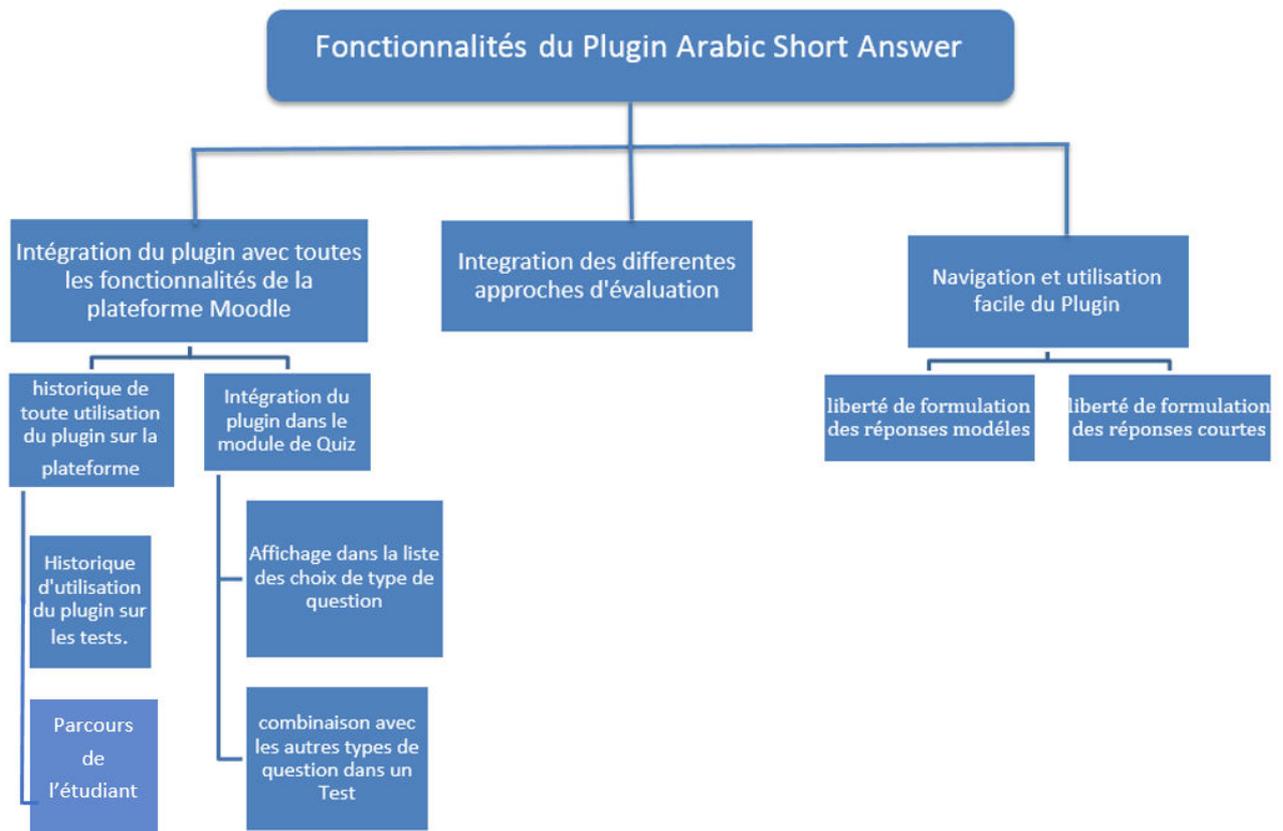


FIGURE 16 : FONCTIONNALITES DU PLUGIN ARABIC SHORT ANSWER

- Des scénarios de cas d'utilisation de ces fonctionnalités, considérant les deux rôles :
 - Étudiant.
 - Enseignant.

C'est un modèle de départ très simpliste auquel nous ajoutons rapidement des altérations. En fait, il existe deux scénarios de cas d'utilisation :

1. Le premier scénario se résume sur l'installation du Plugin, c'est-à-dire, ce que l'administrateur consulte dans la partie visible de l'écran lors de l'installation du Plugin. Grâce à l'intégration de notre Plugin, ce dernier exploite toutes les fonctionnalités de la plateforme Moodle et interagit avec le système de quiz comme tous les Plugins par défaut de cette plateforme. Voir la figure 17.
2. Le deuxième scénario consiste à résumer la partie qui suit l'installation de Plugin, c'est-à-dire, le rôle des enseignants et étudiants avant et après le test (quiz). Notre Plugin facilite la tâche aux enseignants pour élaborer le test comme il répond à toutes les attentes des étudiants afin de réaliser un test en ligne en toute sécurité. Voir la figure 18.

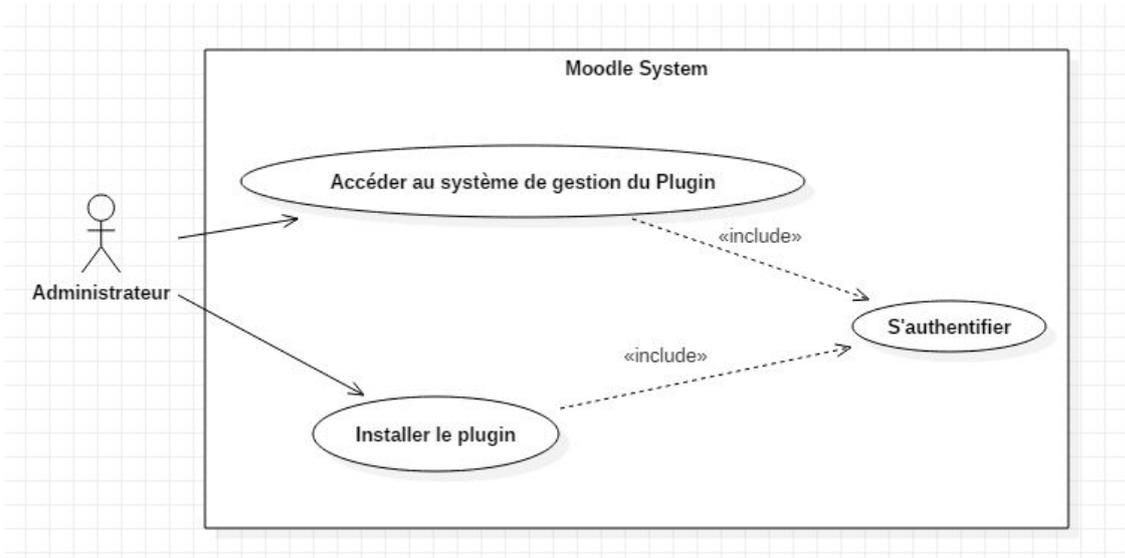


FIGURE 17 DIAGRAMME DE CAS D'UTILISATION « INSTALLATION DE PLUGIN »

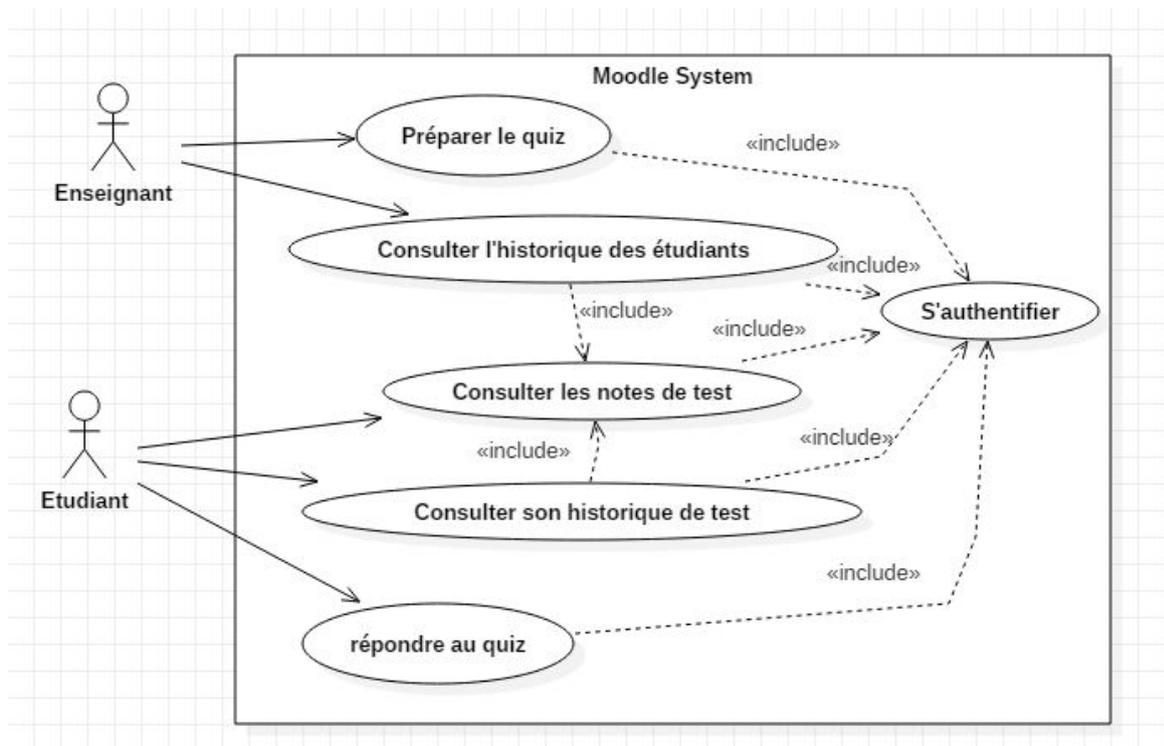


FIGURE 18 DIAGRAMME DE CAS D'UTILISATION « ROLE DES ENTITES DE TEST »

1.3 Phase 3 : Constitution et implantation de modèle de données pour le Plugin

Jeu de tables de Plugin :

Le jeu de tables de Plugin est en grande partie libre. Il existe cependant trois règles à suivre [23]:

- ❖ Il existe dans chaque module de type de question une table nommée `{ $CFG->prefix } { qtype_pluginname_options }` (exemple, si le préfixe est le préfixe par défaut 'mdl_' : **mdl_qtype_arabicshortanswer_options** en considérant **arabicshortanswer** comme le nom de notre Plugin). Cette table contient les paramètres mémorisés pour chaque instance (une instance est obtenue à chaque fois qu'on insère une question dans un cours). La seule différence entre la table « `mdl_qtype_pluginname_options` » de notre plugin « **arabicshortanswer** » et celle « `mdl_qtype_shortanswer_options` » de module par défaut de Moodle « **ShortAnswer** » se résume dans la colonne « **usecase** ». Cette dernière est un paramètre booléen de la table du **Plugin shortanswer** de Moodle, elle peut être délicate lorsque la capitalisation est importante. Par exemple, « Accepterons-nous **Ban Ki-moon** ou **ban ki-moon** comme réponse ? » Si le paramètre **usecase** est activé, les deux réponses sont considérées correctes, sinon l'une des réponses sera considérée comme étant réponse fausse. Dans la table de notre Plugin **Arabic ShortAnswer**, nous n'aurons pas besoin de ce paramètre car la langue Arabe ne contient pas les lettres majuscules dans son alphabet.
- ❖ Cette table contient une seule catégorie de champ :
 - Les champs obligatoires
 - INT(10) **id** auto_increment UNSIGNED PRIMARY KEY : l'identifiant numérique d'instance de notre type de question.
 - INT(10) **questionid** UNSIGNED : la clef étrangère sur la question où l'instance est implantée.

