

**UNIVERSITÉ SAAD DAHLAD DE BLIDA -1-**

**Faculté des Sciences**

Département d'Informatique

## **MÉMOIRE DE MAGISTER**

Spécialité : Informatique Répartie et Mobile (IRM)

# **EVALUATION AUTOMATIQUE DES REPONSES DANS LE CONTEXTE D'UN EXAMEN EN LIGNE**

Par

**Rachda DJELLATOU**

Devant le jury composé de :

H. ABED	Professeur, U.Blida-1-	Présidente
N. BOUSTIA	Maître de Conférences(A), U.Blida-1-	Examinatrice
W. K. HIDOUCI	Professeur , ESI	Examineur
Dj.BENNOUAR	Maître de conférences(A) , U.Bouira	Promoteur

Blida, Octobre 2015

## RÉSUMÉ

Il s'agit ici d'entamer une recherche dans le monde de la correction automatique des réponses aux questions dans le contexte d'un examen en ligne. Cette recherche touche le domaine du traitement du langage naturel, mais ne prend pas le langage naturel dans toutes ses dimensions, mais il considère de manière très sérieuse le contexte, ou domaine, dans lequel le langage naturel est utilisé. Ainsi, le processus de réponse à une question de géographie ou une question de programmation en Java devra considérer fortement le vocabulaire et la construction de phrases dans ces deux domaines

## ABSTRACT

This is to start a search in the world of automatic correction of answers to questions in the context of an online exam. This research would affect the field of natural language processing, but should not take the natural language in all its dimensions, but rather consider very seriously the context or domain in which natural language is used. Thus, the process of responding to a question of geography or a question of Java programming will strongly consider the vocabulary and sentence construction in both areas

## ملخص

ابحث في عالم التصحيح التلقائي للأجوبة على الأسئلة في سياق الامتحان عبر الإنترنت. وهذا من شأنه البحث في مجال معالجة اللغة الطبيعية، ولكن لا ينبغي أن تأخذ اللغة الطبيعية في جميع أبعادها، وإنما تنظر بجدية جدا إلى السياق أو المجال الذي يستخدم اللغة الطبيعية. وبالتالي، فإن عملية الرد على سؤال الجغرافيا أو مسألة البرمجة في جافا يعتمد على بناء المفردات والجمل

## DÉDICACES

Aux êtres les plus chers à mon cœur, ma mère et mon père, qui m'ont toujours tout donné : amour, affection, soutien,...Je ne saurai tout citer, ni trouver les mots pour les remercier, que Dieu me les garde inchaAllah !

Mon Mari !

Mon jeune enfant.

mes chères et adorables sœurs, merci pour votre soutien et vos encouragements

!

toute ma famille !

mes enseignants de l'USDB!

tous ceux qui ont contribué à ma formation !

tous ceux qui m'ont aidé et soutenu !

Rachda, qui vous remercie tous !

## REMERCIEMENTS

Tout d'abord, je présente mes sincères remerciements et ma plus grande reconnaissance à mon directeur de mémoire Dr. Bennouar Djamel, maître de conférences à l'Université de BOUIRA, pour avoir accepté de m'encadrer et pour m'avoir orienté, aidé et encouragé tout au long de la réalisation de ce travail. Je lui adresse encore mes vifs remerciements pour la pertinence de ses conseils et remarques, pour son soutien moral et pour tout le respect et la confiance qu'il m'a toujours manifestés.

Je présente tous mes respects et mes sincères remerciements aux membres du jury qui ont accepté d'évaluer ce travail.

## TABLE DES MATIERES

Résume.	
Abstract	
ملخص.	
Dédicaces	
Remerciements	
Table des matières	
Liste des figures	
Liste des tableaux.	
1. Introduction générale.....	12
Chapitre 1 : les approches d'évaluation de l'apprenant.....	15
1.1. introduction.....	15
1.2. L'évaluation de l'apprenant.....	17
1.3. Objectifs de l'évaluation.....	18
1.4. Les type d'évaluations :.....	19
1.5. Les types de questions.....	20
Chapitre 2 techniques pour l'évaluation automatique des réponses courtes	22
2. Les techniques utilisées.....	22
2.1. Introduction.....	24
2.2 contexte.....	25
2.3 la similarité entre texte.....	26
chapitre 3: état de l'art sur l'évaluation automatique de réponses courtes.....	36
3.1. Introduction.....	36
1) solution a plate forme pédagogique sous licence libre.....	37
1. moodle (modular object-oriented dynamic learning environment) :.....	38
2.claroline.....	39
3.efront.....	40
4.atutotr.....	40
5.chamilo.....	41
6.olat.....	42
7.openelms e-learning.....	42
8.hotpotatos :.....	43
b. solution académique.....	44
1. Introduction.....	44
2.the intelligent essayassessor (iea) :.....	44
3.e-rater (electronic essay rater ).....	45
4.c-rater (conceptual rater) : .....	46
5.intellimetric : .....	48
6.criterion : .....	48
7.sear(schema extract analyze and report) : .....	48
8.iems (intelligent essay marking system) .....	48
9.apexassessor.....	48
10.automatedtext marker (atm) .....	49

11.bayesian essay test scoring system (betsy) : .....	50
12.automark.....	51
13.autotutor.....	52
14.ps-me(paperless school free -text marking engine) .....	52
15.carmelc.....	53
16.erb(evaluatingresponseswith bleu) .....	53
17.atenea(willow) .....	54
18.summary street.....	54
19.texas.....	55
c. évaluation des solutions existantes: .....	59
chapitre 4 l'évaluation automatique basée sur les indications de l'enseignant.....	61
4.1. approche définie.....	61
4.2. introduction.....	62
4.3. les objectifs : .....	63
4. détermination des éléments a considérer par l'enseignant :.....	64
5. le système de graduation automatique définie.....	65
1. spécification des réponses référence et les indications enseignants :.....	66
2. génération d'un corpus : .....	72
6. la phase de l'évaluation : .....	73
chapitre 5 évaluations de l'approche.....	74
5.1. Introduction : .....	74
5.2. Environnement de réalisation.....	75
5.3-l'interface graphique : .....	82
5.4. résultats.....	86
5.5. discussion : .....	87
conclusion générale.....	89
Bibliographie.....	92



## LISTE DES FIGURES

Figure 1.1 :Taxonomie de Bloom	18
Figure 1.2 : Processus d'évaluation des réponses courtes [3]	20
Figure 2.1 : La distance de Levenshtein	29
Figure 2.2:La distance de JaroWinkler	31
Figure 2.3:méthode d'évaluation par rubrique	39
Figure 3.1 : Le coté configuration de teste sous Atutor	45
Figure 3.2:le Coté Utilisateur chez Atutor en cas : choisir un seul mot	45
Figure 3.3:le Coté Utilisateur chez Atutor en cas : choisirun paragraphe	45
Figure 3.4:La plateforme Chamilo et les réponses courtes	46
Figure 3.5:Interface OLAT sur les réponses Courte	46
Figure 3.6:Copies d'écran d'Intelligent Essay Assessor[46]	49
Figure 3.7:Architecture IEA[47]	49
Figure 3.8 :Architecture de E-rater [10]	50
Figure 3.9:Les quatre étapes d'évaluations dans c-rater[40]	50
Figure 3.10: Les trois modules d'ApexAssessor	53
Figure 3.11:L'analyseur Syntaxique et sémantique d' Automatedtext marker	54
Figure 3.12: Scoring and recalculation window in BETSY[47]	55
Figure 3.13 : Vue d'ensemble de tous les modules impliqués dans le processus d'évaluation effectué par Automark[10]	56
Figure 3.14:Logiciel qui utiliseERB	57
Figure 3.15:un aperçu de l'interface de la version en ligne de Atenea	58
Figure 3.16:Summary Street feedback display[31]	59
Figure 4.1:le systeme de graduation automatique definie	72
Figure 4.2:Un exemple de réponse écrite en utilisant le contenu et les mots clés dynamique	74
Figure 5.1: renommer les instances du module	80
Figure 5.2: modification des tables	80
Figure 5.3 : succès d'installation	81
Figure 5.4 : l'ajout du module à la liste des modules.	81
Figure 5.5 : liste des types de question	82
Figure 5.6:le fichier install.xml	82
Figure 5.7:le fichier de définition des interfaces de notre module	83
Figure 5.8:fichier de programmation d'interface	83
Figure 5.9: Fichier d'implémentation des fonctions	83
Figure 5.10:L'interface native de Module Short Answer	84
Figure 5.11: La partie en commun avec notre module	84
Figure 5.12: les options de configuration du réponse enseignants	84
Figure 5.13:La Zone de Contrôle de la Correction Automatique	85
Figure 5.14:Grading controller	85
Figure 5.15:Multiple Line SubController	85
Figure 5.16:Preprocessing Controller	86
Figure 5.17:Stop word subcontroller	86
Figure 5.18:Character Controller	86
Figure 5.19:Keywords controller	87
Figure 5.20:Espace subcontroller	87

Figure 5.21:Zone spécifique	88
Figure 5.22:La partie de génération du corpus	88
Figure 5.23: cas d'ignoration espace	89
Figure 5.24: cas d'ignoration des accents	89
Figure 5.25: cas des majuscules et minuscules	89

## LISTE DES TABLEAUX

Tableau 1.1 :Taxonomie de Bloom	18
Tableau 1.2: les types de question existant selon Bloom[12]	24
Tableau 3.1:Statistiques du l'utilisation du logiciel Moodle en 2013 et 2014	41
Tableau 3.2:Les propriétés de Moodle vis a vis les réponses courtes	42
Tableau 3.3:Les propriétés de Claroline vis a vis les réponses courtes	42
Tableau 3.4:Les propriétés de eFront vis a vis les réponses courtes	43
Tableau 3.5:Les propriétés de Hotpotatos vis a vis les réponses courtes	46
Tableau 3.6:Exemple des question avec C-rater [50]	50
Tableau 3.7:Tableaux récapitulatif sur les logiciel Acadimique	62
Tableau 4.1: SASF	73
Tableau 4.2 : GPC SASF paramètre classes	74
Tableau 4.3:Exemple des indications dela question mentionée dans la figure 21, specifiéutilisant GPI	76
Tableau 4.4: le dictionnaire généré et les pénalités associées	78
Tableau 4.5: GESA	80

# 1. INTRODUCTION GENERALE

## INTRODUCTION

Grâce à l'évolution technologique, les Technologies de l'Information et de la Communication (TIC) ont pu intégrer divers domaines. L'éducation et la formation n'ont pas échappé non plus à cette avancée avec l'émergence du nouveau secteur de l'enseignement/formation à distance. Ce secteur bénéficie, dans une large mesure, des progrès en cours, notamment l'utilisation de l'Internet comme moyen de communication à distance.

Depuis plusieurs années, les spécialistes de l'informatique et des sciences de l'éducation se sont beaucoup intéressés à ce nouveau secteur par l'élaboration d'outils et de logiciels pour l'apprentissage à distance. Toute fois, ils se sont davantage centrés sur la manière de transmettre les enseignements que sur leur évaluation, alors que celle-ci se situe au cœur même d'un enseignement.[1]

Aujourd'hui, les systèmes de télé-enseignement semblent avoir atteint un bon degré de maturité qui permettrait de drainer un nombre de plus en plus important d'enseignants à l'utilisation de ces techniques pour , d'une part , la diffusion des ressources pédagogique, la communication plus efficace entre apprenant, enseignant et concepteur de cours ; et, d'autre part, le contrôle continu des étudiants.

Le contrôle continu à travers les tests en ligne semble être le point le plus attractif pour les enseignants. Cependant ce point ne semble pas avoir atteint un haut degré de maturité pour qu'il puisse devenir un facteur important d'attraction des enseignants à l'utilisation des plateformes de télé-enseignement.

Au contraire, il semble être aujourd'hui le point le plus faible qui contribue plutôt au désintéressement d'un nombre important d'enseignants, surtout ceux qui ne trouvent pas, dans le contexte de leurs matières enseignées, comment exprimer des questions dans le contexte des techniques traditionnelles telles que les diverses variantes de QCM et les contraintes importantes imposées aux étudiants dans le contexte des réponses trop courtes, réduites souvent à un mot ou une valeur.

### Problématique :

Les réponses courtes, ou l'étudiant est amené à écrire plusieurs mots semble être une technique qui permettrait d'évaluer de manière assez efficace l'acquisition des connaissances d'un cours, et permettre à l'enseignant de voir quels sont les points acquis et les points de difficulté pour les étudiants.

Cependant la situation actuelle des outils de correction automatique de réponse courte ne semble pas avoir atteint un haut degré de maturité et impose aux examinés des compétences et des contraintes importantes.

Ainsi, ces outils ne permettent pas de traiter les faiblesses dans l'orthographe ni les erreurs d'inattention telle que la répétition erronée de lettre, l'oubli d'un accent, l'utilisation d'un mauvais déterminant etc.

### Objectif:

L'objectif de notre travail consiste à étudier les diverses techniques de corrections automatique, notamment la correction de réponse courte et proposer une approche pour une correction automatique de réponse courte plus efficace que les solutions existantes.

L'idée principale de notre approche est d'associer efficacement l'enseignant au processus de correction automatique. Cette association consiste à mettre à la disposition de l'enseignant diverses fonctionnalités qui lui permettront de guider à l'avance le processus de correction automatique.

La conséquence directe de cette association est, d'une part, l'adaptation de la correction automatique au domaine (ou matière) concerné par l'évaluation et, d'autre part, ne nécessiterait pas de mettre en place des algorithmes sophistiqués qui, malgré leur sophistication, ne donneront peut être pas de résultats satisfaisants.

Un autre objectif de taille, consiste à travailler dans le contexte de la communauté Moodle et ajouter à ce dernier un module. Ainsi, la maîtrise du code de ce dernier et de la méthode de mise en place et test de module est un travail important que nous devons achever.

### Organisation du mémoire :

Pour réaliser ce projet, nous avons organisé notre mémoire en six chapitres répartis comme suit :

Le premier chapitre est consacré aux introductions et généralité sur l'évaluation de l'apprenant.

Le deuxième chapitre présente les techniques de classification utilisées pour l'évaluation automatique des réponses en ce qui concerne toutes les méthodes de similarité des textes qui seront illustrées par des exemples ou des algorithmes.

Le troisième chapitre propose un état de l'art sur les outils existants dans le domaine d'E-Learning, Nous distinguerons ceux qui permettent d'effectuer des évaluations automatiques de réponses courtes que nous avons divisées en deux parties :

La première partie concerne les plateformes sous licence libre et la deuxième partie :

Concerne les solutions académiques.

Le quatrième chapitre présente notre approche de L'évaluation automatique basée sur les indications de l'enseignant

Le cinquième chapitre est basé sur l'évaluation de l'approche qui est faite par la construction d'un module qui a été intégré à la plateforme Moodle

Enfin, pour conclure ce mémoire, nous effectuerons un bilan de nos travaux et nous ouvrirons des perspectives de recherche

# CHAPITRE 1 : LES APPROCHES D'EVALUATION DE L'APPRENANT

## 1. Généralité

### 1.1. Introduction

Dans cette partie nous allons avoir un aperçu global sur l'évaluation, les types d'évaluations et les types de questions existants.

### 1.2. L'évaluation de l'apprenant

En (1956), Bloom distingue trois grands domaines d'apprentissage : cognitif, affectif et psychomoteur. Le domaine cognitif décrit la connaissance et le développement des habilités et compétences intellectuelles (le savoir et le savoir-faire), le domaine affectif décrit les aptitudes ou le savoir-être et le domaine psychomoteur décrit les habilités physiques et motrices.

Bloom a proposé une taxonomie qui permet de classer les objectifs pédagogiques en six niveaux. Tous ces objectifs se traduisent en comportements attendus des apprenants et ces niveaux sont classés en fonction de la complexité cognitive de l'objectif visé ; la figure suivante décrit ces niveaux [1]

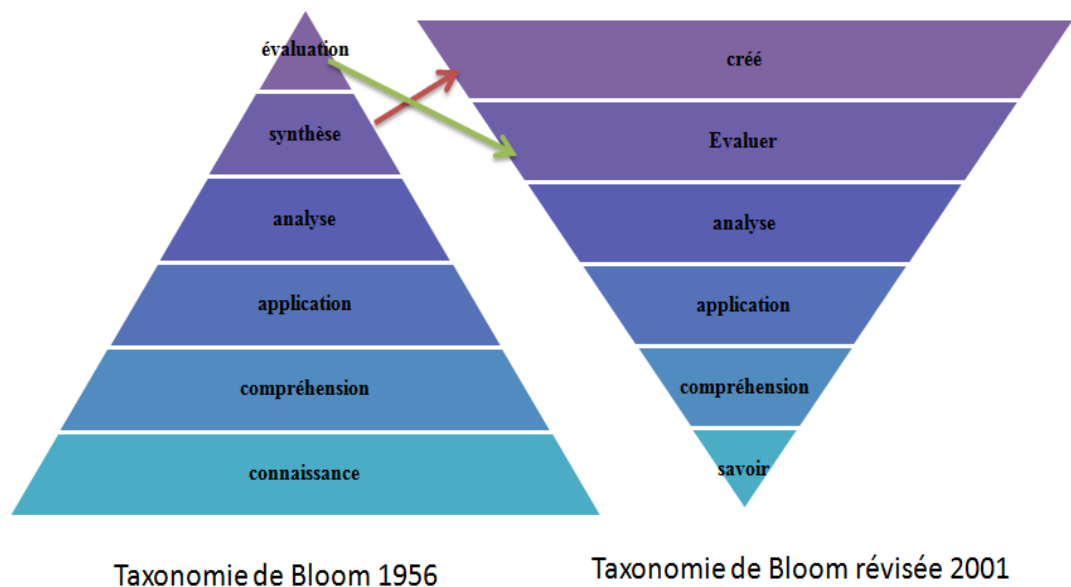


Figure 1.1 :Taxonomie de Bloom

	compétence	question	compétences prouvé	Méthodes d'évaluation
1	connaissance	Énumérer et définir	Rappel de l'information	questionnaire
2	compréhension	Résumer et prévoir	Saisir le sens du texte	un examen
3	application	Illustrer et résoudre	Utilisation pratique de la matière	Un jeu de rôle
4	analyse	Déduire et comparez	Modèles de préavis et des données cachées	Analyser une sortie d'ordinateur
5	synthèse	Intégrer et réécrire	informations Digest	Les questions ouvertes
6	évaluation	Recommande		

Tableau 1.1 :Taxonomie de Bloom

### 1.3. Objectifs de l'évaluation

L'évaluation doit rendre compte du degré auquel les individus en formation répondent à ces intentions ou à des attentes[2] parmi les objectifs que nous pouvons citer :

#### pédagogique :

- fournir à l'étudiant des rétroactions sur la progression de son apprentissage :
  - l'aidant à identifier ses problèmes d'apprentissage.
  - pour lui permettre d'améliorer son rendement.
- guider l'apprenant dans son processus d'apprentissage.
- permettre aux enseignants de réajuster et améliorer leurs enseignements.
-



✚ mesure :

Évaluer les connaissances, la compréhension, les aptitudes ou les compétences des étudiants.