Voici en figures (19 et 20), les deux tables de chaque type de question via **PhpMyAdmin** qu'on vient de citer au-dessus :

				id	questionid			
<input type="checkbox"/>		Éditer		Copier		Supprimer	1	1
<input type="checkbox"/>		Éditer		Copier		Supprimer	2	2
<input type="checkbox"/>		Éditer		Copier		Supprimer	3	3
<input type="checkbox"/>		Éditer		Copier		Supprimer	4	4
<input type="checkbox"/>		Éditer		Copier		Supprimer	5	5
<input type="checkbox"/>		Éditer		Copier		Supprimer	6	6

FIGURE 19 TABLE QTYPE_ARABICSHORTANSWER_OPTIONS VIA PHPMYADMIN

				id	questionid	usecase			
<input type="checkbox"/>		Éditer		Copier		Supprimer	1	1	0
<input type="checkbox"/>		Éditer		Copier		Supprimer	2	2	1
<input type="checkbox"/>		Éditer		Copier		Supprimer	3	3	0
<input type="checkbox"/>		Éditer		Copier		Supprimer	4	4	0
<input type="checkbox"/>		Éditer		Copier		Supprimer	5	5	1
<input type="checkbox"/>		Éditer		Copier		Supprimer	6	6	0

FIGURE 20 TABLE QTYPE_SHORTANSWER_OPTIONS VIA PHPMYADMIN

- ❖ Les autres tables qui ont relation avec notre propre table de type de question **qtype_arabicshortanswer_options**. Chaque classe d'une tentative de question **question_attempts** étant une liste de **question_attempt_steps**. Ceci est naturellement stocké dans trois tables de base de données liées, sauf que comme chaque étape peut contenir une liste arbitraire de paires nom => valeur, nous obtenons une autre table **question** propre à la question elle-même. Cette dernière contient plusieurs tentatives de question. La table pour notre type de question est reliée directement avec la table **question**. Ceci est montré dans la figure au-dessous :

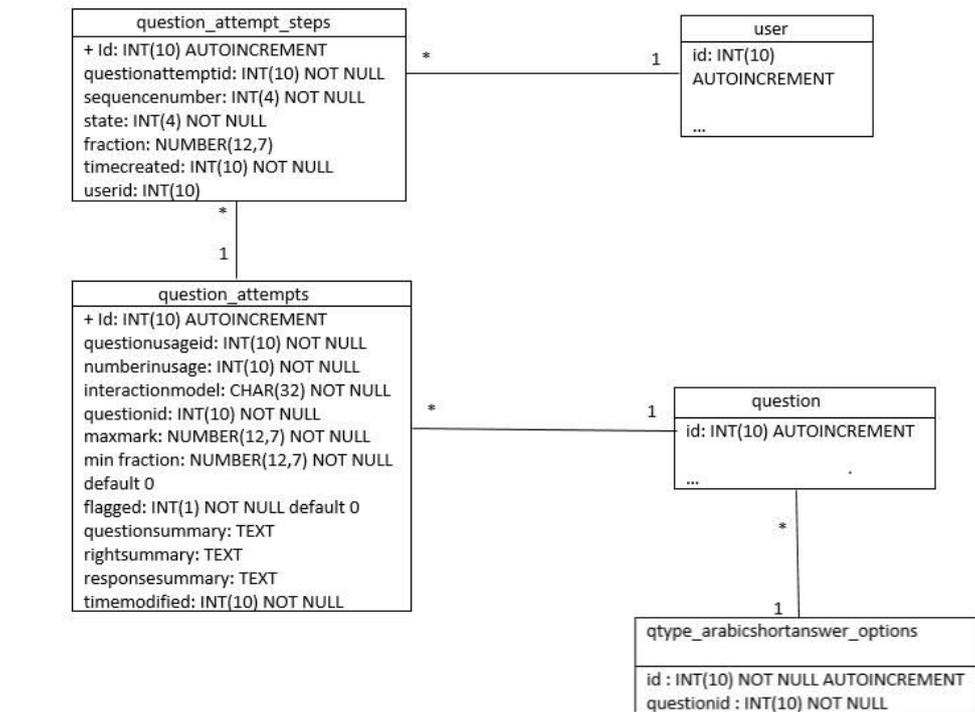


FIGURE 21 MODELE E/A DE LA BASE DE DONNEE DU MOTEUR DE QUESTION

1.4- Phase 4 : Déploiement sur le Serveur distant et Flask Rest API

Cette dernière phase concerne l'approche sémantique développée l'année passée par le binôme pour comparer entre les couples des réponses courtes des enseignants et étudiants. Cette approche est développée en langage PYTHON et vu que le système de Moodle supporte que le langage PHP, nous avons réfléchi à intégrer leur application dans un serveur distant qui supporte le langage requiert. La figure 22 résume cette phase.

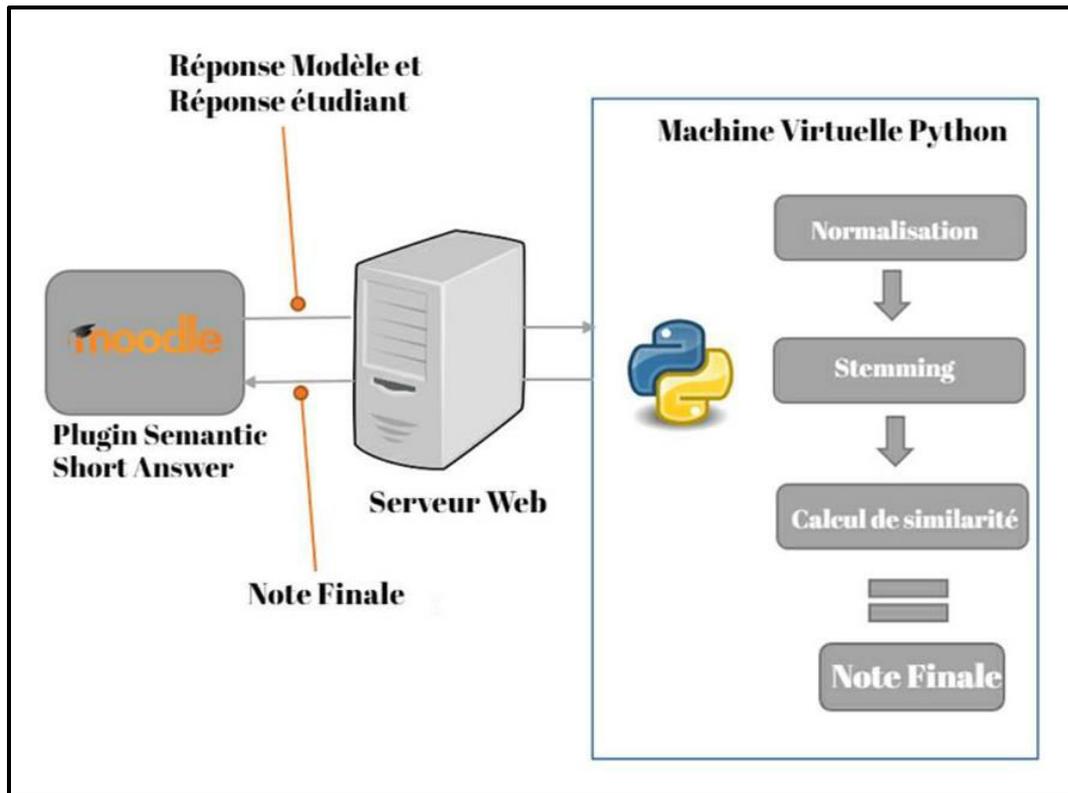


FIGURE 22 VU D'ENSEMBLE DU PLUGIN « ARABIC SHORTANSWER »

PythonAnywhere est un environnement de développement intégré en ligne (IDE) et un service d'hébergement Web (Platform as a service) basé sur le langage de programmation Python. Fondé par Giles Thomas et Robert Smithson en 2012, il fournit un accès en navigateur aux interfaces de ligne de commande Python et Bash¹¹ basées sur un serveur, ainsi qu'un éditeur de code avec mise en évidence de la syntaxe. Les fichiers de programme peuvent être transférés vers et depuis le service à l'aide du navigateur de l'utilisateur. Les applications Web hébergées par le service peuvent être écrites à l'aide de n'importe quel Framework¹² d'application basé sur WSGI¹³ comme Flask dans notre cas. PythonAnywhere a été créé par Resolver Systems, qui a également produit Resolver One, un programme de tableur basé sur Python.

Concept de Flask API : [24]

Flask est un Framework web, ou plutôt, un micro-framework. Ce “micro” signifie simplement que Flask ne fait pas tout. Cela signifie aussi que pour en faire plus que ce qu'il permet de base, il faudra installer des extensions. Heureusement, celles-ci sont nombreuses, de qualité, et très bien intégrées. Flask est un ensemble de modules qui vont faciliter la

¹¹ Bash est un interpréteur en ligne de commande de type script.

¹² désigne un ensemble cohérent de composants logiciels structurels, qui sert à créer les fondations ainsi que les grandes lignes de tout ou d'une partie d'un logiciel (architecture).

¹³ La Web Server Gateway Interface (WSGI) est une spécification qui définit une interface entre des serveurs et des applications web pour le langage Python.

programmation de notre application web dynamique, celle qui contient l'approche. En effet, il suffit que notre application suive la norme WSGI¹⁴

Flask fournit un modèle simple pour le développement web. Une fois importé en python, Flask peut être utilisé pour gagner du temps dans la construction d'applications Web. Il garde le noyau simple mais extensible. Il n'a pas la couche d'abstraction de base de données, validation de formulaire ou tout autre composant. Flask supporte les extensions. Ces extensions existent pour les objets relationnels mappeurs, validation de formulaire, traitement du téléchargement et plus.

Comme illustré sur la figure 23, nous avons réussi à trouver le **Flask API** qui joue le même rôle de Jobe REST API (CodeRunner) qui est un remplaçant instantané de Flask (Flask est un Framework open-source de développement web en Python) qui fournit une implémentation d'API navigable et donne correctement le contenu des réponses négociées et l'analyse intelligente des requêtes. A travers le Flask API, nous pouvons offrir des services à d'autres logiciels ou plateformes web (Moodle plugin par exemple) mais **comment se fait l'interaction ?**

Nous avons utilisé le cUrl pour dialoguer avec notre Flask API en PHP. Des lignes de codes de cUrl en PHP dans le fichier « question.php » de notre plugin servent à récupérer le contenu d'une ressource accessible par un réseau informatique. La ressource est désignée à l'aide d'une URL. Nous avons aussi utilisé un serveur distant PythonAnywhere qui est un environnement en ligne de développement intégré (IDE) et un service d'hébergement Web (Platform as a service) basé sur le langage de programmation Python seulement pour pouvoir garantir le transfert de nos ressources en ligne dont l'approche déjà développée en python par le binôme.

A ce stade, nous avons réussi d'intégrer facilement un programme python en FLASK (qui est un Framework de développement web en python) dans un IDE en ligne PythonAnywhere, d'interagir avec notre FLASK API à travers des lignes de code cUrl en PHP et de transférer des ressources via l'URL. Ces ressources vont désigner les réponses courtes et la note correspondante. Nous reviendrons à cette partie avec plus de détail dans la partie de développement.

¹⁴ https://fr.wikipedia.org/wiki/Web_Server_Gateway_Interface



FIGURE 23 CONNEXION AVEC LE SERVEUR DISTANT PYTHONANYWHERE A TRAVERS FLASK API

2. Développement du Plugin d'évaluation automatique des réponses courtes :

Le développement de notre Plugin consiste à développer un nouveau type de question qui fonctionne avec le moteur de question Moodle présenté dans la partie 5 de chapitre 2. Comme nous avons déjà affirmé que la documentation Moodle encourage les développeurs de lire le code de certains des types de questions existants inclus dans la base de code Moodle. Cela nous guide à réaliser notre Plugin Arabic ShortAnswer et l'adapter avec le moteur de question de la plateforme. Il existe également de nombreux autres exemples de types de questions dans le référentiel GitHub de « Open University » [25]. Et encore plus d'exemples de types de questions dans la catégorie de type de question de Moodle de la base de données des plugins [26].

2.1 Modèle de plugin de type question :

Le simple modèle de type de question (Template vide) fourni par la documentation de Moodle, est une alternative pour commencer à développer un Plugin de type question et fonctionne correctement. Bien qu'il ne fasse pas de classification ou notation réelle (grading) ou recueille des entrées claviers des étudiants. Notons également que tout le code du moteur de questions contient de nombreux commentaires du documentation PHP qui devraient expliquer le but de chaque classe et méthode. Ce document [13] donne un aperçu des points clés pour démarrer. Il ne tente pas de dupliquer tous les détails dans les PHPdocs. Ce document suppose que nous comprenons le modèle de données, dans lequel une tentative de question comprend plusieurs étapes et chaque étape possède des propriétés telles qu'un état et une marque, ainsi qu'un tableau de données soumises.

2.2 Disposition du fichier du Plugin :

Le Plugin « Arabic ShortAnswer » comme d'autres types de questions réside dans un dossier question/type. La mise en page de ce dossier suit la mise en page typique d'un plugin Moodle. Dans le dossier question/type/arabicshortanswer/ nous allons programmer notre type de question et remplir nos fonctions. Voici les fichiers du Plugin :

- **edit_arabicshortanswer_form.php**

Le formulaire de Moodle, est utilisé pour définir le formulaire de modification de notre question. Ce fichier PHP contient la classe **qtype_arabicshortanswer_edit_form** qui hérite de la classe **question_edit_form**. Ce tableau au-dessous contient les fonctions qui interagissent dans ce fichier avec leurs paramètres et rôles.

FONCTIONS	ROLES	Paramètres d'entrée	Paramètres de retour
protected function definition_inner(\$mform)	Formulaire d'ajout d'une question.	\$mform : Elément de formulaire.	N/A
protected function data_preprocessing(\$question)	Effectue un prétraitement nécessaire sur les données transmises à la fonction set_data () avant qu'il soit utilisé pour initialiser le formulaire.	\$question : l'objet question représente les données transmises au formulaire.	\$question : l'objet question les données modifiées.
public function qtype()	Récupère le nom de type de question	N/A	Chaine de caractère 'arabicshortanswer'

TABLEAU 7 : FONCTIONS DU FICHIER EDIT_ARABICSHORTANSWER_FORM.PHP

- **questiontype.php**

Définit la classe **qtype_arabicshortanswer** qui hérite de la classe **question_type**. Cela doit étendre cette dernière. Ce fichier est utilisé pour sauvegarder ou supprimer

des questions de la base de données. Le tableau au-dessous contient les fonctions nécessaires avec leurs paramètres et rôles.