✚ certification :

Certifier le niveau de réalisation atteint (compétences acquises) par l'apprenant pour un public visé :

- employeur.
- institutions éducatives.

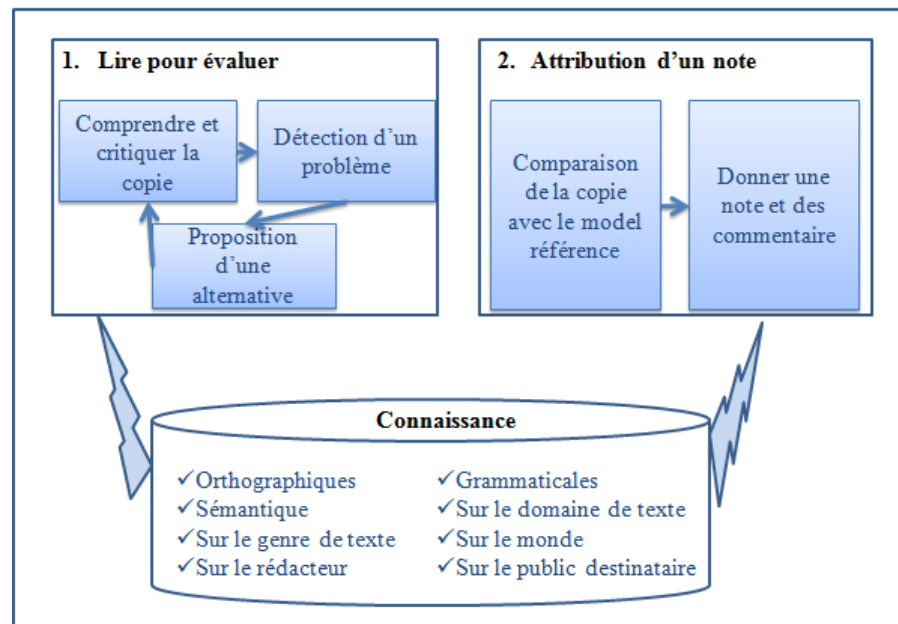


Figure 1.2 : Processus d'évaluation des réponses courtes [3]

1.4. Les types d'évaluation :

On peut faire un certain nombre de distinctions importantes en ce qui concerne l'évaluation. La liste ci-après n'est pas exhaustive mais seulement indicative.

a. évaluation formative :

Le terme d'évaluation formative a été proposé par M. Scriven[3][4]. C'est une évaluation qui se produit à n'importe quel moment au cours d'un processus d'apprentissage[1].

Dans ce genre d'évaluation, le plus important est de les aider, interagir entre l'évaluateur (éventuellement humain ou machine) et l'étudiant au cours du processus d'apprentissage,[5]

b. évaluation sommative :

Destinée à faire des bilans, la somme des compétences et des connaissances des étudiants après une séquence d'enseignement ou après une période donnée, le but est donc de découvrir s'ils ont assez de compétence pour passer le module, unité, chapitre, etc ou pas. [6][7][4]

c. évaluation alternative :

La définition de ce type d'évaluation a été introduite par Hancock qui dit « *Évaluation alternative est un processus continu impliquant l'élève et l'enseignant à porter des jugements sur les progrès de l'élève dans la langue en utilisant des stratégies non conventionnelles* »[8].

L'exemple serait de demander à des étudiants de gérer un projet de date limite, et de les interviewer une fois par semaine pour découvrir leur progrès. Son objectif principal est d'améliorer l'apprentissage des étudiants en les laissant employer leurs capacités de recherche et avec leurs ressources [9]

d. évaluation holistique :

Il signifie que l'instructeur donne des points globaux pour la réponse basée sur des critères multiples. Par exemple, pour un étudiant de l'EEL (l'anglais en tant que deuxième langue), l'examineur tiendrait compte de la qualité d'écriture, le vocabulaire, la grammaire, le modèle, le contenu et l'épellation et puis, avec une certaine équation qui les combine tout, calculerait les points holistiques. [9]

e. évaluation analytique :

Elle se compose des évaluations multiples des composants du matériel écrit. En notant pour chaque thème considéré comme une note.

Définir les éventuels critères de comparaison (est-ce que les techniques utilisées dans la littérature sont efficaces dans le cadre d'une plateforme de télé-enseignement, le domaine d'application, la langue, y a t'il un outil qui a été réalisé, le degré d'intervention de l'évaluateur avant de lancer l'évaluation automatique, sur quels aspects intervient l'évaluateur, prise en charge de la connaissance correcte de la langue de réponse par ceux qui sont évalués etc.)[9]

### 1.5. Les types de questions

Les paragraphes suivants décrivent l'ensemble des types de questions existant dont :

### 1) Questions fermées

L'inconvénient majeur de ces réponses est le fait que l'étudiant doit répondre obligatoirement à ce qui est proposé, il ne peut ni modifier ni s'exprimer avec sa propre réponse :

a) question à choix simples (multiple choice question (mcq)):

C'est un type de question dans lequel les étudiants doivent choisir quelle est la réponse correcte parmi plusieurs.

b) question à choix multiples (multiple response question (mrq)):

Elle est très semblable à une question de MCQ avec la différence principale que la clef n'est pas unique et plusieurs réponses pourraient être exactes.

c) question hot-spot :

Dans ce type de question l'étudiant doit indiquer la réponse en pointant vers une partie spécifique de l'image qui est la solution à la question. Ces types de question sont insuffisants pour évaluer les étudiants [5][11] .

d) question de type association ou appariement

Ces questions proposent à l'apprenant deux listes de choix où il devra faire correspondre ou associer des éléments de la première liste avec un ou plusieurs éléments de la deuxième liste.

e) question de type réarrangement

Ces questions proposent à l'apprenant de mettre dans un certain ordre une liste d'éléments qui font partie de l'énoncé. Celui-ci contient des propositions dans un ordre quel- conque. L'apprenant doit alors les « réarranger » dans le bon ordre.

### 2) Questions semi-ouvertes

Ce type de questions est un compromis entre le type précédent et le type suivant, il ne permet pas à l'apprenant de s'exprimer dans son propre langage, mais, celui-ci à la possibilité de répondre de plusieurs façons différentes. La

réponse à ce genre de question doit être brève et précise et pour laquelle on donne des orientations. Nous distinguons :

1. question à réponse courte

L'étudiant répond à la question par un ou plusieurs mots. L'enseignant pose des questions claires et précises qui imposent des réponses brèves et spécifiques. Généralement, les réponses correspondent à des mots ou de courtes expressions.

2. question de type « carte conceptuelle »

La carte conceptuelle est une représentation graphique des idées et des relations entre ces idées de manière à faciliter la visualisation de l'ensemble.

3) Questions ouvertes

C'est le type le plus complexe de question à évaluer. Les étudiants doivent écrire un texte au sujet de matières demandées avec la seule aide de leurs propres expériences et connaissance. Nous distinguons deux types de réponses : les réponses longues et les réponses courtes.

Les réponses courtes sont des réponses brèves et strictes à un sujet bien défini (avec un mot ou une phrase), le professeur peut évaluer l'étudiant avec un ensemble de question que l'on définit ici en suivant la taxonomie de Bloom tel que montré dans ce tableaux :

	Taxonomie de Bloom	Habilités ciblées	Exemples de questions
Niveaux 1	Acquérir des connaissances	identifier, nommer, énumérer, rappeler, décrire, énoncer, déterminer, nommer	- Qui...? - Qu'est-ce...? - Décris... - Nomme
Niveaux 2	Comprendre	raconter, interpréter, ressortir, résumer, réviser, déduire, expliquer	- Quelle est l'idée principale...? - Expliquer ce que signifie... - Quels sont les faits? - Quelles sont les opinions?
Niveaux 3	Appliquer	appliquer, préparer, simuler, découvrir, utiliser, s'exercer	- Pourquoi...est-il important? - Qu'est-ce qui t'arriverait si...? - Comment...?
Niveaux 4	Analyser	comparer, ordonner, opposer, classer, distinguer, déduire, mettre en perspective	- Comment...se compare-t-il à...? - Quel est le rapport entre...et...? - Quelles sont les causes et les effets de...?
Niveaux 5	Évaluer	évaluer, justifier, critiquer, apprécier, juger	- Qu'est-ce qui est plus important et pourquoi? - Lequel est le plus logique? - Quels sont les forces et les défis? - Quel est ton point de vue au sujet de...? - Évalue les effets et les causes de... - Quelle est ton opinion sur...?

Niveaux 6	Créer	élaborer, concevoir, composer, organiser, résoudre planifier, produire, réaliser	<ul style="list-style-type: none"> <li>- Quelle est la solution à...?</li> <li>- Quel effet...a-t-il sur...?</li> <li>- Élabore un organisateur graphique pour...</li> <li>- Quelles sont les solutions de rechange à ...</li> <li>- Invente un nouveau...avec ce que tu as trouvé?</li> </ul>
-----------	-------	--	--

Tableau 1.2: les types de question existant selon Bloom[12]

On peut trouver les réponses courtes que l'on traite dans notre étude dans les quarts niveaux de Bloom et parfois le niveau cinq.

## **CHAPITRE 2 TECHNIQUES POUR L'EVALUATION AUTOMATIQUE DES REPONSES COURTES**

### **2. LES TECHNIQUES UTILISEES**

#### 2.1. Introduction

L'évaluation automatique des réponses courtes est basée fondamentalement sur la comparaison de la réponse de l'enseignant à la réponse de l'étudiant. Selon le degré de similarité entre les deux réponses, une évaluation de la réponse de l'étudiant peut être déduite.

La similarité entre textes est ainsi le cœur des approches d'évaluation automatique de réponses courtes. Dans ce chapitre ; nous présentons certaine approche permettant la comparaison de textes, les approches présentées ont été sélectionnées pour répondre au mieux au contexte et donner une aperçu des méthodes les plus utilisées dans le contexte de notre étude.

#### 2.2 Contexte

Notre étude a pour but de développer un outil de correction semi automatique pour l'évaluation des étudiants lors les examens en ligne. Notre problème est la comparaison des textes dont l'un c'est la réponse de l'étudiant et l'autre c'est la réponse modèle du professeur. Les textes sont saisis à l'aide d'un outil de connexion que ce soit un ordinateur, une tablette ou un Smartphone, les calculs de résultats se font en temps réel sur un serveur distant. Les résultats dans ce cas doivent être pertinents et obtenus très rapidement. Nous avons procédé comme piste d'étude à la similarité entre les textes.

#### 2.3 La similarité entre texte

Evaluer la similarité entre deux documents est une problématique dans plusieurs disciplines comme l'analyse de données textuelle, la recherche d'informations ou l'extraction de connaissances à partir de données textuelles. Dans chacun des domaines, la similarité est utilisée pour différents objectifs.

En analyse des données textuelles (ADT) : les similarités sont utilisées pour la description et l'exploration de données.

En recherche d'information : l'évaluation de la similarité entre document et requête est utilisée pour identifier les documents pertinents par rapport au besoin d'informations exprimé par les utilisateurs

En text mining les différents similarités sont utilisés pour produire des représentations synthétiques de vastes collections de documents [10].

#### a) Similarité syntaxique

Consiste à comparer deux chaînes de caractère en mesurant leur similarité en utilisant des approches algorithmiques. Parmi les mesure de similarité nous citons la distance de Lavenstein [11]; le coefficient de Dice[12] ,l'indice de Jackard[13] , le Cosinus [14] etc.

#### 1. la similarité basée sur les chaînes de caractère[14]

La mesure est basée sur les séquences de chaînes et la composition de caractères pour comparer la similarité. Une mesure de la chaîne est un indicateur qui mesure la similarité ou de dissemblance (distance) entre deux chaînes de texte pour l'appariement de chaîne ou la comparaison, on les décompose en deux catégories mesure basée caractère et mesure basée terme.

#### a. La mesure basée caractères :

##### i. la distance de Hamming

La distance de Hamming,  $d(z_1 ; z_2)$ , entre deux mots de même longueur  $z_1$  et  $z_2$  (c.-à-d. deux n-uplets dans le cas le plus général) est le nombre de symboles (c.-à-d. de positions) pour lesquels  $z_1$  et  $z_2$  diffèrent[15] . L'algorithme de Needleman et Wunsh est basé sur cette distance.

Exemple :

- ✓ La distance de Hamming entre 101010 et 111010 est un car ces deux mots diffèrent seulement par la deuxième position.
- ✓ La distance de Hamming entre 1010 et 0101 est quatre car ces deux mots diffèrent par toutes les positions.
- ✓ La distance de Hamming entre 1010 et 111010 n'est pas définie (et ne sera pas considérée).



### ii. la distance de Needleman- Wunsch

L'algorithme de Needleman-Wunsch est un algorithme qui effectue un alignement global maximal de deux chaînes de caractères. Il est couramment utilisé en bioinformatique pour aligner des séquences de protéines ou de nucléotides. L'algorithme a été présenté en 1970 par Saul Needleman et Christian Wunsch dans leur article « *A general method applicable to the search for similarities in the aminoacidsequence of twoproteins* ».[11]

### iii. la distance de Levenshtien

La distance de Levenshtein est une distance mathématique donnant une mesure de la similarité entre deux chaînes de caractères. Elle est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne à l'autre. Elle a été proposée par Vladimir Levenshtein en 1965. Elle est également connue sous les noms de distance d'édition ou de déformation dynamique temporelle, notamment en reconnaissance de formes et particulièrement en reconnaissance vocale[11].

Exemple :

		m	e	i	l	e	n	s	t	e	i	n
l	0	1	2	3	4	5	6	7	8	9	10	11
e	1	1	2	3	3	4	5	6	7	8	9	10
v	2	2	1	2	3	3	4	5	6	7	8	9
e	3	3	2	2	3	4	4	5	6	7	8	9
n	4	4	3	3	3	3	4	5	6	6	7	8
s	5	5	4	4	4	4	3	4	5	6	7	7
h	6	6	5	5	5	5	4	3	4	5	6	7
t	7	7	6	6	6	6	5	4	4	5	6	7
e	8	8	7	7	7	7	6	5	4	5	6	7
i	9	9	8	8	8	7	7	6	5	4	5	6
n	10	10	9	8	9	8	8	7	6	5	4	5
n	11	11	10	9	9	9	8	8	7	6	5	4

Figure 2.1 : La distance de Levenshtein

### iv. la distance de smith-waterman

L'algorithme de Smith-Waterman est un algorithme d'alignement de séquences utilisé notamment en bioinformatique. Il est par exemple utilisé pour aligner des séquences de nucléotides ou de protéines. Cet algorithme a été

inventé par Temple F. Smith et Michael Waterman en 1981. L'algorithme de Smith-Waterman est un algorithme optimal qui donne un alignement correspondant au meilleur score possible de correspondance entre les acides aminés ou les nucléotides des deux séquences. Le calcul de ce score repose sur l'utilisation de matrices de similarité ou matrices de substitution.[14]

#### v. La distance de Jaro Winkler

La distance de Jaro-Winkler mesure la similarité entre deux chaînes de caractères. Il s'agit d'une variante proposée en 1999 par William E. Winkler, découlant de la distance de Jaro (1989, Matthew A. Jaro) qui est principalement utilisée dans la détection de doublons.

Plus la distance de Jaro-Winkler entre deux chaînes est élevée, plus elles sont similaires. Cette mesure est particulièrement adaptée au traitement de chaînes courtes comme des noms ou des mots de passe. Le résultat est normalisé de façon à avoir une mesure entre 0 et 1, le zéro représentant l'absence de similarité.[11]

L'indice de Jaro-Winkler(Winkler, 1990)(Winkler, 2006), appelé aussi distance de Jaro- Winkler, est basé sur l'indice de Jaro (Jaro, 1989)(Jaro, 1995). Il complète l'indice de Jaro en tenant compte de l'ordre des caractères. Les valeurs obtenues sont plus élevées s'il s'agit de chaînes de caractères ayant le même préfixe.

La similarité de deux chaînes de caractères C1 et C2 précédentes est calculée suivant l'indice, ou distance, de Jaro de la manière suivante :

$$S_R = \frac{1}{3} \left( \frac{m}{|I1|} + \frac{m}{|I2|} + \frac{m-t}{m} \right) \quad (1)$$

Et ici le nombre de caractères de C1 et C2 correspondants entre eux, le nombre de caractères

Mt transposés, L1 est la longueur de C1 et L2 la longueur de C2. L'éloignement maximum admis pour les caractères, en-deçà duquel on pourra dire que les caractères sont correspondants est :

$$e = \left( \frac{\max(|I1|, |I2|)}{2} \right) - 1 \quad (2)$$

L'indice de Jaro-Winkler est calculé suivant la formule suivante :

$$S_W = S_R + L \cdot p(1 - S_R) \quad (3)$$

S étant l'indice de similarité de Jaro, L la longueur du préfixe commun aux deux chaînes de R Caractères, et p un facteur de pondération, destiné à favoriser ou pas les chaînes ayant un préfixe commun. La valeur proposée par Winkler pour p est 0,1. La longueur L du préfixe commun aux deux chaînes ne doit pas excéder Quatre.

Si l'on reprend l'exemple donné pour les indices de similarité précédents, nous pourrions construire avec les chaînes de caractères C1 et C2 le tableau suivant :

		1	2	3	4	5	6	7	8	9	10	11	12
		i	n	f	o	r	m	a	t	i	q	u	e
1	i	1	0	0	0	0	0	0	0	1	0	0	0
2	n	0	1	0	0	0	0	0	0	0	0	0	0
3	f	0	0	1	0	0	0	0	0	0	0	0	0
4	o	0	0	0	1	0	0	0	0	0	0	0	0
5	r	0	0	0	0	1	0	0	0	0	0	0	0
6	m	0	0	0	0	0	1	0	0	0	0	0	0
7	a	0	0	0	0	0	0	1	0	0	0	0	0
8	t	0	0	0	0	0	0	0	1	0	0	0	0
9	i	1	0	0	0	0	0	0	0	1	0	0	0
10	o	0	0	0	1	0	0	0	0	0	0	0	0
11	n	0	1	0	0	0	0	0	0	0	0	0	0

Figure 2.2: La distance de JaroWinkler

L'éloignement maximum admis pour les caractères correspondants est ici :

$$e = \left( \frac{\max(|l1|, |l2|)}{2} \right) - 1 = \frac{12}{2} - 1 = 5 \quad (4)$$

Notons que les caractères suivants sont éloignés l'un de l'autre de plus de cinq caractères, et ne seront pas considérés comme correspondants :

- ✓ le premier "i" de "information" et le deuxième "i" de "informatique"
- ✓ le premier "i" de "informatique" et le deuxième "i" de "information"
- ✓ le "n" de "informatique" et le deuxième "n" de "information" ? Le "o" de "informatique" et le deuxième "o" de "information"

Le nombre de caractères correspondants est :

$$M=9$$

Les longueurs respectives des deux chaînes de caractères sont :

$$|l_1|=12 \quad |l_2|=11$$

Le nombre de transpositions est :

$$t=0$$

L'indice de Jaro pour les deux chaînes est :

$$S_R = \frac{1}{3} \left( \frac{m}{|l_1|} + \frac{m}{|l_2|} + \frac{m-t}{m} \right) = \frac{1}{3} \left( \frac{9}{12} + \frac{9}{11} + \frac{9-0}{9} \right) = 0.82 \quad (5)$$

Pour  $l=9$  (longueur du préfixe commun aux deux chaînes : "informati") et  $p=0,1$  (valeur standard), l'indice de similarité de Jaro-Winkler pour les deux chaînes C1 et C2 sera donc:

$$S_W = S_R + l.p(1-S_R) = 0.82 + 9 \cdot 0.1(1-0.82) + 0.98 = 0.98 \quad (6)$$

#### vi. la similarité utilisant les n-gram

Un n-gramme est une séquence de n words qui apparaissent successivement dans le texte. Un n-gramme de taille « un » s'appelle un « uni gramme », un n-gramme de taille « deux » s'appelle « bi-gramme et » un n-gramme de taille « trois » s'appelle un « tri-gramme ». . n-gram mesures de cooccurrence la façon dont un résumé du candidat chevauche avec un résumé de référence en utilisant une moyenne pondérée [16]

#### vii. BLEU(bilingual evaluation understudy)

Algorithme créé par Papineni et al [17] pour évaluer les systèmes de traduction automatique d'une langue vers une autre langue [18] en utilisant la couverture des n-grammes de la traduction référence [19]; il a été utilisé pour évaluer google traduction et bing a (0.31) et (0.29) respectivement [20]; l'algorithme BLEU a été utilisé dont le principe est de faire une comparaison avec la réponse modèle de l'enseignant et la réponse de l'étudiant avec un indice de pénalité pour pouvoir augmenter sa performance

L'idée principale de la comparaison est de mesurer la proximité du candidat au Modèle référence nommée M-BLEU [21] pour plus de performance sur ce type d'algorithme il est nécessaire de faire un entrepôt de données de modèle référence des réponses de l'instructeur ou l'enseignant ; pour chaque réponse d'étudiant il est nécessaire d'avoir plusieurs réponses type pour augmenter la performance de l'algorithme et comparer la réponse de l'étudiant avec la réponse la plus adéquate .

**b. La mesure basé terme**

**i. La distance de Manhattan (city Block )**

Elle calcule la distance qui serait parcourue pour aller d'un point de données à l'autre si un chemin de grille est suivi. Elle est la somme des différences de leurs composantes correspondantes[14] alors la formule est notée comme suit :

$$d = \sum_{i=1}^n |x_i - y_i| \quad (7)$$

**ii. La similarité cosinus**

Elle mesure la similarité entre deux vecteurs en mesurant le cosinus de l'angle entre eux. [14]

L'indice de similarité dit « Cosine » (Singhal, 2001)(Tan, et al., 2005)est calculé comme étant le cosinus de l'angle formé par deux vecteurs:

$$Sc = \cos(\theta) = \frac{A \cdot B}{|A| |B|} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (8)$$

Le produit scalaire des deux vecteurs A et B est divisé par le produit des normes des deux vecteurs. Le résultat de ce calcul sera toujours compris entre 0 et 1.

Si l'on considère ici aussi les chaînes de caractères C1 = "informatique" et C2 = "information" ainsi que leur union {a, e, f, i, m, n, o, q, r, t, u}, nous pouvons traduire chacune des deux chaînes de caractères en un vecteur, en considérant le nombre d'occurrences de chacun des caractères contenus dans une chaîne par rapport à l'union des deux chaînes.

Nous obtiendrons : C1 = (1,1,1,2,1,1,1,1,1,1,1) et C2 = (1,0,1,2,1,2,2,0,1,1,0)

Cosine se calcule donc pour les chaînes de caractères C1 et C2 de la manière suivante :

$$lc = \frac{(1*1)+(1*0)+(1*1)+(2*2)+(1*1)+(1*2)+(1*2)+(1*0)+(1*1)+(1*1)+(1*0)}{\sqrt{1^2+1^2+1^2+2^2+1^2+1^2+1^2+1^2+1^2+1^2+\sqrt{1^2+0^2+1^2+2^2+1^2+2^2+2^2+0^2+1^2+1^2+0^2}}}$$

$$lc=0.41 \quad (9)$$

#### le coefficient de dice (dice's coefficient)

Elle est définie comme étant le double de l'intersection de deux lots X et Y et est divisée par la somme des deux lot[12]

$$S = \frac{2*|X \cap Y|}{|X|+|Y|} \quad (10)$$

Exemple :

Prenons l'exemple de deux chaînes "hello" et "allo", auquel ont souhaité déterminer le coefficient de Dice. Les bigrammes respectifs sont {he,el,ll,lo} (soit 4 bigrammes) et {al,ll,lo} (soit 3 bigrammes). L'intersection des deux lots sont les valeurs {ll,lo}. Ainsi, l'équation devient:

$$s = 2*2 / (4+3) = 4/7 = 0,57 \quad (11)$$

Grâce à cette valeur de 0,57, il est possible de déterminer que ces deux mots sont à peu près semblables. Plus la valeur est proche de 1 , plus les chaînes sont similaires. Si le résultat est égal à 1, cela signifie que les deux chaînes sont identiques.

Les chaînes de caractères C1 et C2 sont de longueurs respectives 12 et 11. Leur intersection {a,f,i,m,n,o,r,t} est de longueur 8. Le coefficient de Sorensen est :

$$Ss = \frac{16}{23} = 0.69 \quad (12)$$

### iii. l'indice de Jaccard

L'indice et la distance de Jaccard sont deux métriques utilisées en statistiques pour comparer la similarité et la diversité entre des échantillons. Elles sont nommées d'après le botaniste suisse Paul Jaccard.[13]

L'indice de Jaccard (Jaccard, 1912)(Jaccard, 1901)(Tan, et al., 2005)(Tanimoto, 1957)tient son nom du botaniste suisse Paul Jaccard (1868–1944). Il est calculé comme suit:

$$S_J = \frac{A \cap B}{A \cup B} \quad (13)$$

Si l'on considère, à titre d'exemple, les deux chaînes de caractères :

C1 = "informatique" et C2 = "information"

l'union de ces deux chaînes {a, e, f, i, m, n, o, q, r, t, u}est de longueur 11 et leur intersection {a,f,i, m, n, o, r, t} est de longueur 8.

L'indice de Jaccard est dans ce cas :

$$S_J = \frac{8}{11} = 0.72 \quad (14)$$

### iv. matching coefficient

Est une approche très simple à base de vecteurs qui comptent simplement le nombre de termes semblables, (dimensions), laquelle les deux vecteurs ne sont pas nuls[14].

#### b) Similarité sémantique

La similarité sémantique est un concept, selon un ensemble de termes qui se voit attribuer une métrique basée sur la ressemblance de signification/contenu sémantique.

Parmi de telles mesures de similarité, nous trouvons Resnik, LSA (Analyse sémantique Latente), ESA(Analyse sémantique explicite )

### i. La similarité basée corpus

Le principe de similarité basée corpus est de détecter la similarité entre les mots, selon les informations obtenues à partir d'un grand corpus comme Wikipédia .

ils existent plusieurs techniques qui utilisent ce principe tels que LSA, Explicit Semantic Analysis (ESA), Pointwise Mutual Information-Information Retrieval (PMI-IR) et Extraction des mots semblables distributionnels utilisant cooccurrences [20] .

Parmi les techniques les plus importantes : la technique LSA.

#### i. Latent Semantic Analysis (LSA)

C'est un modèle statistique développé par Landauer et al en 1988 qui permet la comparaison de similarité sémantique des mots [22] ; Cette méthode permet de construire des classes thématiques sous la forme de sacs de mots pondérés constitués par le vocabulaire d'un texte. Après différentes extensions, cette modélisation est utilisée pour traiter des données multilingues et en extraire les similitudes thématiques[13].

LSA est utilisé par une matrice qui capture les mots et calcule la fréquence d'apparition dans le texte ; la matrice est transformée en SVD(*Singular Value Décomposition*) ; ce dernier permet de déterminer la similarité entre les mots , il est capable de produire des résultats de score approximatif à des résultats d'experts ; Cependant ; LSA ne prend pas en considération l'ordre des mots. [23][24][25][14]

LSA a été utilisé par différent ITS (les systèmes du tutoring intelligent) comme AutoTutor[26], Intelligent EssayAssessor[27] , Summary Street [28]; Apex [29]etc.

#### ii. L'analyse sémantique explicite (ESA)

C'est une mesure utilisée pour calculer la proximité sémantique entre deux textes arbitraires, la technique basée sur Wikipedia représente des termes (ou textes) comme vecteurs de grandes dimensions, chaque entrée de vecteur présentant le poids de TF-IDF entre le terme et un article Wikipedia .



La proximité sémantique entre deux termes (ou textes) est exprimée par la mesure du cosinus entre les vecteurs correspondants[30]

iii. Point wise mutual information information retrieval (PMI-IR)

Est une méthode pour calculer la similarité entre des paires de mots, il utilise la syntaxe avancée des requêtes de recherche d'AltaVista pour calculer les probabilités.

Le plus souvent deux mots co-occurents l'un près de l'autre sur une page web est élevé, plus leur score de similarité PMI-IR est élevé.

iv. Analyse sémantique saillant[30]

SSA en anglais (Salient Semantic Analysis) intègre une abstraction sémantique similaire et l'interprétation des mots que l'ESA, mais il utilise des concepts saillants recueillies à partir de la connaissance encyclopédique, où un «concept» représente un mot ou une phrase sans ambiguïté avec un sens concret et qui offre une définition encyclopédique.

Dans ce cas, il est déterminé sur la base du mot étant un lien hypertexte dans son contexte, ce qui implique qu'ils sont très pertinents pour le texte donné.

SSA est un exemple de modèle Vecteur d'espace généralisé (GVSM), où les vecteurs représentent les associations de mots-concepts. [30][31]

1) L'algorithme KNN

L'algorithme KNN figure parmi les plus simples algorithmes d'apprentissage artificiel. Dans un contexte de classification d'une nouvelle observation  $x$ , l'idée fondatrice simple est de faire voter les plus proches voisins de cette observation. La classe de  $x$  est déterminée en fonction de la classe majoritaire parmi les  $k$  plus proches voisins de l'observation  $x$ . La méthode KNN est donc une méthode à base de voisinage, non-paramétrique ; Ceci signifiant que l'algorithme permet de faire une classification sans faire d'hypothèse sur la fonction  $y=f(x_1,x_2,\dots,x_p)$  qui relie la variable dépendante aux variables indépendantes.[32]

Pour utiliser l'algorithme KNN dans la correction automatique des essaie, le problème est résolu comme suit :

L'auteur utilise la Catégorisation du texte, ensuite il a transformé la rédaction en vecteur après pré-préparation (adverbe, pronoms, adjectif) ,il calcule la fréquence de chaque caractéristique et les transforme en vecteur de même dimension .

Pour algorithme KNN , toutes les différentes fonctions de l'essai composent un espace vectoriel. Chaque vecteur est exprimé par les poids TF-IDF, après on calcule la similarité entre l'essai test et les essais d'apprentissage avec la formule de cosinus[33]

## 2) L'évaluation Basée sur les Rubriques

Méthode utilisée pour évaluer les rédactions telle que définie dans des rubriques dont chacune est notée par une valeur, le score final est obtenue en additionnant toutes les notes. Comme par exemple NAPLAN comportent dix fonctions à évaluer comme (l'orthographe notée de 0 a six points, vocabulaire du 0 à 5, l'idée ;..ext) et chaque caractéristique est noté par 0 a6 ,à la fin une note globale est attribuée en sommant toutes les fonctions. [34]

	Criteria				Points
	4	3	2	1	
<b>Level of engagement in class</b>	Student proactively contributes to class by offering ideas and asking questions more than once per class.	Student proactively contributes to class by offering ideas and asking questions once per class.	Student rarely contributes to class by offering ideas and asking questions.	Student never contributes to class by offering ideas and asking questions.	
<b>Listening, questioning and discussing</b>	Respectfully listens, discusses and asks questions and helps direct the group in solving problems.	Respectfully listens, discusses and asks questions.	Has trouble listening with respect, and takes over discussions without letting other people have a turn.	Does not listen with respect, argues with teammates, and does not consider other ideas, Blocks group from reaching agreements.	
<b>Behavior</b>	Student almost never displays disruptive behavior during class discussions and group activities.	Student rarely displays disruptive behavior during class discussions and group activities.	Student occasionally displays disruptive behavior during class discussions and group activities.	Student almost always displays disruptive behavior during class discussions and group activities.	
<b>Preparation</b>	Student is almost always prepared with assignments and required class materials.	Student is usually prepared with assignments and required class materials.	Student is rarely prepared with assignments and required class materials.	Student is almost never prepared with assignments and required class materials.	
<b>Problem-solving</b>	Actively seeks and suggests solutions to problems.	Improves on solutions suggested by other group members.	Does not offer solutions, but is willing to try solutions suggested by other group members.	Does not try to solve problems or help others solve problems.	
<b>Group/partner teamwork</b>	Works to complete all group goals. Always has a positive attitude about the tasks and work of others. All team members contribute equally. Performed all duties of assigned team role.	Usually helps to complete group goals. Usually has a positive attitude about the tasks and work of others. Assisted team members in the finished project. Performed nearly all duties of assigned team role.	Occasionally helps to complete group goals. Sometimes makes fun of the group tasks and work of others. Finished individual task but did not assist team members. Performed some duties of assigned team role.	Does not work well with others and shows no interest in completing group goals. Often makes fun of the work of others and has a negative attitude. Contributed little to group effort. Did not perform duties of assigned team role.	
				<b>Total</b>	

Figure 2.3:méthode d'évaluation par rubrique

### 3) NLP

C'est l'application de l'intelligence Artificielle pour l'analyse du langage naturel. Plusieurs techniques utilisées dans ce domaine comme l'analyseur syntaxique pour trouver la structure linguistique d'un texte[35][36] l'analyseur de discours pour analyser la structure du discours dans le texte, mesures de similarité lexicale, pour analyser l'utilisation du mot d'un texte [35][37].

## CHAPITRE 3: ETAT DE L'ART SUR L'EVALUATION AUTOMATIQUE DE REPONSES COURTES

### 3. 1. Introduction

L'évaluation automatique des réponses libres est étudiée depuis les années soixante[38] , , et elle a attiré plus d'attention au cours des dernières années, surtout après l'introduction de E-Learning dont la plupart ne comptent que sur les types simples de questions, telles que des choix multiples, remplir les blancs ou questions « oui / non », mais cette façon d'évaluer est pas assez précise pour mesurer les connaissances de l'étudiant.[39]

Voici quelques plateformes pédagogiques sous-licence libre et des solutions académiques.

#### 1) Solution à plate-formes pédagogiques sous licence libre Moodle (Modular Object-Oriented Dynamic Learning Environment) :

Moodle est un logiciel permettant la mise en place de cours en ligne et de sites web. Il peut être utilisé par un seul professeur pour partager des cours ou servir plusieurs milliers d'étudiants dans une université. Moodle dispose de plus de 100 modules réalisés par des développeurs et des professeurs un peu partout dans le monde. Les outils les plus répandus permettent la mise en place de wiki, forum, gestion des devoirs, questionnaires, examens, messagerie, partage de documents, tchat, etc.

#### Les Statistiques du l'utilisation du logiciel Moodle[40]

	2013	2014
Sites enregistrés	85,505	55,086
Pays	240	231
Cours	8,309,310	7,468,886
Utilisateurs	77,393,569	67,785,570
Enseignants+	1,167,905	
Inscriptions	125,227,120	140,148,293
Messages des forums	140,396,231	132,587,033
Ressources	73,334,754	67,645,918
Questions de tests	235,516,949	261,534,450

Tableau 3.1: Statistiques du l'utilisation du logiciel Moodle en 2013 et 2014

Voici un tableau résume les propriétés de moodle vis avis les réponses courtes :

	Moodle
Le comportement vis-à-vis des espaces	Très sensible
Les accents (en Arabe Achakl)	non
Doublons	Très sensible
Expression ayant un aspect sonore très proche (on, an, am, om,..Dhad et Dad en Arabe)	non
Utilisation des pronoms (le, la el en Arabe etc..)	Très sensible
Conjugaison correcte	non
Association de pondération aux réponses	existe
Pénalité	existe
Association de dictionnaire	non
Construction de dictionnaire particulier contenant les mots à utiliser	oui
Association de correcteur d'orthographe	non
l'impact de l'enseignant dans la préparation des questions	Très fort
Sensible à la casse	Option à coucher

Tableau 3.2:Les propriétés de Moodle vis a vis les réponses courtes

La correction automatique de moodle est faite caractère par caractère ; comparaison du code ascii si les deux chaînes sont les mêmes la réponse est juste sinon elle est fausse quelle que soit la situation.