FONCTIONS	ROLES	Paramètres d'entrée	Paramètres de retour
protected function delete_files(\$questionid, \$contextid)	Supprimer tous les fichiers appartenant à cette question	1- \$questionid : de type int. La question en cours de suppression 2- \$contextid : de type int. Le contexte de la question est	N/A
public function save_question_options(\$question)	Ajouter et sauvegarder la question dans la base de données.	\$question : l'objet question représente les données transmises au formulaire	N/A
Protected function initialise_question_instance(question_definition \$question, \$questiondata)	Initialise les champs communs de question_definition par exemple (temps de création, catégorie, id du contexte, feedback général ...)	1- question_definition \$question : la définition de question que nous créons. 2- Objet \$questiondata : les données de la question chargées depuis la base de données.	N/A
public function get_possible_responses(\$questiondata)	Récupère toutes les réponses possibles à cette question (Dans ce type il n'y a qu'une seule réponse de l'étudiant pour chaque question)	Objet \$questiondata : les données de la question chargées depuis la base de données.	array(\$questiondata->id => \$responses); Ce tableau associatif représente le champ et la valeur. Il contient les identifiants des réponses.
public function delete_question(\$questionid, \$contextid)	Supprime la question de toutes les tables (La table spécifique et la table générale)	1- \$questionid : de type int. La question en cours de suppression. 2- \$contextid : Le contexte auquel cette question appartient.	N/A
public function get_question_options(\$question)	Récupère les données de la question.	\$question : l'objet question représente les données transmises au formulaire	Cette méthode retourne true.

TABLEAU 8 : FONCTIONS DU FICHIER QUESTIONTYPE.PHP

- **question.php**

Ceci contient la définition de la classe **qtype_arabicshortanswer_question**, qui devrait hériter de la classe **question_graded_automatically**. Dans le cas de Plugin par

défaut **Shortanswer**, cette classe hérite de la classe **question_graded_by_strategy** ce qui n'est pas notre cas car le plugin shortanswer suit une stratégie basée sur les expressions régulières. Le rôle principal de ce fichier est de comparer la réponse de l'élève à celle de l'enseignant. Ce fichier contient aussi les données d'entrée et de sortie du Plugin. Les données de sortie sont les réponses modèles et étudiants tandis que le score final de l'évaluation de chaque couple de réponses est une donnée d'entrée. Voici les fonctions concernées de ce fichier.

FONCTIONS	ROLES	Paramètres d'entrée	Paramètres de retour
Public fonction <code>__construct()</code>	Constructeur par défaut qui hérite du constructeur de la classe mère question_graded_automatically	N/A	N/A
public fonction <code>get_expected_data()</code>	Quelles données peuvent être incluses dans la soumission du formulaire quand un étudiant répond à une question dans son état actuel ? (Données attendues) La classe <code>question_attempt</code> appelle à une fonction statique, <code>question_attempt::get_field_prefix()</code> est automatiquement ajouté au début.	N/A	Retourne <code>array('answer' => PARAM_RAW) : array string nom de variable =>PARAM ...</code> constante ou, comme cas spécial à ne pas utiliser uniquement, la constante <code>question_attempt::USE_RAW_DATA</code> prend toutes les données brutes soumises appartenant à cette question.
public fonction <code>summarise_response(array \$response)</code>	Envoie la réponse soumise par l'étudiant vers la base de données	Tableau <code>\$response</code> : des réponses, comme cela pourraient être passé à la fonction <code>grade_response()</code>	<code>\$response['answer']</code> : une chaîne de caractère contient un résumé en texte brut de cette réponse, qui pourrait être utilisé dans les rapports.
public fonction <code>is_complete_response(array \$response)</code>	Vérifie si l'état de la réponse étudiant est complet ou incomplet	Tableau <code>\$response</code> : des réponses, comme cela pourraient être passé à la fonction <code>grade_response()</code>	Elle retourne un booléen si cette réponse est une réponse complète à cette question
public fonction <code>get_validation_error(array \$response)</code>	Détermine si la réponse de l'étudiant est valide ou pas.	Tableau <code>\$response</code> : des réponses, tel que retournés par la fonction statique <code>question_attempt_step::get_qt_data()</code>	Elle retourne un booléen, si cette réponse peut être notée. Si elle retourne FAUX, il faut générer une description pour ce problème.

public function is_same_response (array \$prevresponse, array \$newresponse)	Si l'étudiant change sa réponse et qu'elle est identique à sa réponse précédente, la nouvelle réponse peut être écarté et ne pas pris en considération.	1- array \$prevresponse : les réponses précédemment enregistrées pour cette question, tel que retournées par la fonction statique question_attempt_step::get_qt_data() 2- array \$newresponse : les nouvelles réponses, dans le même format.	Cette méthode retourne un booléen si les deux ensembles de réponses sont les mêmes - c'est-à-dire si le nouvel ensemble des réponses peut être rejeté en toute sécurité.
public function get_correct_response()	Retourne la réponse modèle de l'enseignant.	N/A	array('answer' => \$answer) Cette méthode retourne la réponse modèle de l'enseignant
public static function checkInput(\$var)	Assure la Sécurité des réponses soumises par l'étudiant et l'enseignant	\$var : une réponse (modèle/étudiant), avant l'application de quelques fonctions PHP de sécurité des flux de données.	Cette méthode retourne une réponse (modèle/étudiant) bien sécurisée.
protected static function safe_normalize(\$string)	Normalise une chaîne Utf-8 en FORM_C, en évitant les pièges de PHP	\$string : la chaîne de caractère d'entrée (réponse).	\$normalised : La chaîne de caractère normalisée.
public function grade_response(array \$response)	Evalue la réponse de l'étudiant et retourne la note et l'état de la question. Dans cette fonction, nous gérons les entrées et les sorties de Plugin. Nous envoyons les couples de réponses au serveur distant et nous recevons la note finale.	array \$response : Réponses, telles que retournées par la fonction question_attempt_step::get_qt_data()	array(float, integer) : La fraction et l'état

TABLEAU 9 : FONCTIONS DU FICHIER QUESTION.PHP

- **renderer.php**

Ceci contient la définition de la classe **qtype_arabicshortanswer_renderer**, qui devrait étendre la classe de base **qtype_renderer**. Il Permet d'afficher l'interface pour répondre à la question et le feedback. Nous trouvons aussi du code html dedans.

FONCTIONS	ROLES	Paramètres d'entrée	Paramètres de retour
public function formulation_and_controls(question_attempt \$qa, question_display_options \$options)	Génère l'output de la question.	1- question_attempt \$qa : Nous stockons dans cet objet, notre tentative pour répondre à la question. 2- question_display_options \$options : Nous stockons toutes les options d'une question dans cet objet.	\$result : Retourne l'output de la question.
public function specific_feedback(question_attempt \$qa)	Affiche les feedback	question_attempt \$qa : Nous stockons dans cet objet, notre tentative pour répondre à la question.	Retourne et affiche le feedback
public function correct_response(question_attempt \$qa)	Affiche la réponse modèle après une tentative	question_attempt \$qa : Nous stockons dans cet objet, notre tentative pour répondre à la question.	Retourne et affiche la réponse modèle.

TABLEAU 10 : FONCTIONS DU FICHIER RENDERER.PHP

- **lang/en/qtype_arabicshortanswer.php**

Ceci définit les langages que notre Plugin supporte. Mise à part la langue par défaut qui la langue anglaise, nous avons ajouté deux autres langues que notre Plugin peut supporter tel que le français et l'arabe. Ce sont des chaînes de caractère pour notre type de question. Nous pouvons aussi inclure d'autres langues. Le fichier de langue doit définir au moins les chaînes de caractère donnant un nom au modèle. La figure 24 montre le contenu du fichier **qtype.arabicshortanswer.php** :

```

20 * @package    qtype
21 * @subpackage semanticanswer
22 * @copyright  2019 Snoussi El Hareth & Madani Abderraouf
23 * @license    http://www.gnu.org/copyleft/gpl.html GNU GPL v3 or later
24 */
25
26 $string['pluginname'] = 'اجابة قصيرة دلالية';
27 $string['pluginname_help'] = 'فقط أدخل إجابة قالب وسيتم تعيين النتيجة تلقائيًا';
28 $string['pluginname_link'] = 'question/type/semanticanswer';
29 $string['pluginnameadding'] = 'إضافة سؤال';
30 $string['pluginnameediting'] = 'تعديل سؤال';
31 $string['pluginnamesummary'] = 'سؤال يقوم تلقائيًا بتعيين استجابات الطلاب بناءً على التشابه الدلالي مع استجابة نموذج المعلم';
32 $string['test'] = 'test';
33 $string['pleaseenterananswer'] = 'الرجاء إدخال إجابة';
34 $string['answer'] = 'إجابة: {$a}';
35 $string['correctansweris'] = '{$a}: الجواب الصحيح هو';
36 $string['modelanswer'] = 'الجواب النموذجي';
37 $string['modelansweris'] = '{$a}: الجواب النموذجي هو';
38 $string['provideanswer'] = 'يجب عليك تقديم إجابة';
39 $string['verygood'] = '! جيد جدا';
40 $string['good'] = '! جيد';
41 $string['satisfactory'] = 'لا بأس';
42 $string['average'] = 'متوسط';
43 $string['poor'] = 'ضعيف';
44 $string['failed'] = 'ضعيف جدا';

```

FIGURE 24 CONTENU DU FICHIER
LANG/AR/QTYPE.ARABICSHORTANSWER.PHP

- **lib.php**

Il y a un rappel (callback) important ici qui est appelé par le moteur de questions pour permettre l'accès aux fichiers utilisés par les questions. Comme avec les autres plugins, nous pouvons mettre du code de bibliothèque ici, mais dans le code moderne de Moodle, il est préférable d'utiliser le fichier classes/dossier. Chaque Plugin nécessite ce même fichier avec sa seule fonction.

- **Les fichiers db/install.xml, db/upgrade.php et db/upgradelib.php**

Pour créer toutes les tables de base de données requises, comme d'habitude. Tout Plugin nécessite les mêmes fichiers. Le système Moodle les génère par défaut.

- **version.php**

Dans ce fichier nous avons appliqué la dernière version de Moodle pour notre Plugin. Nous y trouvons des informations de version pour notre type de question. Nous pouvons également indiquer ici la version n ° de Moodle et / ou les autres types de questions requises par ce module. Augmenter la version de la question déclenchera le code approprié dans le fichier **upgrade.php** à exécuter.

- **styles.css**
Contient les styles utilisés par notre type de question. Ici nous préfaçons tous nos sélecteurs avec **que.qtype_arabicshortanswer** afin que les styles ne soient appliqués qu'à notre type de question, qui est encapsulé dans une div avec la classe 'que' et 'qtype_arabicshortanswer'.
- **pix/icon.svg**
C'est l'icône 16 * 16 px qui apparaît à côté de notre type de question dans le panneau de sélection du type de question. (Peut également être .png ou .gif).

3. Installation du Plugin et Intégration des outils d'évaluation automatique des réponses courtes :

3.1. Intégration et installation du Plugin :

Après la phase de développement du Plugin, il reste que l'intégrer dans la plateforme Moodle. Ce Plugin est un ensemble de fichiers qu'on va l'intégrer à partir de l'interface de Moodle. D'abord, Nous remplaçons toutes les occurrences de YOURQTYPENAME « Le nom de type de question » dans les fichiers par notre nouveau nom correspondant à notre type de question « arabicshortanswer ». Enfin, nous nous redirigeons vers l'interface de Moodle, dans la rubrique « Administration » de la page du tableau de bord de Moodle, nous cliquons sur l'option « Plugins », une liste s'affiche puis nous cliquons sur « Installer des plugins ». Une nouvelle page s'affiche or il reste que glisser notre fichier compressé du Plugin puis cliquer sur le bouton « INSTALLER LE PLUGIN A PARTIR DU FICHER ZIP ». Les figures 25, 26 et 27 montrent ce qu'on vient d'expliquer.

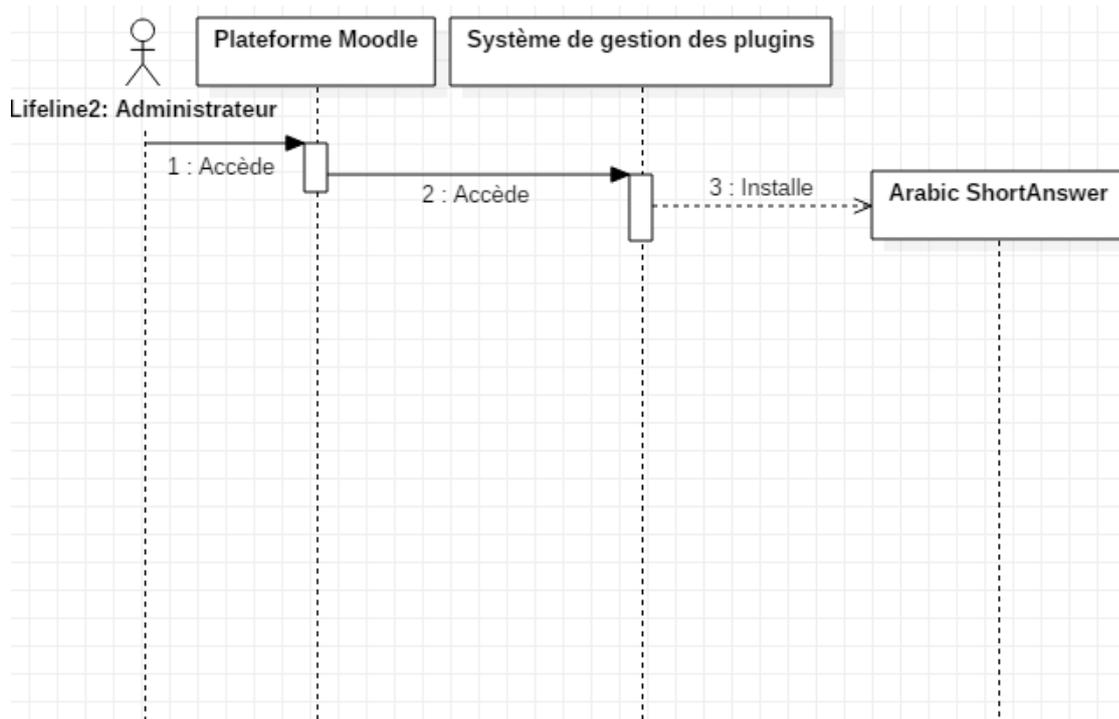


FIGURE 25 DIAGRAMME DE SEQUENCE «SCENARIO: INSTALLER PLUGIN»

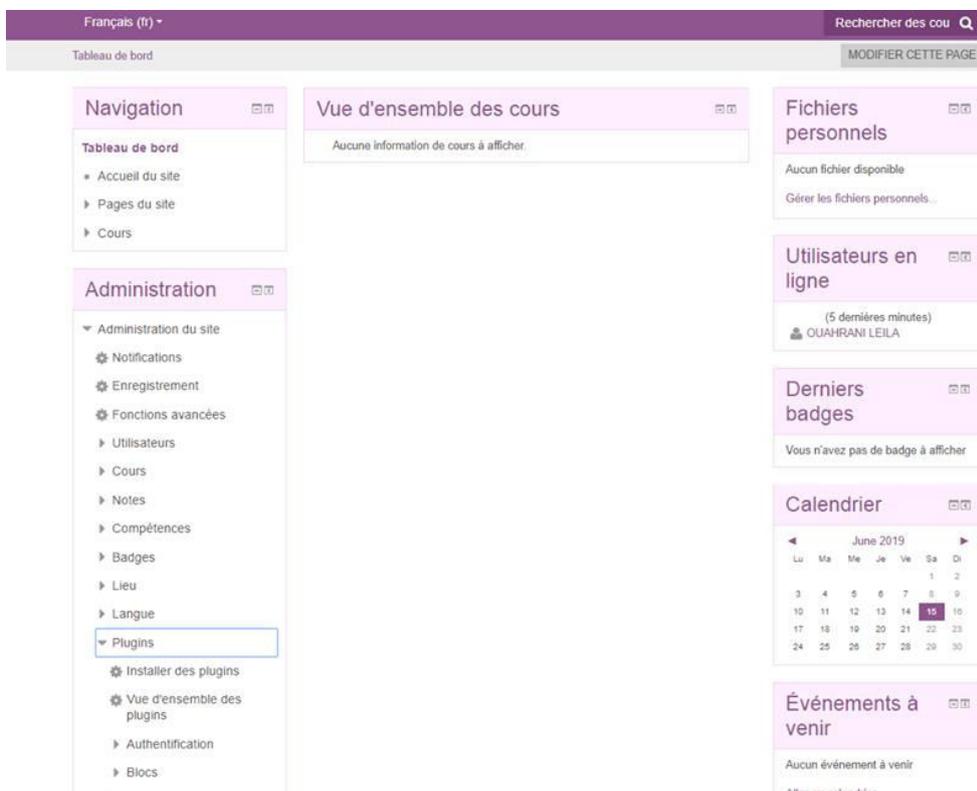


FIGURE 26 LA LISTE DE L'ÉLÉMENT « PLUGINS » DANS LA RUBRIQUE « ADMINISTRATION »

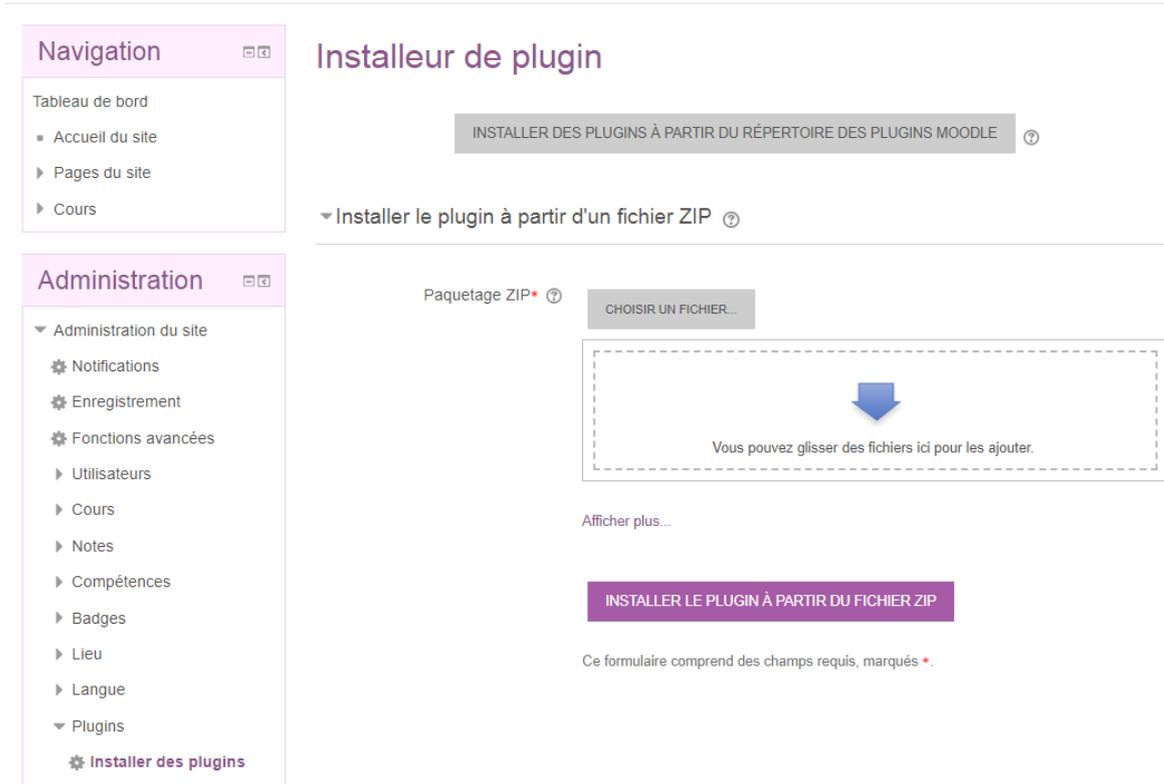


FIGURE 27 L’AFFICHAGE D’UNE NOUVELLE PAGE WEB POUR GLISSER LE FICHIER COMPRESSE DU PLUGIN

3.2 Installation du Framework FLASK et Intégration de l’approche d’évaluation automatique dans le serveur PythonAnywhere :

L’installation de Flask se fait à l’aide de la commande `pip`¹⁵, qui simplifie beaucoup de choses. En plus de cela il faut bien évidemment avoir Python (Flask supporte les dernières versions de Python). Les étapes de l’installation du Flask sont comme suit :

- ✓ Création d’un compte PythonAnywhere.
- ✓ Se rendre dans l’onglet « Web » pour créer notre application FLASK où nous allons intégrer l’approche sémantique.
- ✓ Choisir la version Python et le nom de l’application FLASK
- ✓ Se rendre dans l’onglet « Consoles » pour créer notre environnement virtuel.
- ✓ Consulter la console Bash, en tapant la commande suivante : **`mkvirtualenv flaskapp --python=/usr/bin/python3.6`**
- ✓ Installer notre application Flask à l’aide de la commande : **`pip install flask`**

¹⁵ pip est un gestionnaire de paquets utilisé pour installer et gérer des paquets écrits en Python

Maintenant l'application FLASK est installé, ensuite nous devons intégrer l'approche d'évaluation automatique et ses outils. Du coup nous nous rendons dans l'onglet « Files » qui veut dire « fichiers » pour télécharger tous les fichiers de cette approche et installer toutes les bibliothèques nécessaires dans notre console bash comme nltk et numpy. Enfin, nous nous rendons dans notre fichier mainapp.py afin d'apporter tous les algorithmes nécessaires pour cette approche. Voici quelques figures (28, 29, 30, 31 et 32) qui montrent un peu de tout de ce qu'on vient d'expliquer.

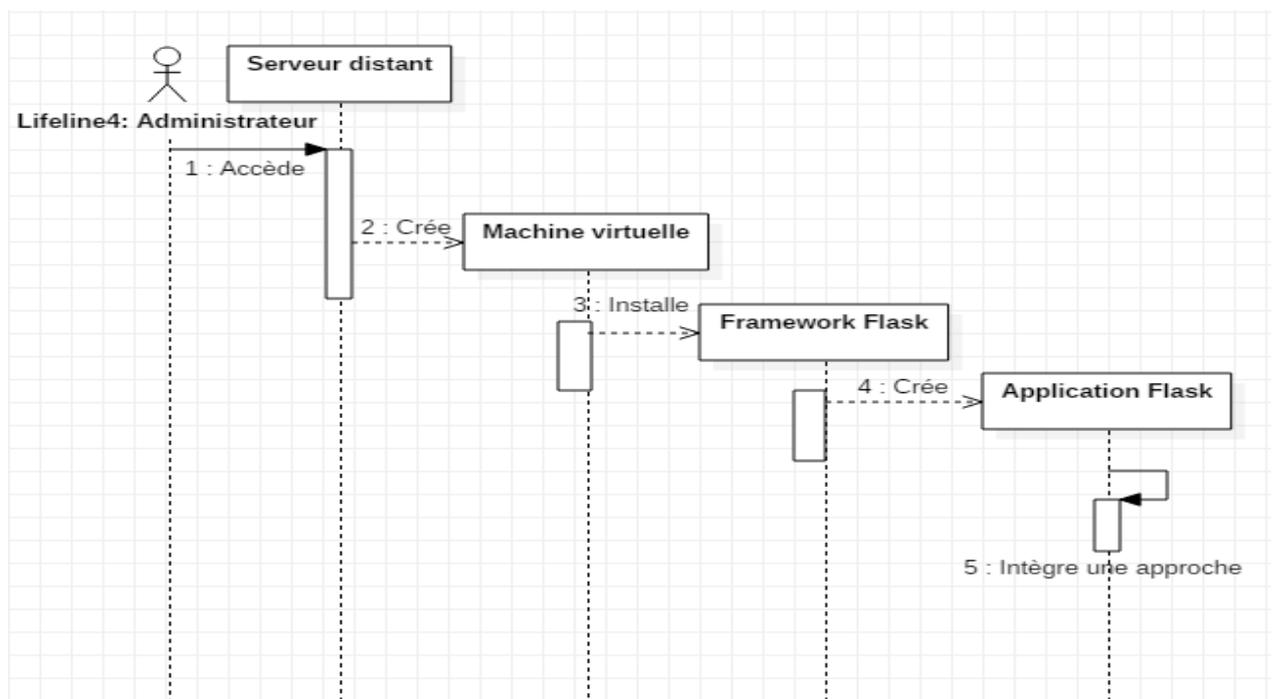


FIGURE 28 DIAGRAMME DE SEQUENCE DE SCENARIO « SCENARIO : INSTALLATION DU FLASK ET INTEGRATION DE L'APPROCHE D'EVALUATION AUTOMATIQUE »