## 2. Claroline

Le comportement du module des réponses courtes de la plateforme Claroline est résumé dans ce tableau :

	Claroline
Le comportement vis-à-vis des espaces	Très sensible
Les accents (en Arabe Achakl)	non
Doublons	Très sensible
Expression ayant un aspect sonore très proche (on, an, am, om,..Dhad et Dad en Arabe)	non
Utilisation des pronoms (le, la el en Arabe etc..)	Très sensible
Conjugaison correcte	non
Association de pondération aux réponses	existe
Pénalité	existe
Association de dictionnaire	non
Construction de dictionnaire particulier contenant les mots à	oui

utiliser	
Association de correcteur d'orthographe	non
l'impact de l'enseignant dans la préparation des questions	Très fort
Sensible à la casse	oui

Tableau 3.3:Les propriétés de Claroline vis a vis les réponses courtes

### 3. eFront

La plateforme eFront offre une possibilité de réponses courtes d'après l'étude des propriétés de ce dernier on résume ce tableaux :

	Efront
Le comportement vis-à-vis des espaces	Très sensible
Les accents (en Arabe Achaki)	non
Doublons	non
Expression ayant un aspect sonore très proche (on, an, am, om,..Dhad et Dad en Arabe)	non
Utilisation des pronoms (le, la el en Arabe etc..)	non
Conjugaison correcte	non
Association de pondération aux réponses	existe
Pénalité	non
Association de dictionnaire	non
Construction de dictionnaire particulier contenant les mots à utiliser	non
Association de correcteur d'orthographe	non
l'impact de l'enseignant dans la préparation des questions	Très fort
Sensible à la casse	non

Tableau 3.4:Les propriétés de eFront vis a vis les réponses courtes

### 4. Atutor

Atutor est une plateforme de E-Learning. Son module de réponses courtes « Le module open ended question » n'a pas d'outil de correction automatique ; la correction est faite par l'enseignant lui-même ; la propriété offerte dans ce module est dans l'espace à utiliser pour répondre, si l'administrateur a choisi un seul mot il nous donne l'espace d'un seul mot sinon selon la taille de la réponse choisie

## Coté administrateur

The screenshot shows the 'Open Ended' question configuration page in Atutor. It includes a 'Category' dropdown menu set to 'Uncategorized', a 'Question ID' field with the value 'q1', and radio button options for 'Answer Size' (One Word, One Sentence, One Paragraph, One Page). There are also sections for 'Optional Feedback' and 'Banquet Content' with text input areas. A 'Save' button is located at the bottom left.

Figure 3.1 : Le coté configuration de testes sous Atutor

## ✓ Coté utilisateur

The screenshot shows the user interface for a 'Question 1: Open Ended' in Atutor. It features a navigation bar with 'Tests and Surveys', 'Create Test/Survey', 'Question Bank', and 'Question Categories'. Below the question title, there is a text input field labeled 'q1' and a 'Back' button.

Figure 3.2:le Coté Utilisateur chez Atutor en cas : choisir un seul mot

The screenshot shows the 'Preview' view of a 'Question 1: Open Ended' in Atutor. It includes the same navigation bar as Figure 3.2. The question ID 'q1' is shown above a large text area for the answer. A 'Back' button is located at the bottom left.

Figure 3.3:le Coté Utilisateur chez Atutor en cas : choisirun paragraphe

## 5. Chamilo

C'est l'une des plateformes qui utilise le type de question à réponse courte mais la correction n'ai pas automatique c'est à l'enseignants de les notés.



Figure 3.4:La plateforme Chamilo et les réponses courtes

## 6. OLAT

La question de texte libre est à disposition seulement dans les questionnaires. La réponse à cette question est indiquée librement par le participant du questionnaire dans un champ textuel de taille variable.

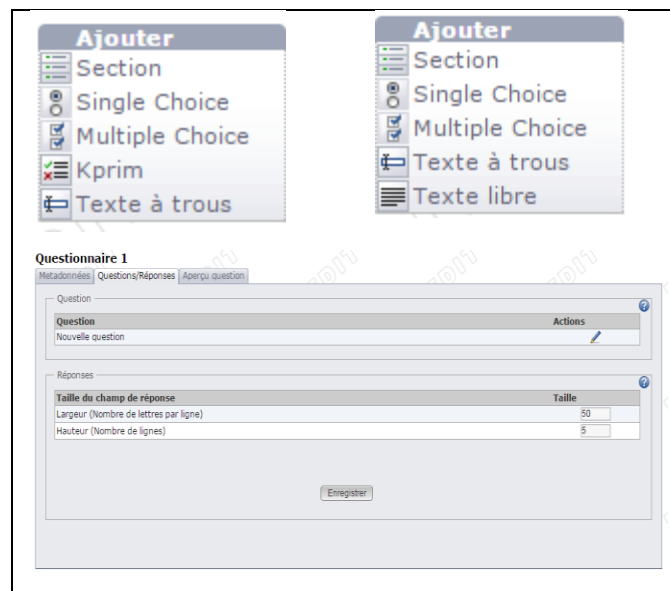


Figure 3.5:Interface OLAT sur les réponses Courte

Il ne fait aucun traitement sur les réponses, il ne fait que la configuration sur le nombre de lettres et des lignes pour l'affichage seulement,

## 7. OpenElms e-learning

Basé sur l'animation flash et la représentation des cours, les types de question sont simples sous forme de QCM. Il ressemble au cours du Cisco.



### 8. Hotpotatos :

Hotpotatos est un logiciel qui contient plusieurs types de question, voici Les résultats de teste qu'on a trouvé pour les questions à réponses courte présentés ci-dessous.

	Hot potatos
Le comportement vis-à-vis des espaces	Très sensible
Les accents (en Arabe Achakl)	non
Doublons	Très sensible
Expression ayant un aspect sonore très proche (on, an, am, om,..Dhad et Dad en Arabe)	non
Utilisation des pronoms (le, la el en Arabe etc..)	Très sensible
Conjugaison correcte	non
Association de pondération aux réponses	existe
Pénalité	N'existe pas
Association de dictionnaire	non
Construction de dictionnaire particulier contenant les mots à utiliser	oui
Association de correcteur d'orthographe	non
l'impact de l'enseignant dans la préparation des questions	Très fort
Sensible à la casse	oui

Tableau 3.5:Les propriétés de Hotpotatos vis a vis les réponses courtes

Pour améliorer les systèmes existant on a pris comme cas d'étude le logiciel open source Moodle ; qui fournit des informations pour les développeurs.

## b. Solution Académique

1. Introduction L'E-testing a commencé dans les années Soixante, parmi les solutions académiques Existants on cite :

### 2. The Intelligent EssayAssessor (IEA) :

Développé dans l'université du COLORADO aux USA à la fin des années 60 par Thomas K. Landauer; basé sur LSA , conçu pour l'indexation des documents et l'extraction de texte.[24], l'évalue la structure, le style ,le contenu en utilisant les techniques de l'Intelligence Artificiel[41]

Le principe de ce logiciel est d'utiliser les étudiants pour la rédaction d'un texte correspondant à ce qu'ils connaissent d'un cours, et les soumettent à l'ordinateur, qui compare ce texte à des textes-cibles préalablement sélectionnés. Après avoir « entraîné » LSA avec un corpus du domaine (cours), le texte de l'élève est comparé à une ou plusieurs copies-types, sélectionnée(s) par l'enseignant. Deux techniques ont été testées, donnant deux types de scores à la copie :

- ✚ le score « holistique », qui compare successivement le texte à noter à une série de copies notées au préalable par un jury. La note de la copie sera celle de la série de copies avec laquelle elle entretient la plus grande proximité, calculée par IEA. Une évaluation de ce calcul de score a été faite à partir de 190 copies de biologie, elle montre une corrélation de 0,80 entre les scores des évaluateurs humains et ceux calculés par IEA.

- ✚ Le score « étalon-or » (gold standard), qui compare le texte à noter avec une copie-modèle idéale, réalisée par exemple par l'enseignant.

La comparaison peut être globale ou paragraphe par paragraphe, de manière à vérifier si l'élève traite correctement chaque notion. Cette méthode est assez proche de la méthode humaine « par impression générale ».

Toutefois, un des avantages majeurs d'IEA, comparé à une évaluation humaine, est que l'étudiant peut soumettre son texte autant de fois que nécessaire, tant que ce dernier n'obtient pas la note voulue. Les auteurs observent d'ailleurs que les notes moyennes des textes soumis à IEA passent de 85/100 note de la première soumission à 92/100.[42][43]

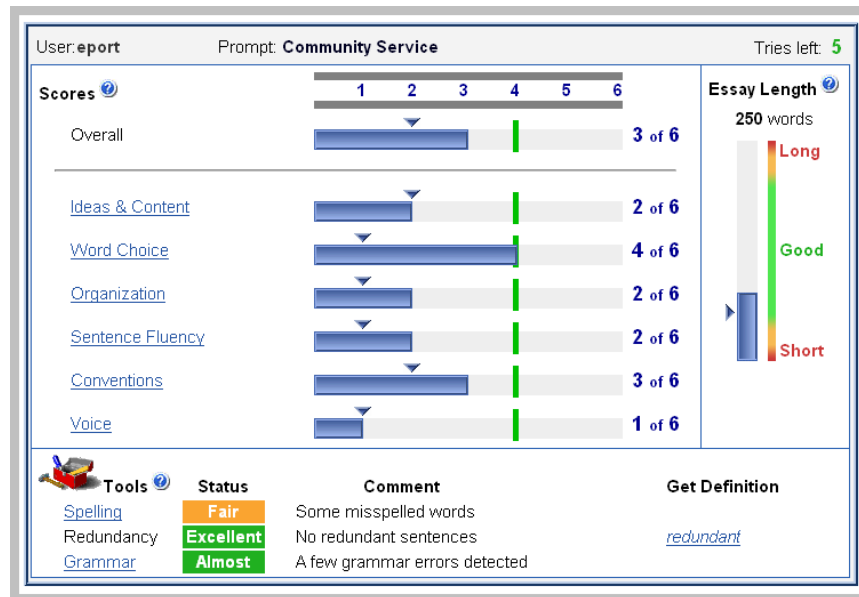


Figure 3.6: Copies d'écran d'Intelligent Essay Assessor[46]

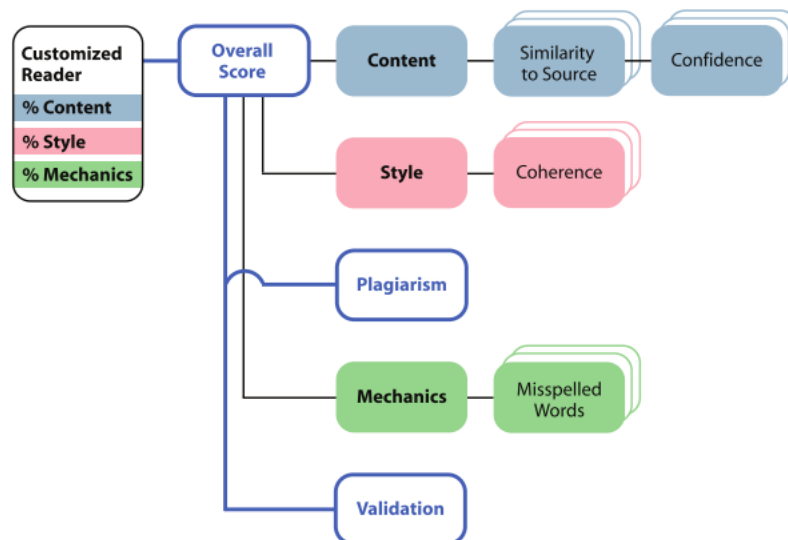


Figure 3.7: Architecture IEA[47]

### 3.E-rater (Electronic Essay Rater )

Est Développé par Burstein en 1998 [35]et utilisé par EducationalTesting Service (ETS) [44].

Il utilise une approche hybride basée sur les méthodes MNLPM(Microsoft Natural Langage Processing)[45]et les techniques de statistiques.

E-rater utilise quatre modules (syntaxe, discours, contenu thématique et complexité lexicale).

Il mesure l'organisation, la structure des phrases et leur contenu ;il est conçu pour mesurer les performances d'écriture [9]

E-rater était utilisé la première fois en 1999 comme un outil d'évaluation dans des examens comme TOFEL et GRE[41]

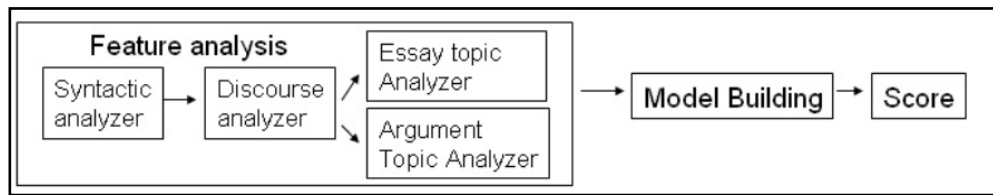


Figure 3.8 :Architecture de E-rater [10]

#### 4.C-rater (Conceptual Rater) :

Est un moteur d'évaluation automatique de réponses courtes développé par Leacock et Chodorow en 2003 à l'ETS (Education Testing Service)[46][47], utilisé dans deux programmes d'évaluation à grande échelle : National Assessment for Educational Progress (NAEP) et à state wide assessment à Indiana [48];il se concentre sur le contenu ; basé sur les techniques NLP ;

C-Rater correspond aux fonctionnalités syntaxiques des réponses des étudiants (i.e., sujet, objet, et verbe) à un ensemble de réponses correctes. Cette méthode ignore l'approche des sacs de mots (BOW) pour faire la différence entre deux phrases .exemple "dog bites man" et "man bites dog" .[49]

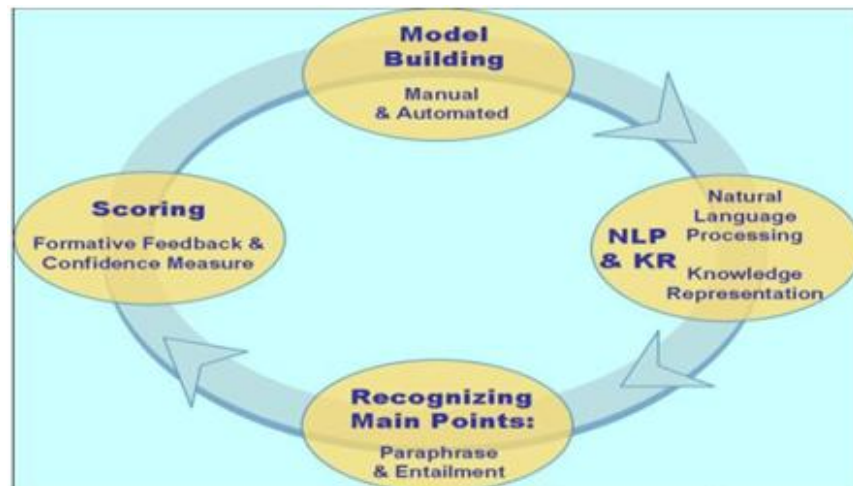


Figure 3.9:Les quatre étapes d'évaluations dans c-rater[40]

Grade	Subject	Question
8	Science 8	<p>Explain how you would design an experiment that would investigate the importance of light to plant growth.</p> <p>Include the type of organisms required, the control and variable, and the method of measuring results.</p>
8	Math 11	<p>A radio station wanted to determine the most popular type of music among those in the listening range of the station. Would sampling opinions at a Country Music Concert held in the listening area of the station be a good way to do this?</p> <p>Explain your answer.</p>
11	Reading comprehension	<p>Compare and contrast what Mama and Walter in A Raisin in the Sun believe to be the most important thing in life or what they “dream” of. Support your choice for each character with dialogue from the excerpt of the play.</p>
Collège	Database management	<p>Differentiate between logical and physical models.</p>

Tableau 3.6: Exemple des question avec C-rater [50]

C-rater est conçu pour les réponses bien définies dans des domaines spécifiques comme les math et les sciences exactes comme montrer dans le tableau.

### 5.IntelliMetric :

Est créé par « the American company Vantage » il est basé sur les approches de l'intelligence artificielle pour la mesure du style et le contenu . , utilise les fonctionnalités de l'intelligence artificielle , les méthodes statistiques et NLP [50][51][52][34]

Intellimetric fournit un score holistique global dans cinq domaines:

1. concentration et sens.
2. Organisation.
3. le contenu et le développement.
4. l'utilisation de la langue et le style.
5. mécanisme et convention.

### 6.Criterion :

Est un outil de correction et d'évaluation des essais sous une interface web. Il est composé de deux systèmes le E-rater et l'outil Critique WritingAnalysis[7];

Il permet de faire un feedback pour l'étudiant ; l'application a été téléchargée par deux cents personnes en décembre 2002.