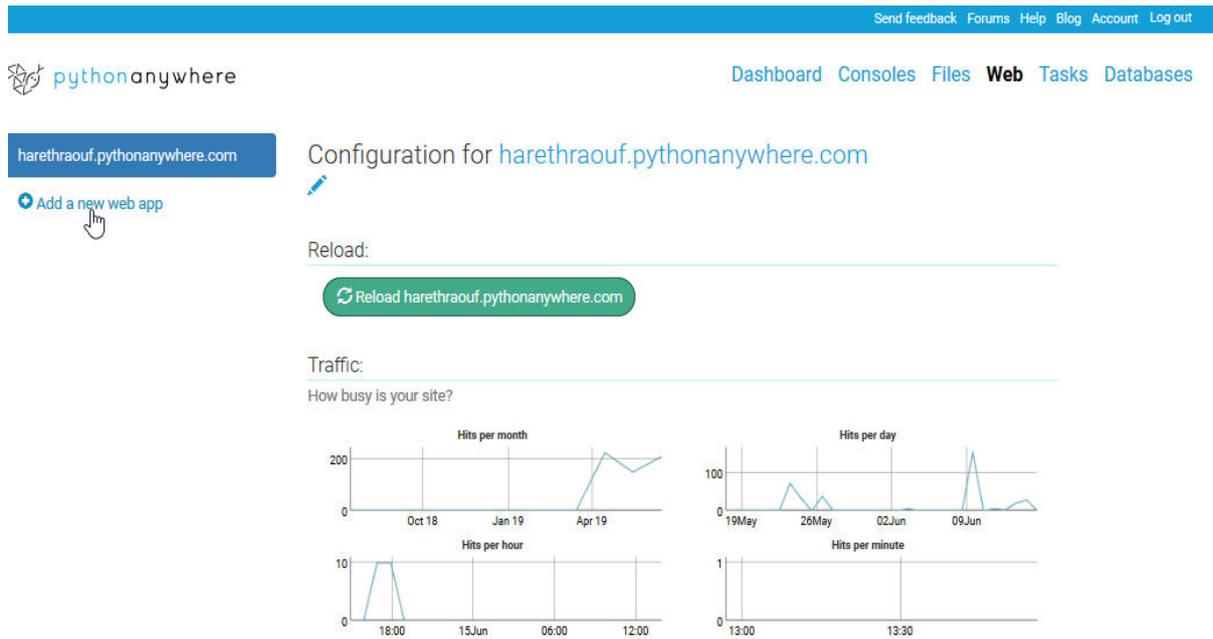


FIGURE 29 AJOUTER UNE NOUVELLE APPLICATION WEB DANS L'ONGLET « WEB »

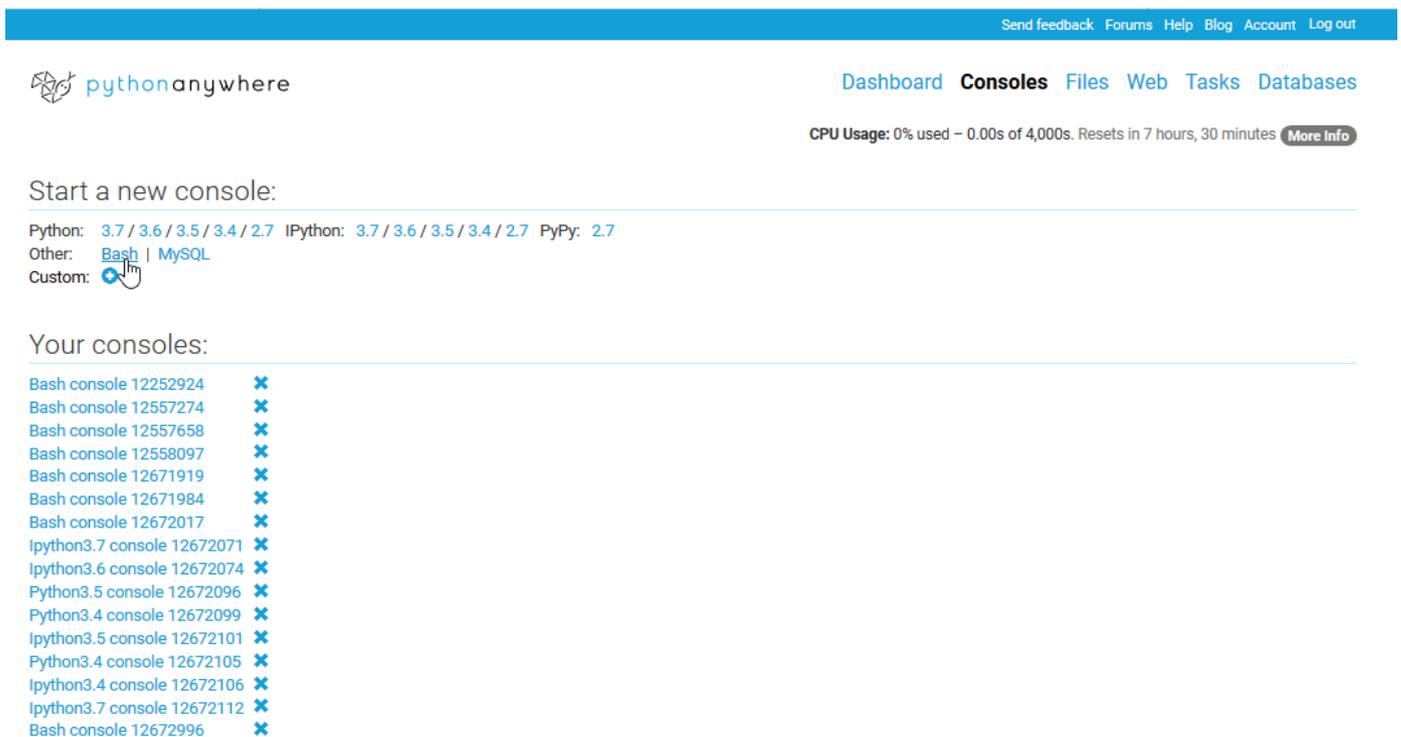


FIGURE 30 SE RENDRE DANS NOTRE CONSOLE BASH DANS L'ONGLET « CONSOLES »

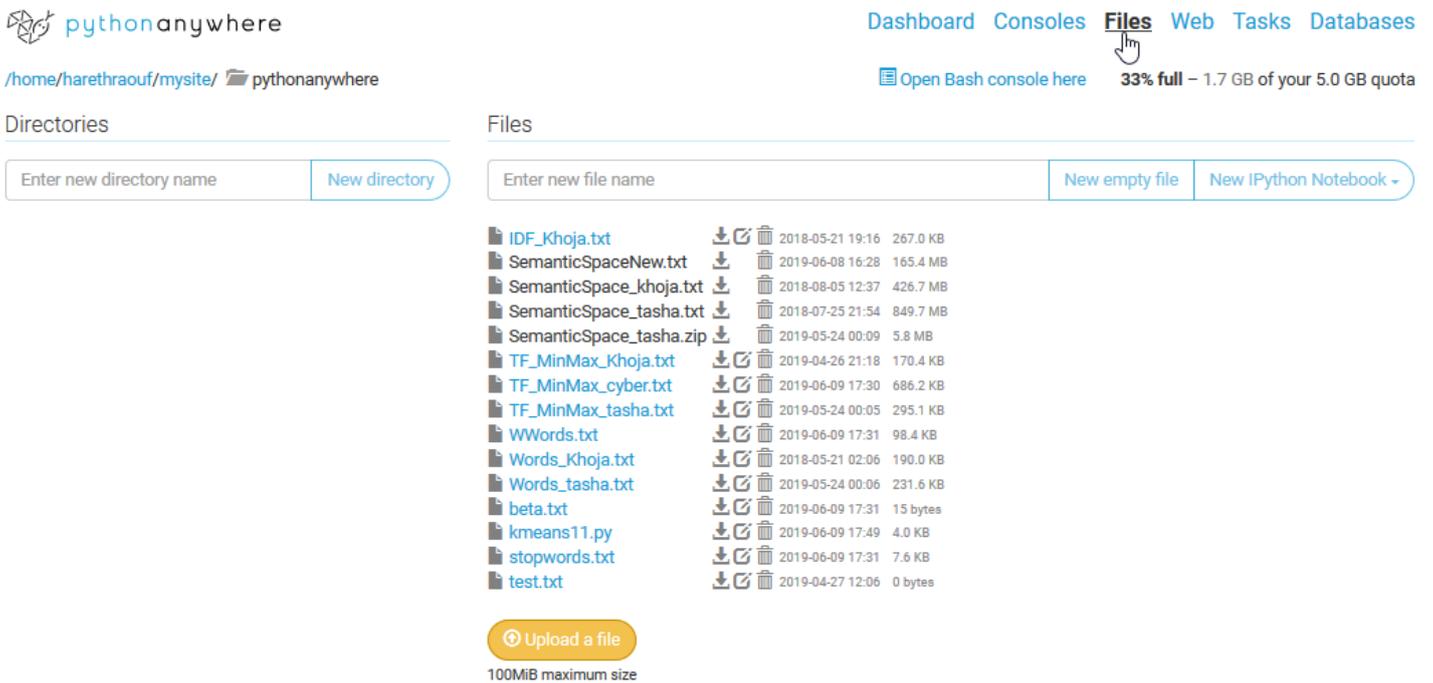


FIGURE 31 LES FICHIERS DE NOTRE APPLICATION FLASK DANS L'ONGLET « FILES »

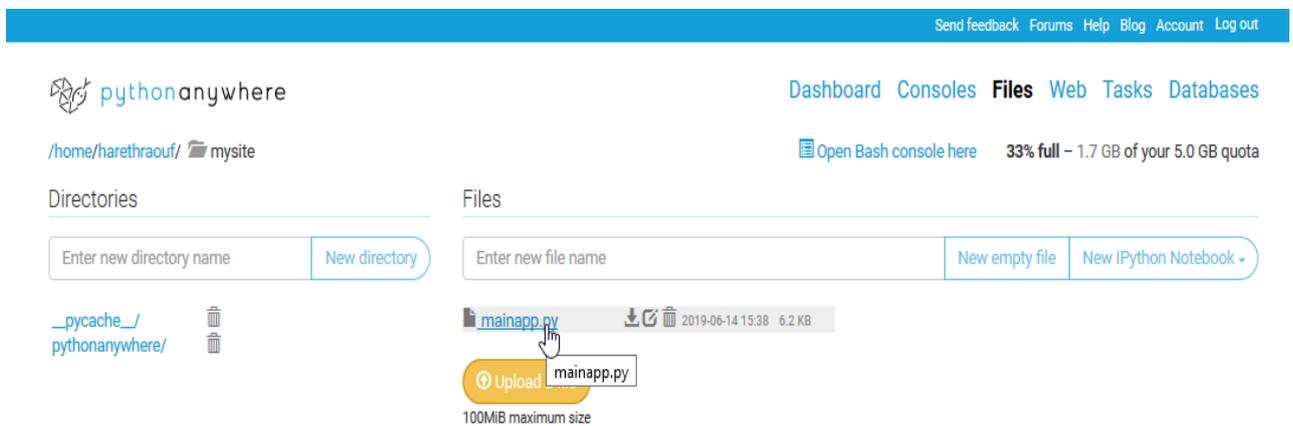


FIGURE 32 LE FICHER.PY PRINCIPAL DE FLASK POUR INTEGRER LES ALGORITHMES DE L'APPROCHE SEMANTIQUE

- **Echange de données entre le Plugin Arabic ShortAnswer et le serveur distant PythonAnywhere :**

En effet, cette partie focalise sur l'échange des requêtes durant le test, la figure 33 détaille ce scénario « Faire le test ». Flask, à sa manière, manipule les requêtes et les réponses HTTP. Une requête HTTP¹⁶ est envoyée par le navigateur au serveur HTTP, ou ce que sont les types GET et POST¹⁷, ou qu'une réponse possède toujours un code d'erreur. Dans notre cas, nous avons besoin d'une requête HTTP de type **GET** pour récupérer la réponse modèle de l'enseignant et celle de l'étudiant. De plus, les autres méthodes HTTP sont disponibles grâce à d'autres outils comme par exemple via **la commande cUrl ou JavaScript**. Dans le fichier **question.php** de notre Plugin, nous avons utilisé des lignes de cUrl PHP pour envoyer un couple de réponse et recevoir le score de la similarité entre ces réponses qui est la note finale.

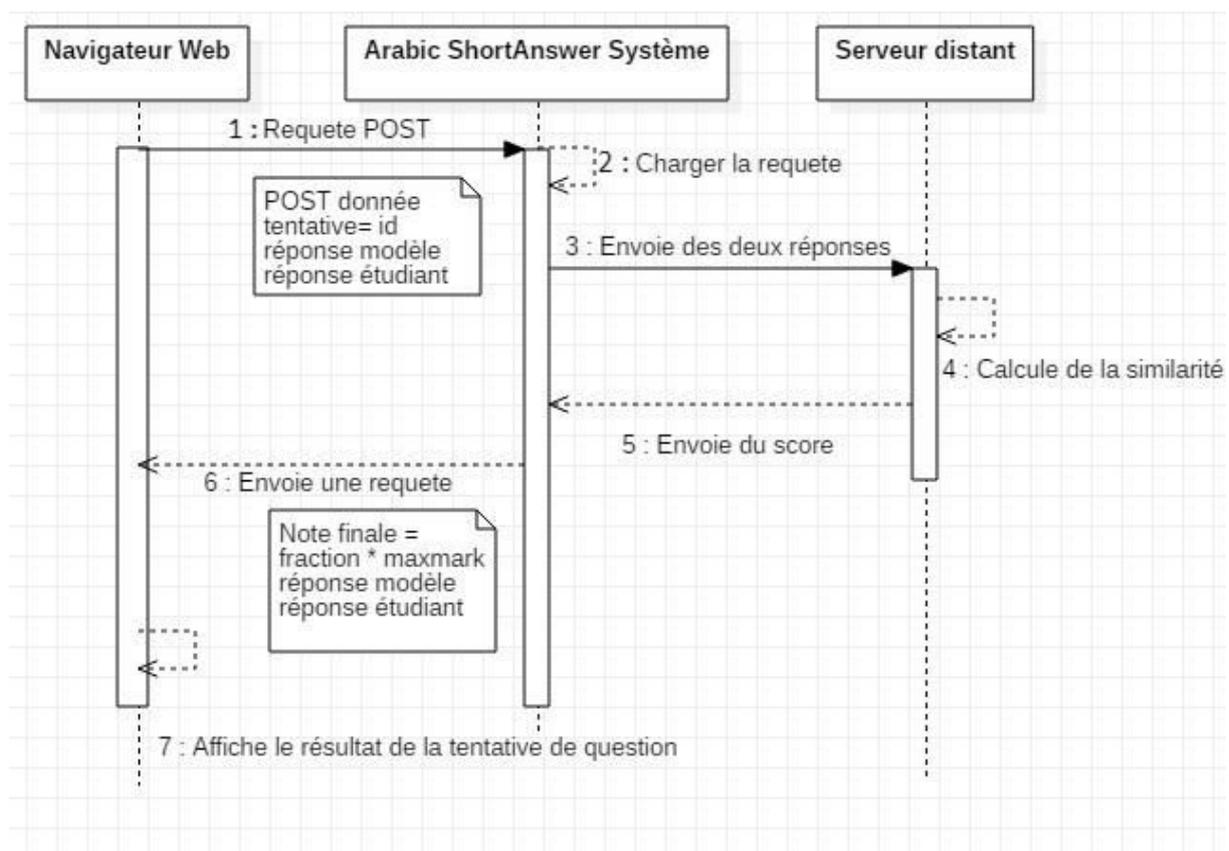


FIGURE 33 DIAGRAMME DE SEQUENCE « SCENARIO : FAIRE LE TEST »

¹⁶ http protocole de communication client-serveur

¹⁷ GET et POST sont des méthodes d'accès définies dans le protocole http.

Voici la figure 34 qui montre les lignes de code de cUrl PHP et la figure 35 qui montre les lignes de code de Flask notant que la variable **\$student** stocke la réponse de l'étudiant tandis que la variable **\$modelresponse** stocke la réponse modèle de l'enseignant.

```

2 $modelresponse = str_replace(array("\r", "\n"), " ", self::safe_normalize($answer));
3
4 $student = str_replace(array("\r", "\n"), " ", self::safe_normalize($response['answer']));
5
6 $curlvar= curl_init(); //Initialise the cURL var
7
8 curl_setopt($curlvar,CURLOPT_RETURNTRANSFER, 1); //Get the response from cURL
9 curl_setopt($curlvar,CURLOPT_URL,'http://harethraouf.pythonanywhere.com/$modelresponse='.self::checkInput($donnee)
10 .'$student='.self::checkInput($data)); //Set the Url
11
12 $content= (string) self::checkInput(curl_exec($curlvar));
13
14 curl_close($curlvar);
15
16 $fraction = (float)$content;
17
18 return array($fraction, question_state::graded_state_for_fraction($fraction));

```

FIGURE 34 LES LIGNES DE CODE EN CURL PHP ASSURANT L'ENVOIE ET LA RECUPERATION DES DONNEES

```

167
168 @app.route('/',methods=['GET','POST'])
169 def home():
170     if request.method == 'GET':
171         x1 = request.args.get('modelresponse')
172         x2= request.args.get('student')
173
174

```

FIGURE 35 LES LIGNES DE CODE EN PYTHON FLASK ASSURANT LA RECUPERATION ET L'ENVOIE DES DONNEES

3.3 Limites du serveur PythonAnywhere :

Les limites de ce serveur se résume sur le fait que ce dernier offre un espace disque de 500 Méga octets, si l'espace sémantique dépasse la taille du stockage de mémoire du serveur, nous devons passer par le paiement en ligne. Aussi, le seul langage de programmation autorisé sur PythonAnywhere est bien que Python. Ce serveur n'accepte aucun autre langage ce qui fait que l'approche sémantique doit être développé en Python seulement.

En réalité, ceci n'est pas une limite en elle-même car initialement les approches ont été développées en python autrement il faudrait investiguer pour un autre cloud qui puisse s'adapter avec les codes sources.

Mais plusieurs raisons, entre autres, peuvent justifier la croissance rapide de Python ces dernières années et le choix pris lors du développement des codes sources relatifs à l'évaluation automatique. À travers un sondage à choix multiples, nous voulons montrer ce qui permet au langage de gagner rapidement du terrain. Parmi ces raisons, nous pouvons citer les suivantes :

- **La facilité d'apprentissage ;**
- **Idéal pour l'enseignement de la programmation.** La croissance de Python ces dernières années est due à son adoption comme une norme pour l'enseignement de la programmation. De nombreux pédagogues apprécient le langage pour sa syntaxe permettant une initiation aisée aux concepts de base de la programmation ;
- **Python est bien comme langage de script ;**
- **Une bonne documentation ;**
- **Une pléthore de bibliothèques** pour de nombreux besoins ;
- **La montée de la data science et du machine learning ;**
- **Une bonne communauté** derrière Python ;
- **La lisibilité du code Python.** Par sa syntaxe, il est plus difficile (mais pas impossible) d'écrire du code illisible en Python. Toutefois, un tel code Python pourrait être plus lisible et propre que son équivalent dans d'autres langages ;
- **La rapidité de développement.** Un programme Python de 50 lignes peut représenter dans d'autres langages, des programmes de plusieurs centaines de lignes. Ce qui fait qu'en fin de compte, même avec un programmeur Python pas assez rapide, nous pouvons gagner beaucoup de temps au niveau du développement ;
- **Python tourne sur presque toutes les plateformes ;**
- **Python est ouvert aux autres langages et technologies.** Python a choisi de collaborer avec les autres langages et technologies comme PHP dans notre cas. Nous pouvons intégrer l'interpréteur Python dans son propre programme, et ajouter le langage comme système de Scripting (c'est le cas de Blender) ou plugin (comme avec Sublime Text). Nous pouvons appeler du code Python depuis d'autres langages et vice-versa.
-

Conclusion :

Le Plugin « **Arabic ShortAnswer** » représente un moule qui peut implémenter n'importe quelle approche dans la plateforme Moodle. Dans la continuité du travail qui a été fait l'année passée, dans le cadre de développement d'une approche sémantique et syntaxique pour l'évaluation automatique des réponses courtes en langue arabe, nous avons réussi à développer la méthodologie qui permet d'intégrer une approche déjà développée, dans un Plugin Moodle ce qui permet de passer à l'échelle dans un environnement d'apprentissage en ligne de manière naturelle.

Chapitre V : Résultats expérimentaux et Évaluation

Notre Plugin représente un moule moyennant certaines entrées et sorties où nous pouvons lui assigner n'importe quelle approche.

Dans ce chapitre, nous présentons des détails supplémentaires concernant la mise en œuvre de notre plugin y compris les détails de l'ensemble de données (dataset utilisés), tous les paramètres d'évaluation automatique que nous avons utilisés pour évaluer les résultats. Enfin, une analyse approfondie des résultats d'études qualitative et quantitative est fournie.

1- Ensemble de données (Data set) :

Un ensemble de données est un ensemble de données collectées dans un but spécifique. Les données peuvent être collectées de nombreuses manières - par exemple, dans le cadre de la prestation de services, d'enquêtes ponctuelles, d'entretiens, d'observations, etc.

Pour évaluer l'approche à utiliser dans notre plugin, 3 datasets sont utilisés (Le Cairo University Dataset [9], STS 250 SemEval 2017 [27] et une partie du dataset de l'examen du Cybercriminalité effectué cette année à l'université de Saad Dahlab.

1.1 Le Cairo University Data set [9]:

- Les questions présentées dans cet ensemble de données couvrent un chapitre du programme officiel égyptien de sciences de l'environnement (ES), qui représente 25% de l'ensemble du programme.
- L'ensemble de données contient 61 questions, 10 réponses pour chacune, avec un total de 610 réponses.
- L'ensemble de données contient un ensemble de réponses et de notes des élèves, qui ont été notées par deux annotateurs qui ont attribué des notes comprises entre 0 et 5.
- L'ensemble de données prend en charge 4 types de questions à réponse courte : "Définir le terme scientifique", "Expliquer", "Quelles sont les conséquences de" et "Pourquoi".
- La corrélation de Pearson (r) l'erreur quadratique moyenne (RMSE) est de 0,86 et 0,69 respectivement entre les deux annotateurs. Dans le tableau 11, nous présentons un exemple de ce Data set.

Question Subject	Question Type	Réponse modèle	Réponse étudiant	Note Manuelle
البيئة الاجتماعية	عرف المصطلحات العلمية	البيئة التي يشترك فيها الإنسان مع أقرانه من البشر	البيئة التي تصف أوجه التعاون والتفاعل بين الجميع من بنى البشر	2.5
لقدرة الجسمانية و الفسيولوجية للحيوانات البحرية التي تعيش في الأعماق	إشرح	تتحمل الضغط الزائد للماء، تتحمل البرودة الشديدة في الأعماق، تستطيع العيش في الظلام الدامس	تساعد القدرات الجسمانية على تكيف الحيوانات في الأعماق التي لا يصل إليها الضوء	2
النظام البيئي المعقد أكثر تماسكا و استقرارا من النظام البيئي البسيط	علل	لأن هذا التعقيد هو أحد العوامل الأساسية في سلامة كل نظام بيئي إذ أنه يحد من أثر التغيرات الإيكولوجية	لأن النظام البيئي المعقد يصعب تدميره لما يوجد لديه من بدائل كثيرة أما النظام البيئي البسيط يسهل تدميره	3

TABLEAU 11 : ECHANTILLON DATASET CAIRO UNIVERSITY [28]

1.2 STS 250 SemEval 2017 [27]

Semantic Textual Similarity (STS) ou La similarité textuelle sémantique mesure le degré d'équivalence sémantique entre deux phrases est une compétition qui se déroule chaque année depuis 2012.

Sur deux phrases, les systèmes participants sont invités à renvoyer un score de similarité continu sur une échelle de 0 à 5, 0 indiquant que la sémantique des phrases est complètement indépendante et 5 signifiant équivalence sémantique. La performance est évaluée en calculant la corrélation de Pearson entre les scores de similarité sémantique assignés par machine et les jugements humains. Dans le tableau 12, nous présentons un exemple de ce Data set.

1 ère phrase	2eme phrase	Note Manuelle
يقف ثلاثة أشخاص على منصة سوداء صلبة بجانب كرسي أسود وقيثارة	تقف امرأة تعزف القيثارة على كرسي أمام البيت الأبيض.	1.8
يمشي الرجل على الشارع مرتدياً سترة بيضاء.	يوجد امرأة في الشارع ترتدي الأخضر.	0.6
رجل يستلقي على الأرض	رجل يستلقي على الأرض في الخارج	4.2

TABLEAU 12 : ECHANTILLON DATASET STS250 SEMEVAL [22]

1.3 Dataset Cyber Criminality Arabic Short Answer Data Set

Partie du Dataset construite à partir de l'examen¹⁸ du cours de cybercriminalité. Cette partie contient 18 réponses modèles ou chaque réponse modèle est comparée à 10 réponses étudiants soit un total de 180 réponses. Dans le tableau 13, nous présentons un exemple de ce Data set. Ce dataset est le résultat d'un travail de PFE en cours non encore soutenu.

Question	Réponse modèle	Réponse étudiant	Note Manuelle
عرف مصطلح الجريمة المنظمة على الانترنت	هي عنة منظم تقوم به جماعات ترتكب أفعالاً تخترق بها القانون للحصول على مكاسب مالية، بطرق وأساليب غير مشروعة تنفذ بعد تدبير وتنظيم	الجريمة المنظمة هي جريمة عنيفة تكون من طرف مجموعة للحصول على المال و تمتاز بالتنظيم والحصول المال	3.5
بماذا تقاس قوة سلسلة الأمن للنظام المعلوماتي؟	تقاس بقوة أضعف حلقة فيها وهي الإنسان (العنصر البشري)	تقاس قوة سلسلة الأمن للنظام المعلوماتي حلقة به بأضعف	2.5
ماهي ابرز طرق انتشار الفيروسات	البريد الإلكتروني و البرامج التي يتم تحميلها من الانترنت خاصة من المواقع المشبوهة	زيارة المواقع المشبوهة جميع البرامج توجد فيها crack المقرصنة و التي تحتوي على فيروسات الإختراق و تحميل برنامج الإنزال	4.5

TABLEAU 13 : ECHANTILLON DATASET CYBER CRIMINALITY ARABIC SHORT ANSWER GRADING

¹⁸ [Lien de l'examen du cours de cybercriminalité](#)

2- Métriques d'évaluation :

Après avoir suivi l'ingénierie des fonctionnalités, la sélection et, bien sûr, l'implémentation de notre modèle et obtenu des résultats sous forme de scores, l'étape suivante consiste à déterminer l'efficacité du modèle basé sur une mesure utilisant des jeux de données(datasets).