### 7.SEAR(Schema Extract Analyze and Report) :

Est un logiciel développé par Christie en 1999 comme résultat de sa recherche de doctorat. Selon ce dernier , la correction des examens se base sur le contenu et le style.[9]

### 8.IEMS (Intelligent Essay Marking System)

Ce système est basé sur les techniques d'indexation du réseau de neurones [24] . Il se concentre sur la correction en fonction du style .Son but est : sommatif et formatif ; et appliqué sur les questions de type qualitatif exemple(biologie, histoire, psychologie ,ex ) .Il se base sur les NLP et les méthodes statistiques[34]

### 9.Apex Assessor

Ce système est intégré dans le système apex web. Il a été créé en 2000 dans le laboratoire des Sciences de L'Education dans l'Université Pierre-Mendé'

de France, si l'étudiant veut apprendre il va juste sélectionner un chapitre sous l'interface web d'apex. À la fin, l'étudiant est assisté par des questions à réponses ouvertes, Apex Assessor est le responsable de la sélection de ces questions, selon l'auteur, le but de ce système n'ai pas sommatif, mais aussi formatif, ce système est basé sur LSA

Le système est composé de trois modules :

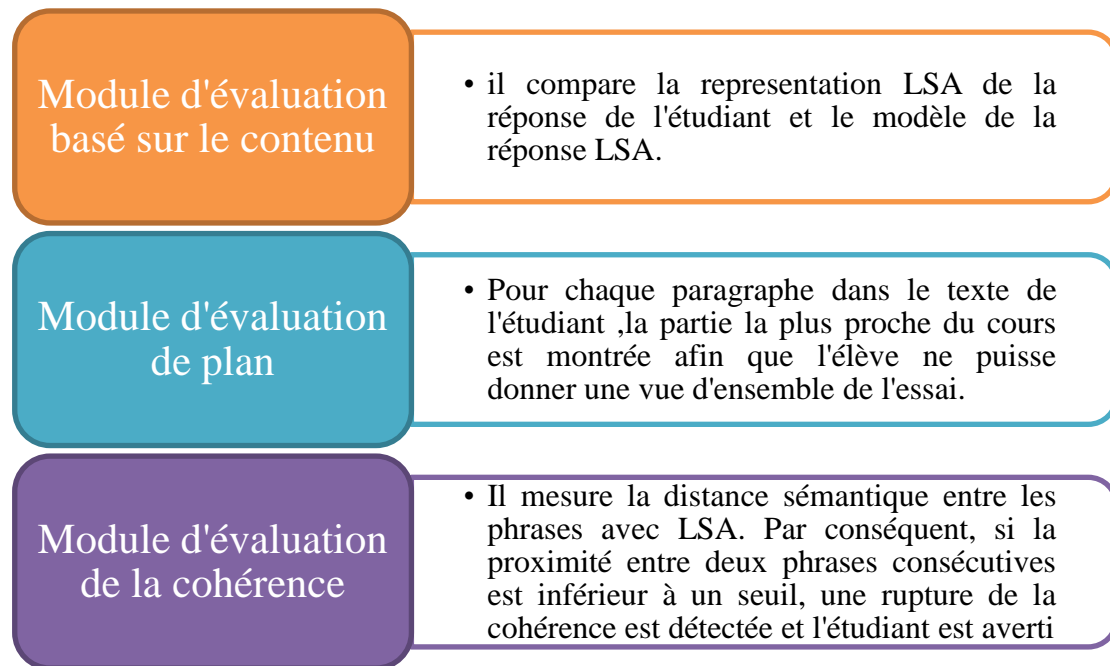


Figure 3.10: Les trois modules d'ApexAssessor

Ce système à un inconvénient : si les réponses s'ont courtes, cela implique un score élevé. Or cela est faux, alors il ont rajouter une contrainte sur le nombre des mot qui doit dépasser 300 mot.[29][42][9][53]

### 10. AutomatedText Marker (ATM)

Créé par Callear, Jerrams-Smith et Soh à l'université Portsmouth dans UK en 2001 , ce système est conçu pour donner deux notes : une pour le contenu et l'autre pour le style. Mais la note finale est donnée par le professeur [9].

Description du système :

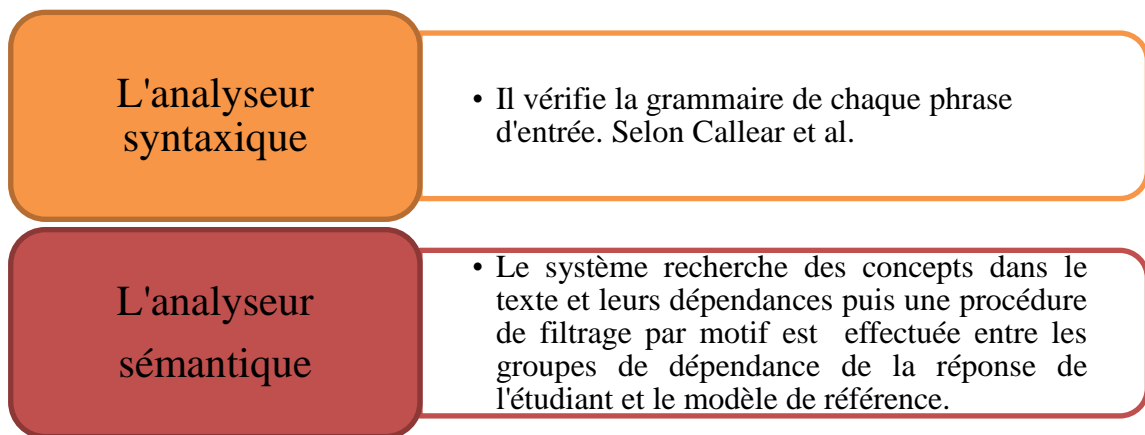


Figure 3.11:L'analyseur Syntaxique et sémantique d' Automatedtext marker

Est définie pour évaluer le concept dans le texte et ces dépendances pour donner une note une pour le style et l'autre pour le contenu.

#### 11. Bayesian Essay Test Scoring sYstem (BETSY) :

Développé par Lawrence M. Rudner dans le Collège Park de l'université Maryland en 2002.

Les deux modèles adaptés par la classification de texte sont :

MBM (Multivariate Bernoulli Model)et BM(Bernoulli Model) , parmi les avantage nous citons :

- ✓ il se peut t'appliquer sur les essais courtes.
- ✓ simple a implémenter
- ✓ ce système est téléchargeable et écrit en power basic.[50]



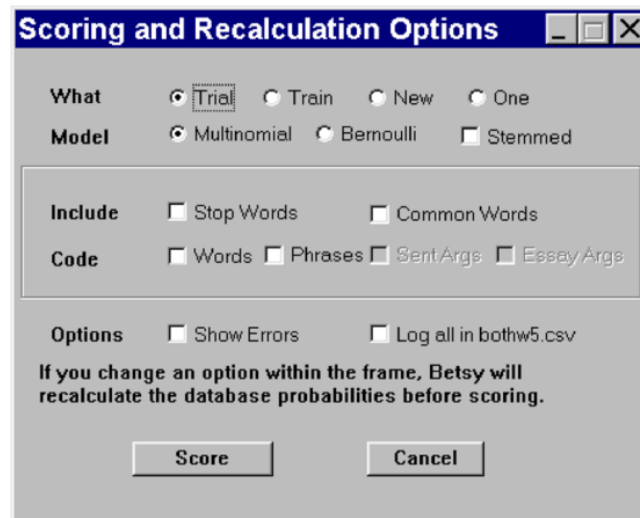


Figure 3.12: Scoring and recalculation window in BETSY[47]

## 12.AUTOMARK

C'est un logiciel développé par Mitchell, Russel, Broomhead, & Aldridge en 2002. Il utilise les méthodes d'analyse statistiques. NLP pour évaluer les réponses ouvertes.

Ce système était mis en développement en trois années, il a été utilisé dans les systèmes commerciaux de E-Learning (ExamOnline) ; Le système était incorporé par des modules de traitement d'Orthographe, de frappe, de syntaxe et de sémantique.[24]

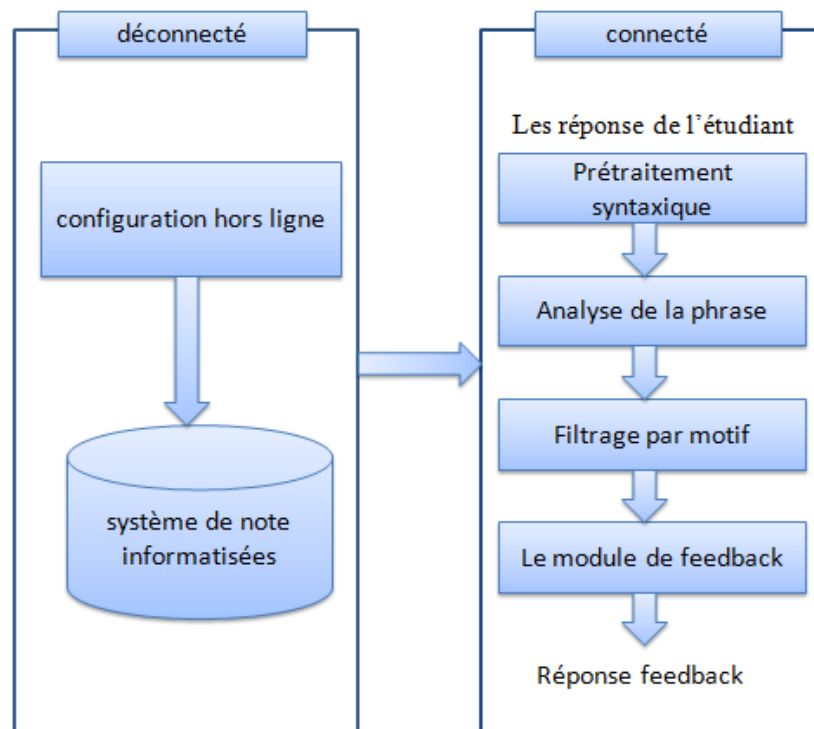


Figure 3.13 : Vue d'ensemble de tous les modules impliqués dans le processus d'évaluation effectué par Automark[10]

### 13.AutoTutor

Développé par Hasting et al en 1999 au Tutoring Research Group de l'université Memphis [53]. Il a été conçu comme un environnement de tutorat.

AutoTutor évite l'approche basée sur les sacs des mots et donc l'ordre des mots est important [49].

L'évaluation dans AutoTutor est le calcul de cosinus maximal entre la réponse des étudiants et l'ensemble du corpus de textes connexes[26].

### 14.PS-ME(Paperless School free -text Marking Engine)

A été conçu comme un composant intégré d'un système de gestion de l'apprentissage basé sur le Web ; la correction ne se fait pas en temps réel à cause des exigences de traitement.

Ce système applique les techniques de NLP pour évaluer les étudiants. Les réponses des étudiants sont téléchargées vers le serveur, chaque réponse est comparée à la réponse modèle ensuite il met une valeur pour chacune et ça sera la note pour l'étudiant par la suite [24].

### 15.carmelctc

CarmelTC est un module d'évaluation de texte libre ; développée à l'université de Pittsburgh par Rose, Roque, Bhembe et Vanlehn en 2002 .

Il utilise les techniques de classification de texte.

La méthodologie de CarmelTC est comme suit: d'abord, la réponse des étudiants est divisée en phrases et la technique bayésienne est utilisée pour trouver la probabilité de la présence de caractéristiques correctes qui représentent chaque phrase.

Un vecteur est généré en fonction de la présence ou l'absence de chaque caractéristique correcte et des règles pour identifier les catégories de phrases sur la base de ces vecteurs.[34]

Le système a été testé avec 126 essais en physique et les résultats sont à 90 % de précision 8 % fausse alèrte [9].

### 16.ERB(EvaluatingResponseswith Bleu)

ERB compare les réponses des étudiants avec le modèle référence en utilisant l'algorithme BLUE [21].Il calcule le pourcentage de chacun des textes de référence qui est couvert par le texte candidat . Un autre changement est que dans un environnement éducatif, il est généralement nécessaire d'utiliser une échelle standard pour les scores, comme entre 0 et 10, ou entre 0 et 100. Les scores de ERB sont toujours entre 0 et 1[54].

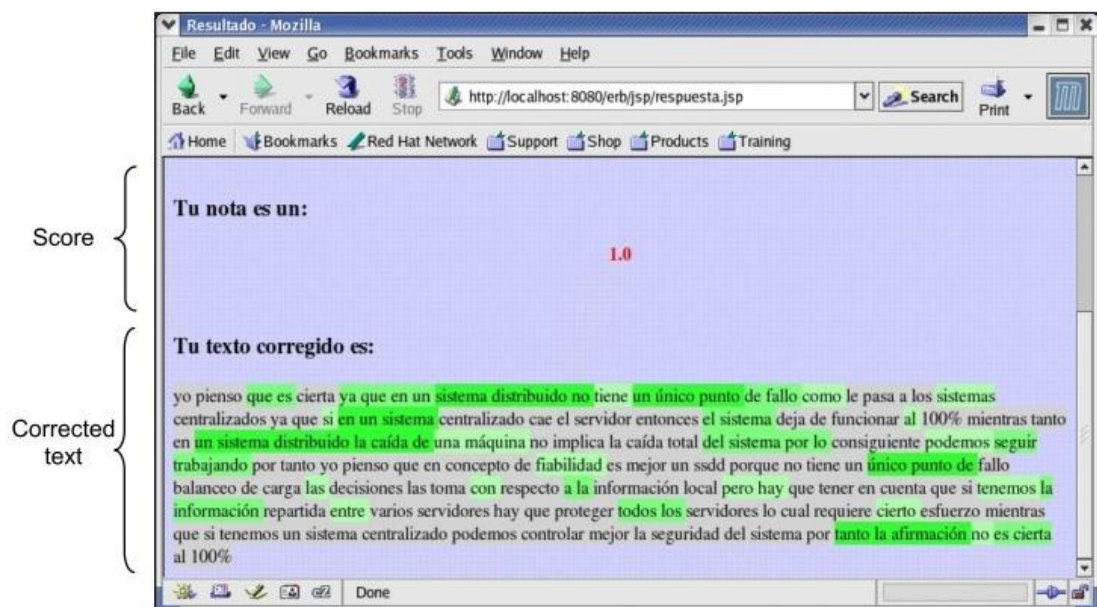


Figure 3.14:Logiciel qui utiliseERB.

### 17. Atenea (Willow)

ATENEA est un système d'évaluation assistée par ordinateur pour évaluer automatiquement les réponses courtes des étudiants[29]. Son nom est la traduction en espagnol d'Athéna ou Athéna, la déesse grecque de la sagesse. Atenea s'appuie sur la combinaison de modules de traitement du langage naturel (NLP) et des procédures d'évaluation basées sur les statistiques. Il a été codé en langage Java sous forme d'application autonome, mais il a aussi une version en ligne qui peut être accessible à partir de n'importe quel navigateur Web connecté ou non à internet. La figure 3.15 montre un aperçu de l'interface de la version en ligne de Atenea.



Figure 3.15: un aperçu de l'interface de la version en ligne de Atenea

Comme on peut le voir, l'entrée Atenea attend la réponse tapée par l'étudiant afin de la comparer avec un ensemble de réponses de référence (réponses idéales) rédigées par les enseignants, qui ont été préalablement stockées dans sa base de données. Il devrait y avoir au moins trois références d'enseignants par chaque réponse des élèves. Ces références peuvent aussi être prises à partir des réponses des meilleurs étudiants, afin d'avoir plus de solutions[54].

### 18. Summary Street

Franzke et Streeter (2006) à l'Université du Colorado à Boulder ont développé [23] Summary Street qui est un outil qui aide les étudiants à écrire de bons résumés en leur retournant un feedback à chaque écriture.

Tout d'abord, un étudiant est muni d'un des conseils généraux sur la façon de rédiger un résumé, l'étudiant choisi un sujet, lit le texte et écrit un résumé. Ensuite, l'algorithme LSA est appliqué pour donner une note globale [53].

Les étudiants ont la chance de faire une grande partie de l'écriture et de la révision avant que leurs enseignants voient le produit final[28].



Figure 3.16: Summary Street feedback display[31]

## 19. Texas

Un système Text-to-Text (Texas) a été introduit en Mohler et Mihalcea en 2009, le score est attribué selon une mesure de la similarité sémantique entre une réponse de l'élève et un modèle de réponse.

Le système a été appliqué à un ensemble de données informatique [14] qui contient 21 questions et 610 réponses d'élèves. Une version améliorée du système Texas a été introduite en Mohler et al. 2011.

L'ensemble de données utilisées dans cette version contenait 80 questions et 2273 réponses. [25]

	Logiciel	contenu	Style	organisation	Critique	Technique utilisé	Développé par	Année D'apparition	précision	Type de réponse	Domaine d'application	feedback
1	IEA	X	X	X		LSA et IA	Thomas K. Landauer	Fin 1960	76%[55]	Long	<ul style="list-style-type: none"> <li>○ psychology,</li> <li>○ medical</li> <li>○ business management</li> </ul>	X
2	E-Rater	X	X	X		Hybride <sup>1</sup>	Burstein, et al	1997	93%[56]	Long	(GMAT® AWA) <sup>2</sup>	
3	C-Rater	X	X	X		NLP	Leacock and Chodorow	2003	84%	courte	<ul style="list-style-type: none"> <li>○ Science</li> <li>○ Math</li> <li>○ Reading comprehension</li> <li>○ Database management</li> </ul>	X
4	Intellimetric	X	X			Hybride <sup>3</sup>	Elliot	1997	84%[52]	Long	GMAT	
5	Criterion 1.0 ,2.0	X	X	X	X	Hybride <sup>4</sup>	Educational Testing Service	2001,2003	82%[7]	Long	TOEFL GMAT programs are	X

<sup>1</sup>NLP et les méthodes statistique

<sup>2</sup>Graduate Management Admission Test® AnalyticalWritingAssessment

<sup>3</sup>NLP et les méthodes statistique

<sup>4</sup>NLP et les méthodes statistique

	Logiciel	contenu	Style	organisation	Critique	Technique utilisé	Développé par	Année D'apparition	précision	Type de réponse	Domaine d'application	feedback
6	SEAR	X	X			NLP	Christie	1999	30%	Long	Histoire	
7	IEMS	X	X			PM <sup>5</sup> +	(Ming, Mikhailov, &Kuan,	2000	80%	Long	plusieurs sujets selon le contenu.	X
8	Apex Assessor	X				LSA	Dessus Lemaire et Vernier	2000	59%	Long	la sociologie	X
9	ATM	X	X			Pattern matching	Callear etJerrams- Smith	2001	NA	Long ,courte	programmation Prolog, la psychologie et la biologie	
10	BETSY	X	X			Hybride <sup>6</sup>	Runder	2002	75%-80% <sup>7</sup>	Long	Toute tâche de classification de texte	
11	Automark	X				NLP	Mitchell, Russel, Broomhead, Aldridge	1999	93 et 96%	Long	produit de e-Learning commerciale	X

<sup>5</sup>Pattern matching+l'algorithme d'indexation par clustring

<sup>6</sup>Les méthodes de classification de texte bayesienne et Statistique

<sup>7</sup>80% avec le model BM[p][r]Et 75% avec le model MBM[r]

	Logiciel	contenu	Style	organisation	Critique	Technique utilisé	Développé par	Année D'apparition	précision	Type de réponse	Domaine d'application	feedback
12	Autotutor	X				LSA	Wiemer-Hastings et al.	1999	28 - 36%[57]	courte	informatique.	X
13	PS-ME	X				NLP	Mason and Grove-Stephenson[9]	2002	NA	Long	National Curriculum Grade and GCSE exam	X
14	CARMEITC	X				TCT <sup>8</sup>	Rose, Roque, Bhembe and Vanlehn	2002	90%[9]	court	physique	
15	ERB	X				BLUE		2004	82%[54]	Long		X
16	ATENEA	X				Hybride <sup>9</sup>	Perez et al.,	2004		courte		
17	Summary Street	X				LSA	Kintsch et al	2000			Ecrire des résumés	
18	Texas .	X				Similarité syntaxiques		2009		courte		

Tableau 3.7:Tableaux récapitulatif sur les logiciel Académique

<sup>8</sup> TCT : Techniques de catégorisation de texte

<sup>9</sup> NLP et les méthodes statistiques



### c. Evaluation des solutions existantes:

Les solutions existantes, qu'elles soient commerciales ou académiques sont caractérisées par les éléments fondamentaux suivants:

#### 1) La forte dépendance d'une langue particulière:

Le quasi totalité des approches dont les résultats reportés étaient satisfaisant ne sont en fait valable que pour une seule langue bien déterminée, plus exactement pour l'anglais.