Différentes mesures de performance sont utilisées pour évaluer nos différents algorithmes, espaces sémantiques et méthodes de similarité sémantiques entre deux réponses. Dans notre cas, nous allons nous concentrer sur ceux utilisés pour déterminer la liaison qui existe entre des variables. Ces derniers sont les scores manuels et automatiques.

2.1 Coefficient de Pearson CP [29]

Il est très important de comprendre la relation entre les variables pour tirer la bonne conclusion d'une analyse statistique. La relation entre les variables détermine la manière dont les bonnes conclusions sont atteintes. En statistiques, étudier la corrélation entre deux ou plusieurs variables statistiques numériques, revient à étudier l'intensité de la liaison ("proportionnalité") qui peut exister entre ces variables.

La **corrélation statistique** est mesurée par le coefficient de corrélation (r). Sa valeur numérique varie de 1,0 à -1,0. Cela donne une indication de la force de la relation.

En général, $r > 0$ indique une relation positive, $r < 0$ indique une relation négative tandis que $r = 0$ indique l'absence de relation (ou que les variables sont indépendantes et non liées). Le coefficient de corrélation est multiplié par 100 pour exprimer un pourcentage de corrélation. Dans le tableau 14, nous présentons les significations de ces valeurs de corrélation.

Valeur de r	Force de la relation
-1,0 À -0,5 ou 1,0 à 0,5	Fort
-0,5 À -0,3 ou 0,3 à 0,5	Modéré
-0,3 À -0,1 ou 0,1 à 0,3	Faible
-0,1 À 0,1	Absente ou très faible

TABLEAU 14 :SIGNIFICATION DES VALEURS DE CORRELATION DE PEARSON [23]

Dans notre cas la mesure du coefficient de corrélation de Pearson a été utilisée pour spécifier la corrélation entre le score automatique et les notes humaines moyennes.

2.2 Erreur quadratique RMSE (Root Mean Squared Error (RMSE) [30]

C'est la deuxième métrique d'évaluation utilisée et elle permet de quantifier une mesure synthétique de l'erreur globale commise. L'algorithme de l'erreur quadratique moyenne permet de calculer cette valeur qui va nous permettre de comparer les valeurs calculées automatiquement (ici les scores ou notes automatiques) et les valeurs attribuées Manuellement (scores manuels attribués par l'expert humain) et plus la valeur tend vers le zéro plus les résultats sont meilleurs.

3- Résultats et discussion :

Cette partie est consacrée à l'évaluation et la comparaison des résultats obtenus par rapport à nos deux métriques entre les notes obtenues manuellement (par l'expert humain) et les notes générées automatiquement afin de choisir la meilleure approche à implémenter dans notre plugin. On va aussi comparer les résultats par rapport à ceux obtenus par Stem léger appliqué sur les couples de réponses de chaque dataset donc soit Isri [31], Tashaphyne [32] ou la combinaison de ces deux derniers.

3.1 Les résultats du modèle CM :

Dans les tables 15 et 16 successivement, nous retrouvons les résultats de l'application du calcul matriciel avec l'espace sémantique Khaleej et CNN successivement sur les trois Datasets, les résultats étaient meilleurs pour l'espace sémantique Khaleej avec les dataset Cairo University et STS250. Pour le dataset cybercriminalité l'espace sémantique CNN a donné de meilleur résultats. Les résultats obtenus par les deux espaces sémantiques restent très proches et comparables.

CM +KHALEEJ					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	74.73	76.73	76.11	77.06
	RMSE	1.45	1.14	1.14	1.26
STS250	CP	61.32	70.36	71.22	68.74
	RMSE	1.41	1.19	1.19	1.22
Cybercriminalité	CP	56.41	60.81	59.25	55.47
	RMSE	1.42	1.31	1.29	1.40

TABLEAU 15 :RESULTATS DU MODELE CM AVEC L'ESPACE SEMANTIQUE KHALEEJ

CM + CNN					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	74.83	76.67	76.26	77.02
	RMSE	1.43	1.14	1.13	1.26
STS250	CP	60.95	70.28	71.21	68.70
	RMSE	1.41	1.19	1.19	1.22
Cybercriminalité	CP	57.14	62.39	58.57	55.61
	RMSE	1.37	1.28	1.32	1.39

TABLEAU 16 :RESULTATS DU MODELE CM AVEC L'ESPACE SEMANTIQUE CNN

3.2 Les résultats de la combinaison (CM,Dice) :

Dans les tables 17 et 18 successivement, nous retrouvons les résultats de l'application de la combinaison du calcul matriciel et l'indice de Dice avec l'espace sémantique Khaleej et CNN successivement sur les trois Datasets. Les résultats étaient meilleurs pour l'espace sémantique CNN avec les dataset Cairo University et STS250. Pour le dataset cybercriminalité l'espace sémantique Khaleej a donné de meilleur résultats. Les résultats obtenus par les deux espaces sémantiques restent très proches et comparables.

La combinaison du modèle CM avec l'indice de Dice a nettement amélioré les résultats sur les trois datasets en notant que le coefficient de Pearson a atteint les 80% de corrélation avec le dataset Cairo University.

CM+DICE+KHALEEJ					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	80.83	79.50	79.74	79.87
	RMSE	1.22	1.16	1.12	1.24
STS250	CP	66.77	71.33	71.08	69.42
	RMSE	1.17	1.14	1.12	1.13
Cybercriminalité	CP	66.18	67.07	69.49	62.05
	RMSE	1.16	1.15	1.12	1.22

TABLEAU 17 :RESULTATS DE LA COMBINAISON (CM,DICE) AVEC L'ESPACE SEMANTIQUE KHALEEJ

CM+DICE+CNN					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	80.98	79.36	80.14	80.00
	RMSE	1.19	1.17	1.08	1.25
STS250	CP	66.57	71.45	70.59	70.14
	RMSE	1.17	1.14	1.12	1.12
Cybercriminalité	CP	67.55	68.78	63.59	66.12
	RMSE	1.16	1.13	1.24	1.19

TABLEAU 18 : RESULTATS DE LA COMBINAISON (CM,DICE) AVEC L'ESPACE SEMANTIQUE CNN

3.3 Les résultats du modèle Somme-vectorielle :

Dans les tables 19 et 20 successivement, nous retrouvons les résultats de l'application somme vectorielle avec l'espace sémantique Khaleej et CNN successivement sur les trois Datasets. Les résultats étaient meilleurs pour l'espace sémantique Khaleej avec les dataset Cairo University et STS250. Pour le dataset cybercriminalité l'espace sémantique CNN a donné de meilleur résultats. Les résultats obtenus par les deux espaces sémantiques restent très proches et comparable.

SV+KHALEEJ					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	75.91	70.21	76.57	73.95
	RMSE	1.15	1.23	1.06	1.15
STS250	CP	48.45	57.79	64.61	62.53
	RMSE	1.50	1.41	1.29	1.29
Cybercriminalité	CP	59.57	57.79	56.84	57.73
	RMSE	1.40	1.41	1.33	1.40

TABLEAU 19 : RESULTATS DU MODELE SV AVEC L'ESPACE SEMANTIQUE KHALEEJ

SV+CNN					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	75.56	70.55	76.07	74.62
	RMSE	1.23	1.08	1.09	1.13
STS250	CP	46.74	57.74	63.94	61.86
	RMSE	1.48	1.44	1.33	1.34
Cybercriminalité	CP	61.30	59.38	61.30	61.57
	RMSE	1.43	1.28	1.29	1.39

TABLEAU 20 : RESULTATS DU MODELE CM AVEC L'ESPACE SEMANTIQUE CNN

3.4 Les résultats de la combinaison (Somme-vectorielle, Dice) :

Dans les tables 21 et 22 successivement, nous retrouvons les résultats de l'application de la combinaison du calcul matricielle et l'indice de Dice avec l'espace sémantique Khaleej et CNN successivement sur les trois Datasets. Les résultats étaient meilleurs pour l'espace sémantique Khaleej avec les dataset Cairo University et STS250. Pour les trois dataset.

La combinaison du modèle SV avec l'indice de Dice a nettement amélioré les résultats sur les trois datasets en notant que le [9] a atteint les 81.45% de corrélation avec le dataset Cairo University.

SV+Dice+KHALEEJ					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	81.45	78.68	80.17	80.11
	RMSE	1.15	1.09	1.14	1.05
STS250	CP	62.28	66.87	68.14	67.55
	RMSE	1.27	1.20	1.17	1.19
Cybercriminalité	CP	67.45	56.57	64.70	64.61
	RMSE	1.14	1.32	1.20	1.18

TABLEAU 21 : RESULTATS DE LA COMBINAISON (SV,DICE) AVEC L'ESPACE SEMANTIQUE KHALEEJ

SV+Dice+CNN					
Dataset	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
Cairo University	CP	81.04	79.00	80.96	80.32
	RMSE	1.11	1.08	1.07	1.02
STS250	CP	66.60	61.69	69.37	66.53
	RMSE	1.21	1.27	1.16	1.20
Cybercriminalité	CP	68.51	68.51	67.82	66.97
	RMSE	1.13	1.24	1.16	1.26

TABLEAU 22: RESULTATS DE LA COMBINAISON (SV, DICE) AVEC L'ESPACE SEMANTIQUE CNN

Discussion des résultats obtenus :

Dataset Cairo University : pour ce dataset le meilleur résultat est obtenu après la combinaison du modèle somme-vectorielle avec l'indice de Dice et le stem Tashaphyne avec un coefficient de Pearson de 81.45 sur l'espace sémantique Khaleej. Ce résultat reste quand-même très proche du meilleur résultat du modèle CM calculé à 80.98 avec le stem Tashaphyne et l'espace sémantique CNN.

Dataset STS250 : pour ce dataset le meilleur résultat est obtenu après la combinaison du modèle CM avec l'indice de Dice et le Stem Isri avec un coefficient de Pearson de 71.45 sur l'espace sémantique CNN sachant que pour le même dataset le meilleur résultat obtenu avec le modèle somme-vectorielle est de 69.37 sur l'espace sémantique CNN et la combinaison de Stem (Isri+Tashaphyne).

Dataset Cyber Criminality Arabic Short Answer Grading Data Set: pour ce dataset le meilleur résultat est obtenu après la combinaison du modèle CM avec l'indice de Dice et la combinaison de stem (Isri+Tashaphyne) avec un coefficient de Pearson calculé à 69.49 sur l'espace sémantique Khaleej ce qui est très proche avec le meilleur résultat obtenu avec la combinaison du modèle somme-vectorielle avec Dice et le stem Tashaphyne sur l'espace CNN calculé à 68.51.

3.5 Comparaison par rapport aux résultats obtenus dans les travaux PFE déjà fait avec le Stem Lourd :

Comme mentionné auparavant, le serveur utilisé en relation avec notre plugin ne tolère pas le langage java d'où l'orientation vers un Stem léger développé en python afin de l'utiliser avec les réponses modèles et réponses étudiants.

Dans la table 23, nous retrouvons la comparaison des résultats du modèle CM avec les espaces sémantiques Khaleej et CNN sur les deux datasets Cairo University et STS250 en introduisant un Stem lourd et un Stem léger, et les résultats ont été meilleures avec les Stem léger sur les deux datasets.

Dataset	Espace	Stem Lourd	Stem Leger
Cairo University	Khaleej	75.92 // 1.17	77.07 // 1.26
	CNN	76.50 // 1.15	77.02 // 1.26
STS250	Khaleej	70.87 // 1.20	71.22 // 1.19
	CNN	70.83 // 1.19	71.21 // 1.19

TABLEAU 23 : COMPARAISON DES RESULTATS DE L'ALGORITHME CM ENTRE STEM LOURD ET LEGER

Dans la table 24, nous retrouvons la comparaison des résultats du modèle SV avec les espaces sémantiques Khaleej et CNN sur les deux datasets Cairo University et STS250 en introduisant un Stem lourd et un Stem léger, et les résultats ont été meilleures avec les Stem lourd sur les deux datasets.

Dataset	Espace	Stem Lourd	Stem Leger
Cairo University	Khaleej	77.50 // 1.05	76.57 // 1.06
	CNN	78.33 // 1.04	76.08 // 1.09
STS250	Khaleej	66.43 // 1.23	64.61 // 1.29
	CNN	66.88 // 1.25	63.94 // 1.33

TABLEAU 24 : COMPARAISON DES RESULTATS DE L'ALGORITHME SV ENTRE STEM LOURD ET LEGER

Dans la table 25, nous retrouvons la comparaison des meilleurs résultats de la combinaison du modèle CM et l'indice de Dice sur les deux datasets Cairo University et STS250 en introduisant un Stem lourd et un Stem léger. Pour le dataset Cairo University le meilleur résultat a été obtenu avec le Stem léger et pour le dataset STS250 le meilleur résultat a été obtenu avec le Stem lourd.

Dataset	Stem Lourd	Stem Leger
Cairo University	80.78 // 1.09	80.98 // 1.08
STS250	71.55 // 1.12	71.45 // 1.12

TABLEAU 25 : COMPARAISON DES RESULTATS DE LA COMBINAISON (CM, DICE) ENTRE STEM LOURD ET LEGER

Dans la table 26, nous retrouvons la comparaison des meilleurs résultats de la combinaison du modèle SV et l'indice de Dice sur les deux datasets Cairo University et

STS250 en introduisant un Stem lourd et un Stem léger. Pour les deux Datasets les meilleurs résultats ont été obtenus avec le Stem lourd.

Dataset	Avec Stem Lourd	Avec Stem Leger
Cairo University	81.49 // 0.98	81.45 // 1.05
STS250	70.87 // 1.11	69.37 // 1.16

TABLEAU 26 : COMPARAISON DES RESULTATS DE LA COMBINAISON(SV,DICE) ENTRE STEM LOURD ET LEGER

Conclusion:

On a atteint une corrélation très importante de 81.45 et une erreur quadratique de 1.05. Ces résultats sont très proches à ceux obtenus par le stem lourd avec une corrélation de 81.49 et une erreur quadratique de 0.98.

3.6 Résultats obtenus avec un espace sémantique spécifique au cybercriminalité :

Pour le Tableau 27, nous avons utilisé un espace sémantique issu d'un corpus de cyber criminalité avec le dataset Cyber Criminality Arabic Short Answer Grading. Le dataset utilisé est une partie du dataset général construit à partir de l'examen de cybercriminalité. Cette partie contient 18 réponses modèles ou chaque réponse modèle est comparée à 10 réponses étudiants soit un total de 180 couples de réponses.

Les résultats obtenus sont meilleurs par rapport aux espaces sémantiques Khaleej et CNN et c'est la combinaison entre le modèle somme-vectorielle et l'indice de Dice avec le stem Isri qui donne le meilleur résultat avec un coefficient de Pearson calculé à 70.51.

Approche	Métrique	Tashaphyne	Isri	Isri+Tashaphyne	Tashaphyne+Isri
CM	CP	56.76	61.04	60.16	56.01
	RMSE	1.41	1.29	1.28	1.39
CM+Dice	CP	67.28	69.75	67.39	61.78
	RMSE	1.16	1.11	1.17	1.22
Sum-vec	CP	61.64	61.18	62.27	63.80
	RMSE	1.42	1.35	1.43	1.29
Sum-vec + Dice	CP	67.42	70.51	68.76	68.77
	RMSE	1.19	1.16	1.20	1.24

TABLEAU 27 : RESULTAT DE TOUTES LES APPROCHES AVEC L'ESPACE SEMANTIQUE CYBERCRIMINALITE

Pour conclure, les résultats obtenus sont des résultats très correctes pour le nouvel espace sémantique de la cybercriminalité qui a donné de meilleurs résultats avec le dataset Cyber Criminality Arabic Short Answer Grading.

Bien que nous soyons conscients que ces résultats peuvent être améliorés en améliorant la qualité de l'espace sémantique, nous les jugeons suffisants étant donné que l'objectif dans ce travail est de tester la fonctionnalité et l'intégration du plugin dans la plateforme.

4- L'évaluation qualitative :

Afin d'effectuer une simulation réelle du fonctionnement du plugin dans un test en ligne réel, un appel à participation avait été lancé auprès des étudiants du Master1 IL afin de participer à des tests¹⁹ créés en ligne sur la plateforme Moodle de Bouira.

Le test est basé sur le cours²⁰ donné dans le module de cybercriminalité enseigné en Master 1 pour les spécialités IL, SIR et SSI du département d'Informatique de l'université de Blida. Les mêmes étudiants ont effectué l'examen de manière traditionnelle en salle et en manuel ont répondu au test en ligne.

L'objectif de cette évaluation qualitative est de tester d'un côté l'intégration du plugin et de voir son comportement à une échelle réelle et d'un autre côté, collecter l'appréciation des étudiants concernant l'environnement du test en ligne et l'apport apporté par rapport à un même test réalisé en traditionnel.

A la fin du test, il y a été demandé aux étudiants de remplir un formulaire d'évaluation (voir annexe)²¹ [Annexe](#).

4.1 Les Tests :

Afin d'assurer la bonne intégration du Plugin deux tests ont été créés :

Le premier test est composé d'un seul type de question qui est le nouveau type développé (Arabic Short Answer) et il contient 17 questions sur le module de cybercriminalité.

Le deuxième test est composé de 3 différents types de questions, avec deux questions de type (Arabic Short Answer), une question à choix multiple, et une dernière question de type (remplir le vide). L'objectif ici est d'évaluer l'intégration du plugin avec d'autres plugins dans un même test. L'approche utilisée pour ces tests c'est le modèle SV avec stem Isri et l'espace sémantique généré à partir du corpus de cybercriminalité.

¹⁹ Test 1 : <http://elearninginfo.univ-bouira.dz/test/moodle/mod/quiz/view.php?id=11>

Test 2 : <http://elearninginfo.univ-bouira.dz/test/moodle/mod/quiz/view.php?id=12>

²⁰ <http://elearninginfo.univ-bouira.dz/test/moodle/course/view.php?id=4>

²¹ Le formulaire d'évaluation : <http://elearninginfo.univ-bouira.dz/test/moodle/mod/data/view.php?id=14>

4.2 Evaluation :

Pour aider à l'évaluation de la participation, à la fin des tests les participants ont rempli chacun un formulaire qui permet de recueillir leurs appréciations sur le fonctionnement du plugin et leur motivation à effectuer des tests en ligne.

La consolidation des réponses aux questions et des remarques formulées par les étudiants participants aux tests nous a permis de faire les constatations suivantes : Les étudiants ont été très positifs à propos du test. Ils apprécient particulièrement le retour immédiat des notes. Ils estiment que la notation instantanée de chaque question aide à maintenir la motivation. Ils prennent ainsi, le temps de bien faire les choses avant de continuer et de valider le test. Ceci en temps normal d'utilisation du plugin dans l'apprentissage du module permet, de pouvoir suivre leur niveau de compréhension du cours.

Dans un contexte d'examen, les étudiants déclarent également qu'ils trouvent utile le retour d'information immédiat (appréciations et des réponses de l'enseignant). En effet, le feedback instantané évite de ne pas savoir comment ils évoluent, et ils apprécient de quitter la salle d'examen en sachant leur note.

D'un point de vue enseignant, sans surprise l'enseignant du module apprécie de n'avoir que peu ou pas de tests de corrections d'examens, en particulier quand la correction concerne un nombre important de copies.

Conclusion Générale

Face à l'absence des outils d'évaluation automatique dans les plateformes de télé-enseignement, nous avons élaboré au travers de ce travail, le Plugin Arabic Short Answer. Le PLUGIN (Arabic Short Answer Question Type Plugin) permet :

- d'intégrer l'évaluation automatique des réponses courtes dans la plateforme de télé-enseignement Moodle,
- d'offrir à l'enseignant la possibilité de construire son test de manière naturelle et uniforme en combinant différents types de questions déjà intégrées sur la plateforme (choix multiple, essais, remplir les vides, ...) avec le nouveau type d'évaluation proposée pour la langue arabe, familiariser les apprenants à un environnement de test et au feedback consolidé sur la même plateforme de cours et enfin évaluer les approches déjà développées par rapport à un passage à l'échelle.

Nous avons dans un premier temps présenté les outils d'évaluation automatique apportées jusqu'à aujourd'hui dans les plateformes de télé-enseignement et nous avons constaté que ces outils sont limités quand il s'agit de considérer les réponses courtes. Ensuite nous avons analysé la plateforme la plus populaire Moodle avec son système de quiz et ses approches d'évaluation automatique des réponses courtes afin d'intégrer un moule améliorant l'évaluation automatique des réponses courtes.

Dans le contexte de notre travail de l'évaluation automatique, le travail déjà réalisé concernant la similarité sémantique et syntaxique nous a permis d'exploiter les différentes approches sémantiques développées en langage Python afin d'intégrer une de ses dernières dans notre serveur distant PythonAnywhere où nous avons réussi à créer une application web basant sur le framework Flask ainsi de réaliser des échanges de données (réponses modèle, réponse étudiant et score) entre Flask et notre Plugin Arabic Short Answer. Ceci a permis le retour immédiat des notes.

L'utilisation des différents datasets et des différentes approches sémantiques d'évaluation automatique des réponses courtes nous a permis de réaliser des différents tests pour voir l'impact de notation automatique sur notre Plugin Arabic Short Answer.

Dans un dernier temps, nous avons effectué une simulation réelle du fonctionnement du Plugin dans un test en ligne réel avec des étudiants. Ceci nous a permis de suivre tout l'historique des étudiants durant le test, constater le bon fonctionnement du Plugin et suivre les fonctionnalités de la plateforme Moodle auprès du notre Plugin Arabic Short Answer.

En perspective, nous pensons étendre l'intégration du plugin dans les questions CLOZE de Moodle. Ce sont des questions imbriquées de plusieurs questions. Bien que la version actuelle de Moodle ne permette pas une telle intégration nous pensons qu'il est possible de modifier le cœur de la question pour pouvoir l'intégrer.

- [13] T. H. e. B. Prathibha, «*Overview of the Moodle question engine* »,» 2016. [En ligne].
] Available: https://docs.moodle.org/dev/Overview_of_the_Moodle_question_engine.
- [14] M. P.-C. J. S. e. T. R. T. Hunt, «*Question Engine 2:Overview* »,» 2015. [En ligne].
] Available: https://docs.moodle.org/dev/Question_Engine_2:Overview .
- [15] M. Abbas, «*Arabic Corpus*,» 2019. [En ligne]. Available:
] <https://sourceforge.net/projects/arabiccorpus/>.
- [16] M. Saad, «*OSAC: Open Source Arabic Corpora*,» 2010. [En ligne]. Available:
] <http://site.iugaza.edu.ps/msaad/osac-open-source-arabic-corpora/>.
- [17] J. B. Lovins, «*Development of a stemming algorithm* »,» vol. 11, n° %1June, 1968.
]
- [18] M. Saad, «*Arabic Computational Linguistics*,» 2019. [En ligne]. Available:
] <https://sourceforge.net/projects/ar-text-mining/files/Arabic-Corpora/>.
- [19] L. M. G. e. D. C. P. D. L. T. Rohde, «*An Improved Method for Deriving Word
Meaning from Lexical* »,» vol. 7, 2004.
]
- [20] C. T. e. C. Berrut, «*Pondération des données incertaines dans les systèmes de recherche
d ’ informations : une première approche expérimentale* », 2006.
]
- [21] J. Macqueen, «*Some methods for classification and analysis of multivariate
observations* »,» vol. 1, n° %1233, 1967.
]
- [22] J. Pratt, «*moodle-qtype_TEMPLATE — GitHub* »,» 2013. [En ligne]. Available:
] https://github.com/jamiepratt/moodle-qtype_TEMPLATE/blob/master/README.md.
- [23] V. Fremaux, «*Développement de Moodle. Méthodologie : conception d'un module
Moodle*,» 7 Avril 2007. [En ligne]. Available:
<https://moodle.org/mod/forum/discuss.php?d=68632>.
- [24] T. Christie, «*Flask API*,» MKDocs, 2017. [En ligne]. Available:
] <https://www.flaskapi.org/>.
- [25] C. C. J. P. M. K. S. M. e. T. H. A. Forth, «*Open University Moodle plugins* »,» 2019.
] [En ligne]. Available: <https://github.com/moodleou>.
- [26] «*Question types* »,» 2019. [En ligne]. Available:
] <https://moodle.org/plugins/browse.php?list=category&id=29>.
- [27] M. D. E. A. I. L.-G. e. L. S. D. Cer, «*SemEval-2017 Task 1: Semantic Textual
Similarity - Multilingual and Cross-lingual Focused Evaluation* ». Vancouver, Canada

-] August 2017.
- [28 W. H. G. e. A. A. Fahmy, «Arabic Short Answer Scoring with Effective Feedback for Students,» vol. 86, n° 12, 2014.
]
- [29 (. 2. 2. S. Kalla, «« Relationship Between Variables »,» Retrieved Jun 24, 2019..
]
- [30 W. Greene, «« Econométrie, Paris, Pearson Education »,» 5e éd, 2005.
]
- [31 K. E. R. a. C. Taghva, «Arabic Stemming without a root dictionary,» University of Nevada, Las Vegas, USA., 2005.
]
- [32 T. Zerrouk, «Tashaphyne, Arabic light stemmer,» 2012.
]

Annexe

Formulaire d'évaluation qualitative

Question 1 : Que pensez-vous du retour immédiat de la note ?

Question 2 : La notation instantanée vous donne plus de motivation à faire des test d'apprentissage en dehors des examens ?

Question 3 : Avez-vous apprécié le fait de retourner sur les réponses avant de valider le test?

Question 4 : Trouvez-vous utile le retour d'information immédiat (la bonne réponse, la remarque sur le niveau de vos réponses) ?

Question 5 : Vous avez passé le même test en salle. Le trouvez-vous meilleur en ligne ?

Question 6 : Remarques et Commentaires