Un nombre très restreint de solutions supporte seulement une des langues suivantes (allemand, espagnol, français).

Cette dépendance est en fait un frein majeur de l'outil pour son exploitation plus vaste.

Pour essayer de pallier cette limitation, certains travaux ont essayé d'adapter certains outils, valable pour la langue anglaise à une autre langue spécifiques.

Deux approches ont été expérimentées:

- La réécriture de tous les modules liés à langue [49] cette tentative n'as pas été fructueuse car elle s'est en réalité soldée par la refonte de la majorité des modules liés à la langue. C'est le cas du logiciel CoMiC-DE [49] qui n'est en fait que la refonte presque totale de CoMiC-EN qui est la version anglaise originale.

- L'utilisation de la traduction automatique d'une langue vers la langue supportée par l'outil. C'est dans ce contexte que se placent les travaux de [58] [59] qui ont utilisé un traducteur espagnol – anglais pour le premier et arabe anglais pour le deuxième. Le constat n'est pas aussi fameux que pour la première approche à cause de la qualité du processus de traduction qui n'est pas toujours excellente. Les imprécisions dues à la traduction ont en fait un impact négatif très sérieux sur les résultats de l'évaluation.

## 2) La forte dépendance de la matière dans lequel l'étudiant est évalué [60]

La totalité des solutions proposées sont spécifiques à un cours particulier. Les tests réels et les résultats jugés par leurs auteurs comme excellent ont été obtenu lors d'examens d'étudiants dans un cours particulier, plus encore sur un aspect particulier du cours.

Ainsi, en programmation informatique les tests réalisés concernent plutôt les aspects généraux tels que donner une définition et non pas des aspects relatifs à la conception à la mise sur pied de solution.

## 3) La dépendance du type de question :

La forme dans laquelle une question est posée et la forme de la réponse attendue sont aussi particulières à chaque outil.

Les outils testés sont en fait valables pour certains types de question et ne sont pas efficaces pour d'autres types de question.

A titre d'exemple certains outils sont reportés comme efficaces pour évaluer des définitions et non pas pour évaluer des questions de type argumenter justification de choix, énumération et description[61].

## 4) La qualité du corpus utilisé :

La qualité des données utilisées est en fait fondamentale. Cette qualité dépend de la mise à jour du corpus, et des valeurs affectées par les experts aux diverses réponses contenues dans le corpus, la note étant en général la moyenne de la note des experts.

Pour diverses raisons, notamment la subjectivité, il a été remarqué que la construction de corpus est en fait un défi.

Dans certains rapports de recherche il a été relevé que dans certains corpus la note donnée par les experts diffère de 4 points sur une échelle de valeur de 5 [62].

La mise à jour du corpus est un autre défi pour permettre d'adapter le corpus aux nouveautés dans le cours et aux nouvelles formes d'évaluation.

Dans les pays en voie de développement, la construction du corpus semble être utopique, du fait que l'évaluation automatique n'est même pas utilisée. Les sous entendus dans une question telle que « selon ce qui a été vu en cours », ou les exercices fait en cours, sont très difficiles à faire supporter efficacement dans un corpus, car ces aspects sont vraiment particuliers à chaque enseignant.

5) Les rejets de notes :

Dans divers domaines de l'informatique telle que la traduction automatique, la recherche d'information, le résultat des approches de similarité de textes utilisés sont exploitables même s'ils ne sont pas de haute qualité. S'ils ne sont pas de la qualité attendue, ils peuvent au moins être exploités comme point de départ par une amélioration et ajustement manuel. Contrairement à ces domaines, les résultats de l'évaluation automatique est un jugement que l'examiné peut accepter ou rejeter et demander une évaluation manuelle.

Les diverses mesures relatives à l'efficacité des diverses solutions n'adresse pas cet aspect. La minimisation à l'extrême du taux de rejet est fondamentale pour la réussite d'une approche d'évaluation en ligne. Cette minimisation permettrait des périodes de consultations très brèves et très rigoureuses.

6) Contenu dynamique des questions :

Un autre aspect de taille qui n'as pas été adressé ni par les techniques de similarité ni par la construction de corpus.

Cet aspect découle du fait que dans certaines réponses, certains mots sont plutôt spécifiques à la personne examiné. A titre d'exemple, dans une question basée sur une image, la première question dit ; donner un nom au petit garçon, la deuxième que fait le petit garçon. La réponse peut être « le garçon joue au ballon » ou « Ali joue au ballon » ou « Omar Joue au ballon ». En réalité la représentation de partie dynamique dans une réponse ne peut pas être supportée par les techniques actuelles de spécification des réponses dans un corpus. Une forme de langage de spécification de réponse pourrait être nécessaire. Le contenu dynamique est souvent rencontré dans les examens portant sur les langages de programmation de systèmes informatiques.

### 7) Les indications de l'enseignant :

En pratique, un examen est en réalité très dépendant de l'enseignant responsable du cours. C'est en fait lui qui est au courant de ce qui a été fait en cours, des sous entendus dans le cours, des exercices vu en cours.

De plus c'est lui qui est en réalité responsable de l'importance à donner à chaque élément de réponse.

À titre d'exemple, dans les premiers tests l'enseignant peut ignorer certaines erreurs, qui pour lui sont naturelles en début du cours.

Par la suite ces erreurs si elles se répètent dans une deuxième période pourraient pénaliser la réponse, par un retrait de pourcentage que seul l'enseignant est capable d'indiquer.

Dans une 3<sup>ème</sup> phase, la même erreur qui était ignoré au départ entraîne l'annulation d'une réponse même si elle contient un nombre important d'éléments justes.

C'est aussi l'enseignant qui indiquera si dans une réponse on peut ne pas considérer les accents ou les homophones et quel serait la pénalité à appliquer si une erreur d'orthographe était commise.

## CHAPITRE 4 L'ÉVALUATION AUTOMATIQUE BASÉE SUR LES INDICATIONS DE L'ENSEIGNANT

### 4.Approche Définie

#### Introduction

Dans ce chapitre nous allons d'une part fixer les objectifs de notre approche pour l'évaluation rapide des réponses et d'autre part illustrer les diverses parties de notre solution.

La solution que nous proposons se base fondamentalement sur un corpus qui est construit automatiquement et qui répond aux aspirations de l'enseignant. La description des réponses de l'enseignant est réalisée en utilisant un langage de spécification de réponses courtes qui reporte en plus des éléments de la réponse courte donnée par l'enseignant les diverses indications de l'enseignant. La construction du corpus enseignant est faite en compilant le contenu de chaque réponse courte donnée par l'enseignant. Le corpus est en fait spécifique à chaque réponse courte donnée par l'enseignant et ainsi plusieurs corpus peuvent être construits pour une question pour laquelle l'enseignant spécifie plus d'une réponse

Certaines indications de l'enseignant peuvent aussi être appliquées à la réponse de l'étudiant. Cette application se traduit par la création d'un jeu de réponses-étudiant qui sera par la suite comparées aux réponses contenu dans le corpus enseignant.

#### Les objectifs :

Les objectifs de notre approche se basent sur les faiblesses et limites des approches actuelles. Le point le plus faible de ces approches est qu'elles ne donnent pas une large importance à l'enseignant qui prépare les questions.

Un autre point faible des approches actuelles est qu'elle ne considère pas la réalité des personnes examinées.

En général, notamment dans les disciplines telles que l'Informatique, l'électronique, le génie mécanique et en général les disciplines technologiques, les personnes examinées ne sont pas des experts en orthographe et en langue.

Cette situation est plus catastrophique pour les personnes poursuivant des études de licence, technicien supérieur ou technicien.

Nous rappelons que l'évaluation vise les étudiants excellents et aussi moyens.

En général, l'évaluation a pour objectif de s'assurer que certaines connaissances ont été acquises par l'étudiant et non pas d'évaluer l'étudiant sur le style de son langage et son écriture sans fautes d'orthographe.

Maintenant, comment devra se comporter le processus d'évaluation automatique sans indication explicite de l'enseignant qui prépare les évaluations, le processus sera inefficace.

C'est en réalité à l'enseignant de la matière d'indiquer lors de la préparation des question/réponses ce qui est important à considérer dans l'évaluation automatique et ce qui ne l'est pas.

Notre objectif est la réalisation d'un outil d'évaluation automatique indépendant de la matière enseignée. Cet outil devra être capable d'évaluer des réponses dans le cas de matière informatique comme la programmation dans un langage de programmation, les réseaux, la compilation ou dans le contexte d'autre matière d'autres disciplines technologiques et des sciences humaines et médicales.

L'outil devra donc permettre à l'enseignant de n'importe quelle matière de faire ressortir ce qui est important à considérer et ce qui ne l'est pas.

#### Détermination des éléments à considérer par l'enseignant :

Dans ce qui suit, nous essayerons de déterminer les situations à partir desquelles nous déterminerons les aspects les plus importants que les enseignants doivent considérer et classer comme important ou non lors de la préparation des questions. Ces situations concernent surtout l'aspect erreur d'orthographe et d'inattention.

- Les espaces : Dans une situation ordinaire, un seul espace est inséré entre les mots lors de l'écriture d'une réponse. C'est ce que fait l'enseignant en définissant la réponse type. Lors des réponses, il est fort probable que l'étudiant ne suive pas ce style d'écriture et ajoute des espaces supplémentaires, au moins par inattention.

- Les accents : souvent celui qui répond ne peut déterminer s'il faut un accent et si cet accent est grave, aigu ou circonflexe dans le cas de la langue française
- Les expressions ayant un fond sonore similaire ou homophones : À titre d'exemple le mot « aire » pourrait être écrit « ère ». Le mot « diffuser » pourrait être écrit « diffuzer ».
- La conjugaison : un verbe qui devrait être conjugué au présent est alors conjugué au passé ou au futur ou très mal conjugué
- Le doublement de mot ou de lettre. Ce phénomène apparaît souvent et est du souvent à une erreur d'inattention
- Utilisation de mauvais déterminant

En plus de ces éléments il est fort probable que selon la nature de la matière enseignée, l'enseignant pourrait introduire de nouveaux éléments permettant de rendre l'évaluation automatique plus efficace. A titre d'exemple, si nous considérons le module de programmation dans un langage évolué, l'enseignant pourra accepter des instructions qui ne respectent pas très rigoureusement le langage de programmation. A titre d'exemple, il pourra indiquer que le point-virgule à la fin de l'écriture d'une instruction est optionnel et qu'un bloc, commençant par une accolade ouvrante et se terminant par une accolade fermante et contenant une instruction est considéré comme correct. Ainsi il serait intéressant de permettre à l'évaluation automatique des réponses courtes de s'adapter aux spécificités des matières. Ainsi chaque matière donnera lieu à un style particulier de préparation des questions et à l'introduction d'une fonctionnalité de gestion de style pour ajouter, mettre à jour ou supprimer des styles.

Dans ce qui suit nous présentons comment divers éléments seront considéré par l'évaluation automatique.

#### Le système de graduation automatique définie

Dans la phase préparation des questions, l'enseignant introduit un ensemble des réponses références (RR) et un ensemble d'indications basée sur les préférences de l'enseignant (IA).

Cette étape donne un ensemble important de réponses proposées complètes ou incomplètes définie un corpus avec la correspondance de grade de chacun.

On va l'appeler ce dernier Grading Engine Reference Answer (GERA) ; le même processus est appliqué à la réponse étudiant et on va l'appeler Grading Engine Student Answer(GESA).

Le processus d'évaluation consiste à comparer les deux fichiers générés (GERA)et (GESA) et permettre d'avoir une note pour l'étudiant d'après deux méthodes différentes :

- Associé à l'étudiant la note de la réponse de GERA correspondante, si la similitude calculée est supérieure ou égale au Seuil de similarité.
- Signaler la réponse comme une réponse qui doit être examinée manuellement par l'enseignant si le facteur de similarité est supérieur ou égal au Manual GradingThreshold similarity factor.

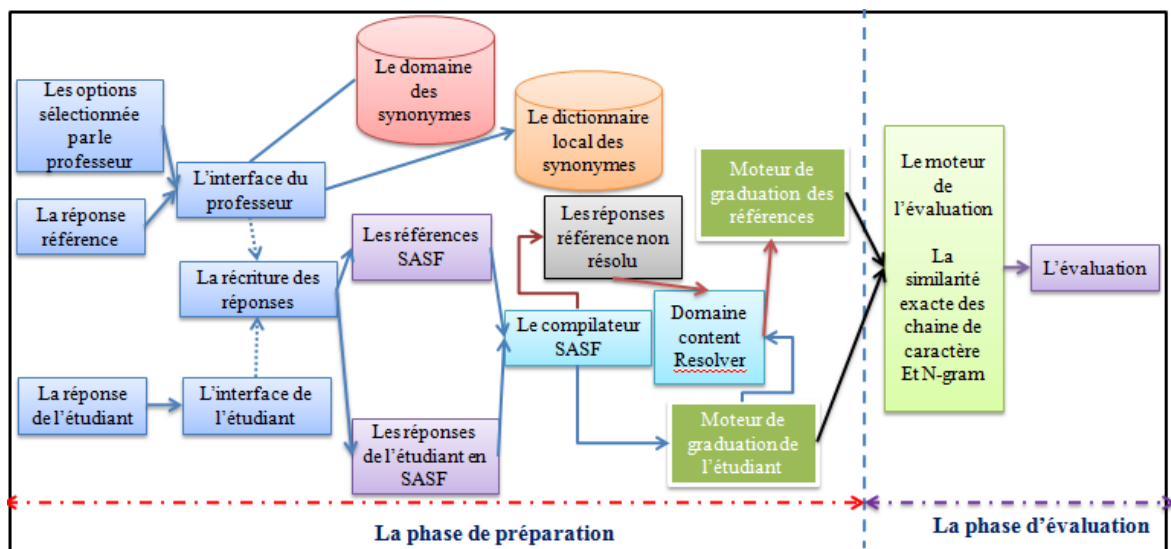


Figure 4.1:le système de graduation automatique défini

### 1. Spécification des réponses référence et les indications enseignants :

Une fois les réponses référence sont complètement définies, l'enseignant commence l'association d'un certain nombre d'indications avec chaque réponse.



Grâce aux indications, l'enseignant tente de prédire les diverses formes de réponses des étudiants qui peuvent être considérées comme correctes ou partiellement correct.

Grâce à cette tâche de prédiction, ce qui peut prendre beaucoup de temps, l'enseignant va essayer de préparer le meilleur environnement pour AGS (AutomaticGrading System).

L'utilisation de balises XML et expressions régulières, même par un professeur des sciences informatiques n'est pas facile du tout.

Comme d'autres systèmes basés sur XML, au lieu de l'utilisation directe de notation XML et d'expression régulière complexe, les enseignants ont fournis dans AGS avec une interface simple pour préciser leurs indications: SASL.

La spécification des réponses de référence et leurs indications associées sont réalisées en utilisant un langage basé sur XML appelé Short AnswerLangage Specification (SASL) .

Les indications des enseignants sont spécifiés en XMLtags ou expression régulière méta-caractères.

L'objectif principal de la spécification des réponses références accompagnées d'un certain nombre d'indications est de guider efficacement la préparation et les phases de classement.

L'annotation commune SASF (tableau 4.1) a été introduite pour faciliter la spécification de diverses indications simples directement dans le texte de la réponse de référence.

Teneur en réponse courte dynamique et exemples de mot clés de ces indications de l'enseignant simples (figure 4.2).

SASF Common Annotations	Role	XML tag
<code>\$list{}</code>		
<code>\$alphaword</code>	nom alphabétique	<DPARTcontent="alphaonly", style="allowcase", minsize = 1, penalty="5" /DPART>
<code>\$alphanumword</code>	nom alphanumérique	<DPART content="alpha", style="allowcase", minsize = 1, firstchar="alpha" /DPART>
<code>\$alphaword</code>	Nom alphanumérique avec le premier caractère en majuscule	Yes
<code>\$alphanumword</code>	Nom alphanumérique avec le premier caractère en minuscules	Yes
<code>\$alphanumword</code>	Nom alphanumérique avec majuscule uniquement	Yes
<code>\$integer</code>	entier	
<code>\$integerrange(min, max), \$integerrangex(min,max)</code>	intervalle	
<code>\$real.xx.err</code>	Valeur réelle	
<code>\$acronymword</code>	Designation Acronyme	
<code>%%thekeyword</code>	désignation de mots-clés	<KEYW value=
<code>\$word[start=3,maxcar=10]</code>	Nombre de caractères importants	Yes
<code>#!synonymlookup (disctionnary)</code>		

Tableau 4.1: SASL

La question:
According the java source code writing style, write only the first line of the constructor used to initialize the last name and the firstname of an object instantiated by the following Personne class.
<pre>public class Personne {     String matriculate, nom, prenom ; }</pre>
La réponses référence
%%public %%Personne (String \$alpha-id1, String \$alpha-id2) %%{

Figure 4.2:Un exemple de réponse écrite en utilisant le contenu et les mots clés dynamique

Les indications de l'enseignant sont spécifiées en utilisant l'interface grading policy interface (GPI) .

Chaque réponse a une note maximale désignée. La nécessité de considérer ou d'ignorer l'ordre des mots-clés ; la sensibilité à la casse, stop words; les accents des caractères, l'homophonie, la synonymie etc.

La table suivante (tableaux 4.2) reporte les différentes classes et les fonctions supportées dans notre outil.

Certaines indications comme la considération du synonyme ou de l'homophone sont associées à un dictionnaire interne sous le contrôle de l'enseignant.

Finalement Lorsqu' on active une propriété on les associe à un facteur de pénalité.

Classe indication	Son Role
Grading controller	-Note maximale -seuil - outils des similarités
Multiple Line SubController	-Multi ligne -maximum des lignes -nombre de ligne
Preprocessing Controller	Appliqué aux deux réponse(enseignant et étudiant) par exemple : -enlever stop word -remplacer stop word -remplacer les caractères
Stop word subcontroller:	-definition des stopword -ignoré ou remplacer les stop word
Character Controller	-sensibilité à la casse. -ignoré les répétitions de caractere -remplacement d'un caractere par une autre -ignoréles accent
Keywords controller	-Soulignez mot important dans une réponse (mots clés) -ignorer / utiliser les mots-clés -Accepter une réponse basée sur les mot-clé,
Expression subController	-Définir l'opérateur
Synonym subcontroller	-construire un dictionnaire interne
Spelling controller	-Fixer et permettre la correction d'erreur en utilisant des références déterminées (réponses de référence des enseignants, le dictionnaire locale)

Tableau 4.2 : GPC SASF paramètre classes part1/2

Classe indication	Son Role
Homophone controller	-Définir l'homophone considéré comme similaire à celui proposé dans le modèle de l'enseignant
Extra Answer controller	-Activer la détection des mots qui ne font pas partie de la réponse (pas dans le dictionnaire local), -la répétition de mots clés -duplication des réponses - retirer mot supplémentaire,
Dynamic part controller	-définir des mots qui ne sont pas utilisés comme mots-clés, contrôleur de style (mots non autorisés, le caractère utilisé, le nombre de chiffres de précision),
Guideline subcontroller	Le texte à afficher pour guider la réponse
SASF subController	redéfinir le jeu de caractères et de la politique pour les constructions définies par l'utilisateur,
Grading threshold	Indiquer le seuil de similarité nécessaire pour accepter pour examen manuel

Tableau 4.2 : GPC SASF paramètre classes part 2/2

### Le Compilateur SASF

Les réponses des étudiants ; la réponse référence et les indications du professeur spécifiées utilisant SASL annotation sont traduits par le moteur de réécriture de réponse (ARE)et transmis en description SASL complet .

Une SASF Description de réponse contient la réponse de texte original et les indications diverses qui doivent être appliquées dans le processus de construction du moteur de classement et les réponses des étudiants (figure 4.3).

Les indications spécifiées ne sont pas appliquées de manière systématique aux réponses référence et aux réponses étudiant. Certains d'entre eux, comme "accepter mot manquant" (facteur de pénalité est inférieur à 100% pour les mots-clés obligatoire) cibler la réponse des enseignants seulement.

	les indications SASL	Les valeurs par et les pénalités associées par défaut dans GPI		Les options sélectionnées par l'enseignant	
01	Preprocess: Pack text	ON	0%	ON	0%
02	Extra Char / Word (Student answer only)	ON	Shared word 30%, char 10%	ON	word 20% char 5%
03	Keyword Order	ON	100%	ON	100%
04	Keyword missed (public, )	ON, all	Shared (100%) :public = 33%, Personne = 33%, { = 33%)	ON, select	Not shared: Public=10% { = 5% Personne 100%
05	Misspell	ON, Local	Shared: number of misspell/100	ON	5%
06	Dynamic Content style	ON, all	0%	On, all, not cumulative	5%
07	Key char in Keywords	OFF, all	0%	ON, select: Public (S3, M5) Personne (S4, M9)	Public : 5%, Personne: 5%
08	Character Repetition	OFF	0%	ON	5%
09	Case sensitivity	ON, all	0%	OFF, all	10%

Tableau 4.3:Exemple des indications de la question mentionnée dans la figure 4.2, spécifiée utilisant GPI

## 2. Génération d'un corpus :

Afin de générer le corpus des réponses possibles on a utilisé un générateur de cas possibles ; suivant l'indication de l'enseignant ; prenons l'exemple d'acceptation des mots vides ou espaces.

Algorithme :

1. Récupération de la réponse enseignant ;
2. Remplacement des espaces avec le code correspondant ;
3. Génération d'une matrice telle que :
  - a) Nombre de colonne = nombre des erreurs à récupérer soit « n » ;
  - b) Nombre de ligne = 2n ;
4. Remplissage de la matrice avec les 2n cas possible en (0 et 1) ;
5. la jointure de chaque cas avec la réponse du professeur ;
6. calcul de pénalité en fonction nombre de 1.(le professeur à le choix de modifier la valeur) ;
7. remplacement du « 0 » par le vide et les « 1 » par l'erreur correspondante.

Ce tableau présente les résultats désirés :

Id	Les réponses générées par l'outil	Pénalité associée	Numéro associée à l'indication
0	public Personne ( String \$alpha-id1,String \$alpha-id2 ){	0	Correcte
1	public Personne(String \$alpha-id1,String \$alpha-id2){	0	Correcte
2	public Personne(String \$alpha-id1,String \$alpha-id2)	5	04
3	pub Personne(String \$alpha-id1,String \$alpha-id2){	5	05
4	public Pers(String \$alpha-id1,String \$alpha-id2){	5	05
5	public Persone(String \$alpha-id1,String \$alpha-id2){	5	05
6	Personne(String \$alpha-id1,String \$alpha-id2){	10	04
7	pub Pers(String \$alpha-id1,String \$alpha-id2){	10	05+05
8	pub Personne(String \$alpha-id1,String \$alpha-id2)	10	05+04
9	public Per(String \$alpha-id1,String \$alpha-id2)	10	05+04
10	public Persone(String \$alpha-id1,String \$alpha-id2)	10	05+04
11	pub Persone(String \$alpha-id1,String \$alpha-id2){	10	05+05
12	public personne(string \$alpha-id1,string \$alpha-id2){	10	09
13	Personne(String \$alpha-id1,String \$alpha-id2)	15	04+04
14	Pers(String \$alpha-id1,String \$alpha-id2){	15	04+05
15	pub Pers(String \$alpha-id1,String \$alpha-id2)	15	05+05
16	Persone(String \$alpha-id1,String \$alpha-id2){	15	04+05
17	pub Persone(String \$alpha-id1,String \$alpha-id2)	15	05+05
18	public personne(string \$alpha-id1,string \$alpha-id2)	15	09+04
19	pub personne(string \$alpha-id1,string \$alpha-id2){	15	05+09
20	public pers(string \$alpha-id1,string \$alpha-id2){	15	05+09
21	public persone(string \$alpha-id1,string \$alpha-id2){	15	05+09
22	Pers(String \$alpha-id1,String \$alpha-id2)	20	04+05+04
23	Persone(String \$alpha-id1,String \$alpha-id2)	20	04+05+04
24	personne(string \$alpha-id1,string \$alpha-id2){	20	04+09
25	pub pers(string \$alpha-id1,string \$alpha-id2){	20	05+05+09
26	pub personne(string \$alpha-id1,string \$alpha-id2)	20	05+09+04
27	public per(string \$alpha-id1,string \$alpha-id2)	20	05+09+04
28	public persone(string \$alpha-id1,string \$alpha-id2)	20	05+09+04
29	pub persone(string \$alpha-id1,string \$alpha-id2){	20	05+05+09
30	personne(string \$alpha-id1,string \$alpha-id2)	25	04+09+04
31	pers(string \$alpha-id1,string \$alpha-id2){	25	04+05+09
32	pub pers(string \$alpha-id1,string \$alpha-id2)	25	05+05+09+04
33	persone(string \$alpha-id1,string \$alpha-id2){	25	04+05+09
34	pub persone(string \$alpha-id1,string \$alpha-id2)	25	05+05+04
35	pers(string \$alpha-id1,string \$alpha-id2)	30	04+05+04
36	persone(string \$alpha-id1,string \$alpha-id2)	30	04+05+09+04

Tableau 4.4: le dictionnaire généré et les pénalités associées(GERA)

La principale tâche du compilateur SASF est la production d'un ensemble ordonné de réponses références appelées la réponse référence classifiée (GERA)

(tableau 4.4) et un ensemble ordonné de réponses étudiants appelé les réponses classifiées (GESA) (Tableau 4.5).

Pour réaliser cette tâche, la SASL localise la prochaine indication à réaliser en utilisant la classe d'indication de priorité.

Une fois localisée, l'indication est appliquée à la réponse originale et aussi aux réponses produites dans l'étape de prétraitement précédente. Lorsqu'il est appliqué sans aucune pénalité, la réponse produite remplace l'original dans les phases de graduation.

En raison de l'application successive de leurs indications et pénalités, la première réponse produite est associée à la note maximale et la dernière réponse produite est associée à la note minimale.

Quand une réponse de référence contient un contenu dynamique, le compilateur SASF produit un ensemble des réponses non résolues de référence (tableau 4.4) qui doit être soumis à un processus de résolution afin de produire GERA.

Le processus de résolution du contenu dynamique, appelé Dynamic Content réponse Resolver (DCRR), est activé quand une réponse étudiant est compilée.

La DCRR essaie de localiser dans la réponse de l'étudiant le contenu correspondant. Une fois localisée, le résolveur vérifie si les propriétés de contenu dynamique diverses sont présentes ou pas, et si nécessaire on applique la pénalité correspondante.

Dans le cas de l'exemple rapporté par (tableaux 4.4 et 4.5), la chaîne « nom » et la chaîne « prénom » va remplacer le contenu dynamique \$ alpha-id1 et alpha-id2.

La DCRR doit assurer que la chaîne « nom » est différent de la chaîne « prénom » et tous ont été écrits en utilisant seulement des caractères alphabétiques minuscules. Après l'application de processus prétraitement (0% de pénalité) et extra char (5%) et extra word detection (20% de pénalité) on obtient

les tableaux des réponses générés de l'étudiant en appliquant les paramètres associées par l'enseignant (tableau 4.5),

GESA specific penalties: 30% (Added to basic GERA answer penalties)	
Style= 5%	: personne, Nom
Extra Word = 20%	: class
Extra Char = 5%	: ;

N°	Les réponses générées	Pénalité
0	class publique personne ( string Nom, string prenom );	0
1	class publique personne(string Nom,stringprenom);	0
2	publique personne(string Nom,stringprenom)	0
3	publique pers(string Nom, string prenom)	5
4	pub personne(string Nom, string prenom)	5
5	publique persone(string Nom,stringprenom)	5
6	pub pers(string Nom, string prenom)	10
7	pub persone(string Nom, string prenom)	10
8	publique personne(string nom,stringprenom)	10
9	publique pers(string nom, string prenom)	15
10	pub personne(string nom, string prenom)	15
11	publique persone(string nom,stringprenom)	15
12	pub pers(string nom, string prenom)	20
13	pub persone(string nom, string prenom)	20

Tableau 4.5: GESA

#### La phase de l'évaluation :

Le moteur d'évaluation peut être fourni avec un certain nombre d'outils de comparaison telle que l'utilisation des algorithmes de similarité qui peuvent être appliqués successivement, selon les indications de l'enseignant ou partiellement.

Le processus considère une à une les réponses du GESA ordonnées, selon leur position dans l'ensemble. Pour chaque réponse de l'étudiant considéré.

Les outils de comparaison tentent de localiser la réponse la plus proche dans le GERA

Lorsque la comparaison donne un résultat supérieur au seuil du classement (MGT), la réponse est étiquetée pour être soumise à un examen manuel par le professeur.



Si la comparaison donne une valeur supérieure ou égale au seuil de classement automatique (AGT), la réponse est classée avec le grade de la réponse de GERA.

Une fois terminée, le processus de comparaison ne conserve que la réponse satisfaisant le seuil automatique de classement (l'AGT préfixé).

Parmi ces réponses, le moteur de classement choisi la réponse de GERA ayant la plus petite position et attribue sa note à la réponse de l'élève.

Après l'application des deux algorithmes (lavenshtein et n-gramme) ensuite on va choisir le grade maximum entre les deux.

Cette réponse est celle ayant la plus petite peine parmi les réponses gardées.

## CHAPITRE 5 EVALUATION DE L'APPROCHE

### 1. Introduction :

L'évaluation de l'approche s'est faite par la construction d'un module qui a été intégré à la plateforme Moodle et qui a été expérimenté partiellement en grandeur nature. Le choix de Moodle est en fait du à deux objectifs :

- Profiter des autres aspects que fourni moodle, notamment la construction facile de tests, le test en ligne et tous les autres aspects liés à la gestion des apprenant et la sécurité du système. ;

- Voir réellement si l'approche préconisée par Moodle pour la construction de plugin est efficace pour la construction de module qui serait complexe comme le notre.

Dans ce chapitre nous allons montrer le contexte dans lequel nous avons réalisé notre module. Ensuite nous présenterons les options implémentées pour la réalisation et les divers tests.

### 2. Environnement de réalisation

L'outil a été réalisé dans le contexte de la plateforme Moodle. Il représente ainsi une extension de la plateforme Moodle.

C'est ce qui est souvent appelé par plugin. Le langage utilisé est celui utilisé par Moodle, c'est-à-dire le langage PHP.

La conception et réalisation d'un plugin Moodle suit une démarche bien précise qu'il est nécessaire de maîtriser.

La plus grande complexité que nous avons rencontré c'est le fait que Moodle soit réalisé complètement en orienté objet et que ce dernier soit composé d'un nombre très important de fichiers interdépendants.

Ainsi malgré la présence d'une méthode claire de mise en place de nouveau module pour Moodle, la mise en place réelle d'un nouveau module est

très délicate et nécessite un temps important pour la compréhension et la mise en œuvre.

Dans ce qui suit nous allons illustrer comment un module peut être créé sous Moodle. Par la suite nous exposerons, par les exemples et les diverses fonctionnalités de notre outil

### 3. Modules et plugins de « Moodle » :

Il existe plusieurs catégories de plugins telle que:

- Thème: Pour changer le look des pages du cours.
- Format de cours: Pour changer la structure (layout) des pages du cours.
- Activités: La principale activité des étudiants dans un cours.....

#### Création d'un nouveau module Moodle :

Pour ajouter un nouveau module dans la plateforme Moodle on a suivi les étapes d'installation d'un plugin de la communauté Moodle.

1/ téléchargement du fichier du plugin vierge compressé (NEWMODULE.zip).

2/ Décompression de l'archive et lecture du fichier README.txt pour suivre les étapes d'installation.

3/ Renommer le dossier NEWMODULE (dans ce cas on a donné à notre plugin le nom 'nvtype'). Le nom du dossier du module doit être en minuscules et ne contient que des caractères alphabétiques.

4/ Modification de tous les fichiers dans le répertoire et ses sous-répertoires et changement de toutes les instances de la chaîne "newmodule" par 'nvtype'.

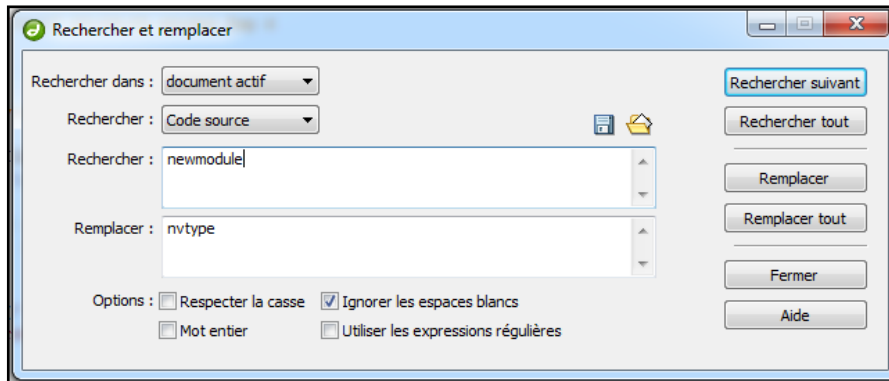


Figure 5.1: renommer les instances du module

5/ Mettre le dossier de 'nvtype' dans le dossier /moddu répertoire de Moodle (dans notre cas on a placé le dossier dans moodle/question/type).

6/ Modification des tables du module dans Paramètres>Administration du site> Développement>éditeurXMLDB.

question/type/numerical/db	[Créer] [Charger] [Modifier] [Enregistrer] [Doc] [XML] [Annuler] [Unload] [Supprimer]
question/type/nvtype/db	[Créer] [Charger] [Modifier] [Enregistrer] [Doc] [XML] [Annuler] [Unload] [Supprimer]
question/type/random/db	[Créer] [Charger] [Modifier] [Enregistrer] [Doc] [XML] [Annuler] [Unload] [Supprimer]
question/type/randomsamatch/db	[Créer] [Charger] [Modifier] [Enregistrer] [Doc] [XML] [Annuler] [Unload] [Supprimer]

Figure 5.2: modification des tables

7/ On trouve les tables du module créé avec succès dans Paramètres> Administration du site>Notifications.



Figure 5.3 : succès d'installation

8/ lors de la visite de Administration>Plugins>Modules d'activités> Gérer les activités, on trouve que ce nouveau module a été ajouté à la liste des modules installés.





 Numerical qtype_numerical	Standard	2012061700		Uninstall
 Mon type qtype_nvtype	Add-on	2012061700		Uninstall
 Random qtype_random	Standard	2012061700		Uninstall
 Random short-answer matching qtype_randomsamatch	Standard	2012061700		Uninstall

Figure 5.4 : l'ajout du module à la liste des modules.

Voilà, notre nouveau module à été ajouté avec succès à la liste des types de question.

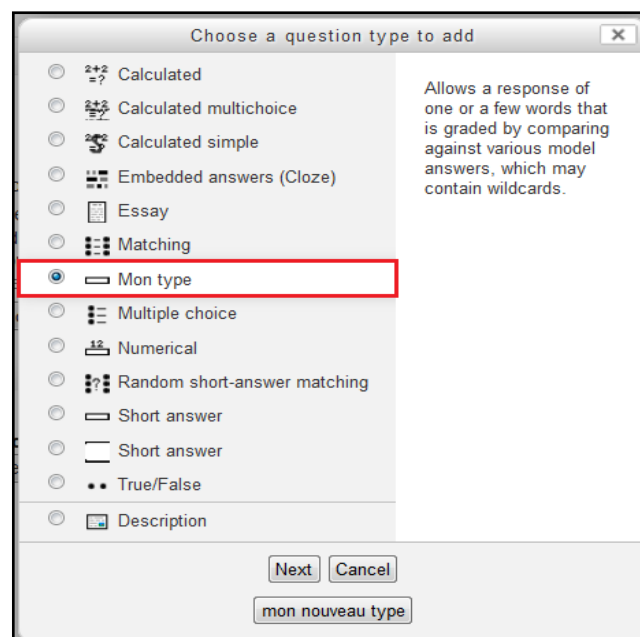


Figure 5.5 : liste des types de question

Après le succès de ces étapes, il faut ajouter notre propre code pour que le plugin arrive à réaliser nos buts.

Codage :

La base de données XML :

Pour rendre possible l'installation de notre module sur n'importe quelle plateforme moodle « 2.x », un fichier XML a été implémenté pour la configuration de la base de données. Ce fichier doit porter le nom *install.xml*.

```

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="../../../../lib/xml/db/xml/db.xsd"

<TABLES>
<TABLE NAME="qtype_rshorttr_options" COMMENT="Options for rshorttr questions">
<FIELDS>
<FIELD NAME="id" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="true"/>
<FIELD NAME="questionid" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="Foreign key references question."/>
<FIELD NAME="espace" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction Negliger es" />
<FIELD NAME="punctuation" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="punctuation." />
<FIELD NAME="titespace" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="Negliger tous les espace." />
<FIELD NAME="accent" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction Negliger ad" />
<FIELD NAME="doubl" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction Negliger do" />
<FIELD NAME="detre" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction Negliger de" />
<FIELD NAME="majmin" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction Negliger l" />
<FIELD NAME="ordre" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction Negliger l" />
<FIELD NAME="stopword" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction stop word" />
<FIELD NAME="synonymy" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction synonyme" />
<FIELD NAME="homophone" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction homonyme" />
<FIELD NAME="separator" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="utilisation de la fonction separae" />
<FIELD NAME="separateur" TYPE="char" LENGTH="255" NOTNULL="true" DEFAULT="" SEQUENCE="false" COMMENT="pour supprimer un ou plusieurs s" />
<FIELD NAME="repo" TYPE="char" LENGTH="255" NOTNULL="true" DEFAULT="" SEQUENCE="false" COMMENT="utilisation des remplacements" />
<FIELD NAME="blanc" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="laisser un blanc seulement entre caract" />
<FIELD NAME="supespace" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="Supprimer tous les espaces." />
</FIELDS>
<KEYS>
<KEY NAME="primary" TYPE="primary" FIELDS="id"/>

```

Figure 5.6:le fichier install.xml

Le fichier de configuration de l'interface :

Ce fichier, est écrit en php, contient toutes les définitions de chaque option dans notre module ce fichier avec le nom qtype\_shorttr.php

```

1 <?php
2 $string['header1']='La Zone de Controle de la Correction Automatique ' ;
3 //interface for Djamel bennouar réalisé par rachda djellatou
4 $string['gradingcontroller']='Grading controller ';//01
5 $string['multiplelinesubcontroller']='Multiple Line SubController ';//02 header
6 $string['multipleshortanswer']='Multiple short Answer';
7 $string['multiplelineorder']='Multiple line order';
8 $string['maxline']='Max line';
9 $string['numbershortanswer']='Number of Short answer';
10 $string['detectextrashortanswer']='Detect extra Short Answer';
11 $string['shortanswergrade']='Short Answer Grade';
12 $string['shortanswerpenaltyifnotgraded']='Short Answer Penaltyif not Graded';
13 ////////////////////////////////////////////////////
14 $string['preprocessingcontroller']=' Preprocessing Controller ';//03
15 $string['packingtext']='Packing text';
16 $string['removingstopwords']='Removing stop words';
17 $string['replacestopword']='Replace stop word';
18 $string['replacechar']='Replace char';
19 ////////////////////////////////////////////////////
20 $string['stopwordsubcontroller']='Stop word subcontroller';//04
21 $string['stopword']='Stop word';
22 $string['addstopword']='Add Stop Word definition';
23 $string['ignorestopword']='Ignore StopWord';
24 $string['stopwordtoignore']='Stop Word to ignore';
25 $string['replacestopword']='Replace Stop Word';
26 ////////////////////////////////////////////////////
27 $string['charactercontroller']=' Character Controller';//05
28 $string['casesensitivity']='Case Sensitivity';
29 $string['packcharrepetition']='pack char repetition';
30 $string['ignorecharacteraccent']='ignore character accent';

```

Figure 5.7:le fichier de définition des interfaces de notre module

Le fichier de développement d'interface :

Ce fichier est le responsable sur l'IHM de notre module ; il est nommé edit\_rshorttr\_form.php.



## Réalisation

Notre stratégie est de prendre des réponses qui sont notées par 0 sur l'outil de Moodle et notées correctes ou partiellement correctes par le professeur et les avoir réutilisées par notre outil et voir les résultats.

Notre outil se base sur les problèmes fréquemment rencontrés par les professeurs qui nous ont orientés sur ces derniers dont on cite :

- ✓ Problèmes des espaces ;
- ✓ Problèmes des accents ;
- ✓ Problèmes de doublement de caractère ;
- ✓ Problème des déterminants ;
- ✓ Problèmes des séparateurs ;
- ✓ Utilisation des mots-clés.

## Comparaison de Module Short Answer Native et Notre Module :

### 1-L'interface Graphique :

#### Cas Le Module Short Answer Native :

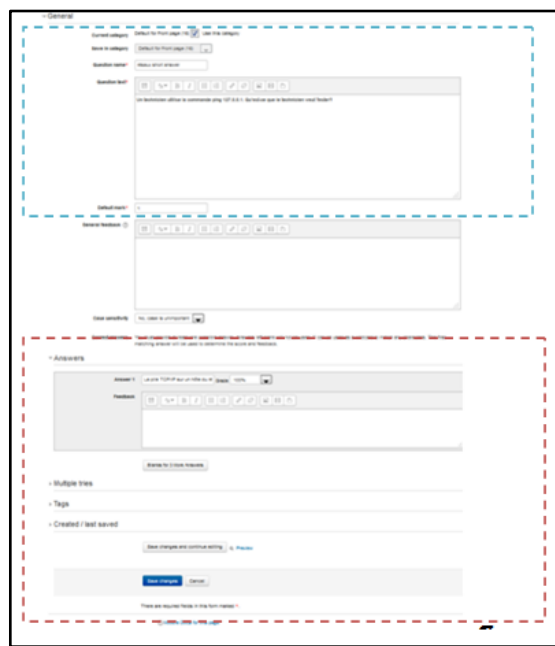


Figure 5.10:L'interface native de Module Short Answer



## Cas Notre Module Réalisé :

### La zone commune :

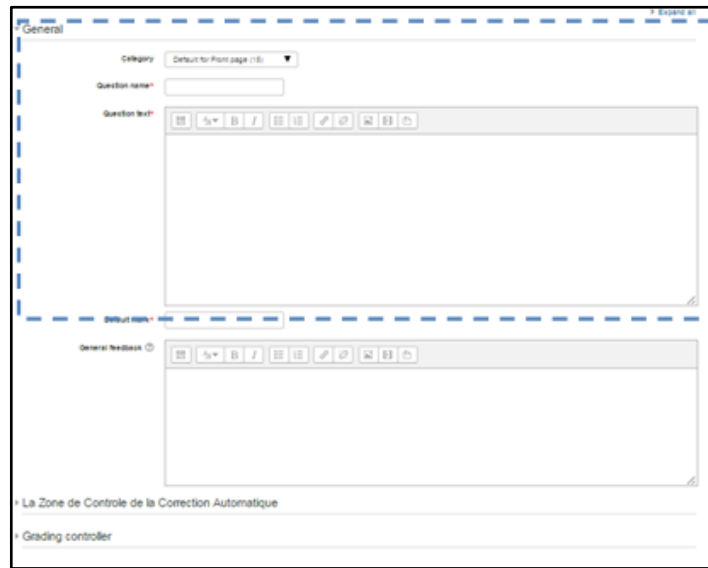


Figure 5.11: La partie en commun avec notre module

### La zone du contrôle des questions automatiques :

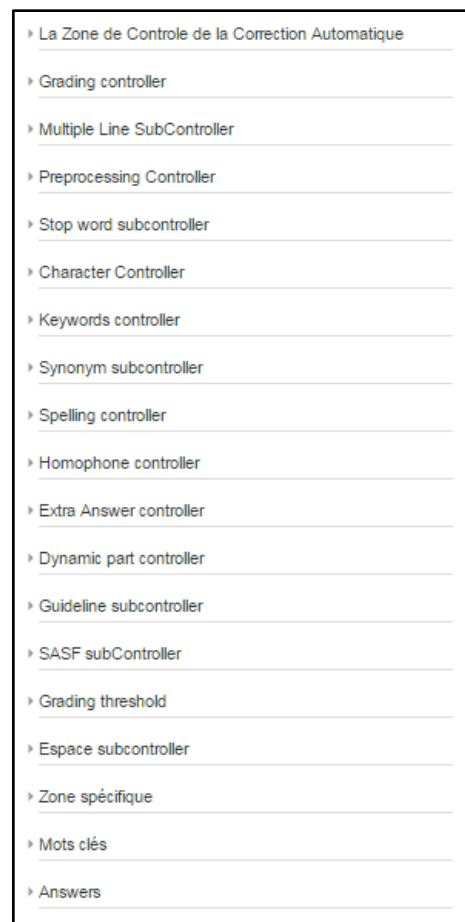


Figure 5.12: les options de configuration du réponse enseignants

## La Zone de Contrôle de la Correction Automatique :

▾ La Zone de Contrôle de la Correction Automatique

- Negliger espaces
- Negliger accents
- Negliger doublement
- Negliger déterminants
- Case Sensitivity Controller
- supprimer tous les espaces
- Punctuation subController
- Stop Word SubController
- Synonymy Controller
- Homophone Controller
- Separator Controller

▸ Grading controller

---

▸ Multiple Line SubController

Figure 5.13:La Zone de Contrôle de la Correction Automatique

## Grading controller

▸ La Zone de Contrôle de la Correction Automatique

▾ Grading controller

[[outilsdessimilarités]]

[[Maximum grade]]  ▼

[[maxattachmentsize]]  ▼

Yes  No

▸ Multiple Line SubController

Figure 5.14:Grading controller

## Multiple Line SubController

▸ Grading controller

▾ Multiple Line SubController

Multiple short Answer  ▼

Multiple line order  ▼

Max line

[[numbershortanswer]]

Detect extra Short Answer  ▼

[[penalty]]  ▼

[[detectduplicateshortanswer]]  ▼

[[shortanswergrade]]

[[shortanswerpenaltyifnotgraded]]

Figure 5.15:Multiple Line SubController

## Preprocessing Controller

▼ Preprocessing Controller

---

Packing text  ▼

Removing stop words  ▼

[[stopword]]  ▼

[[replacestopword]]  ▼

Replace char  ▼

Figure 5.16:Preprocessing Controller

## Stop word subcontroller

▼ Stop word subcontroller

---

[[stopword]]  ▼

[[addstopword]]

[[ignorestopword]]  ▼

[[stopwordtoignore1]]  ▼

[[replacestopword]]  ▼

Figure 5.17:Stop word subcontroller

## Character Controller

▼ Character Controller

---

Case Sensitivity  ▼

pack char repetition  ▼

[[ignorecharacteraccent]]  ▼

Figure 5.18:Character Controller

### Keywords controller :

▼ Keywords controller

Si un Etudiant n'a pas répondu a la bon réponse vous pouvoir utiliser des mots clés ,pour activer cette option ajouter dans la zone de réponse "[%key]" en suit les mots clés séparé par un espace . Exemple: [%key] mot1 mot2

Trouver les mots clés en ordre

ignore/consider Keyword order  ▼

consider keyword only based answer  ▼

Figure 5.19:Keywords controller

### Espace subcontroller

▼ Espace subcontroller

Annuler un seul espace . ?

[[penaltyforusingoption]]  ▼

Annuler les espaces en plus . ?

[[penaltyforusingoption]]  ▼

Annuler tous les espaces . ?

[[penaltyforusingoption]]  ▼

Figure 5.20:Espace subcontroller

### Zone spécifique

▼ Zone spécifique

Supprimer un|plusieur séparateur

Remplacer par un autre séparateur

Figure 5.21:Zone spécifique

## Générer Corpus :

Réponse Professeur \*

Réponse Généré \*

☰ A ▼ B I ☰ ☰ ☰ ☰ ☰ ☰ ☰ ☰

R1: pile TCP/IP sur -----P1:0  
R2: pile TCP/IP sur -----P2:1  
R3: pile TCP/IPsur -----P3:1  
R4:pile TCP/IPsur -----P4:2  
R5: pileTCP/IP sur -----P5:1  
R6:pileTCP/IP sur -----P6:2  
R7: pileTCP/IPsur -----P7:2  
R8:pileTCP/IPsur -----P8:3

Figure 5.22:La partie de génération du corpus

## Teste de quelque fonctionnalité

✓ cas d'ignoration espace :

**Preview question: mon module-réseaux**

1er testreponce  
2eme testreponce 1er testreponce  
2eme testreponce

**Question 1**  
Correct  
Marked out of 1.00

Un technicien utilise la commande ping 127.0.0.1. Qu'est-ce que le technicien veut Tester?

Answer:  ✓

The correct answer is: La pile TCP/IP sur un hôte du réseau

**Preview question: réseaux short answer**

**Question 1**  
Incorrect  
Marked out of 1.00

Un technicien utilise la commande ping 127.0.0.1. Qu'est-ce que le technicien veut Tester?

Answer:  ✗

The correct answer is: La pile TCP/IP sur un hôte du réseau

Figure 5.23: cas d'ignoration espace

✓ cas d'ignorance des accents

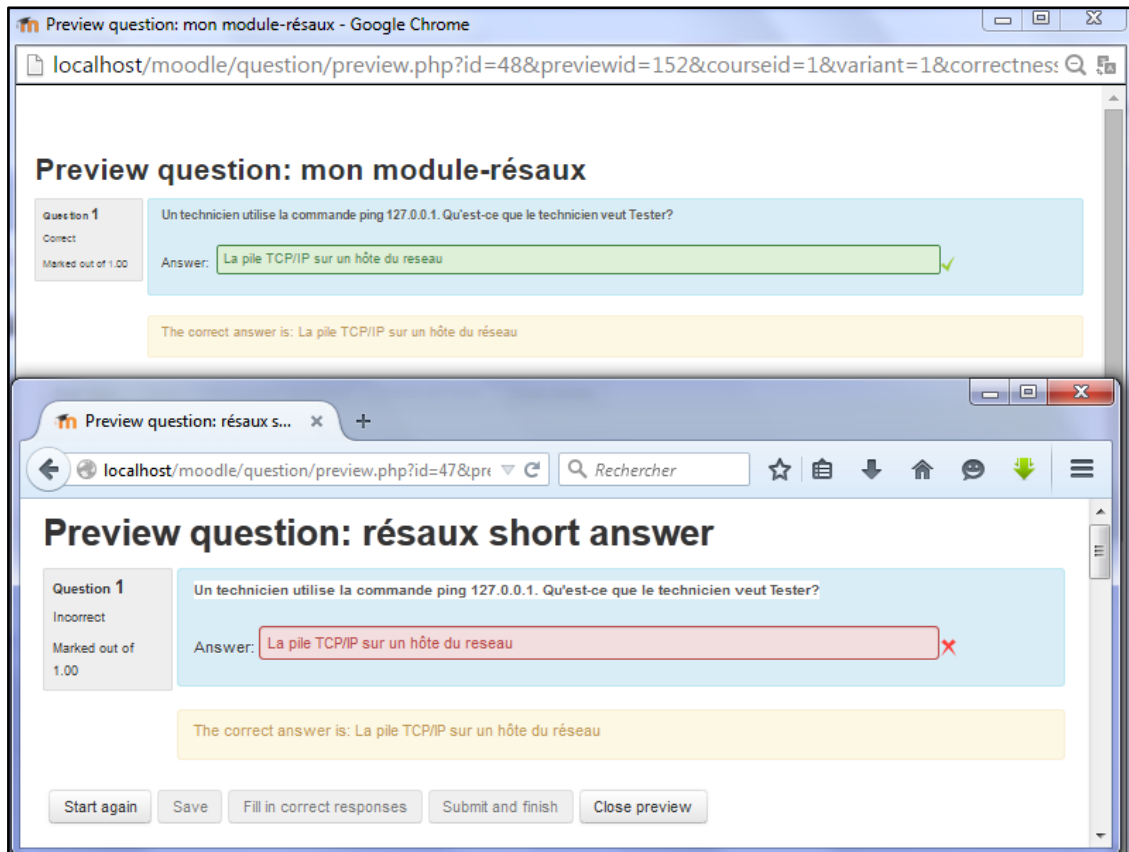


Figure 5.24: cas d'ignorance des accents

✓ cas des majuscules et minuscules

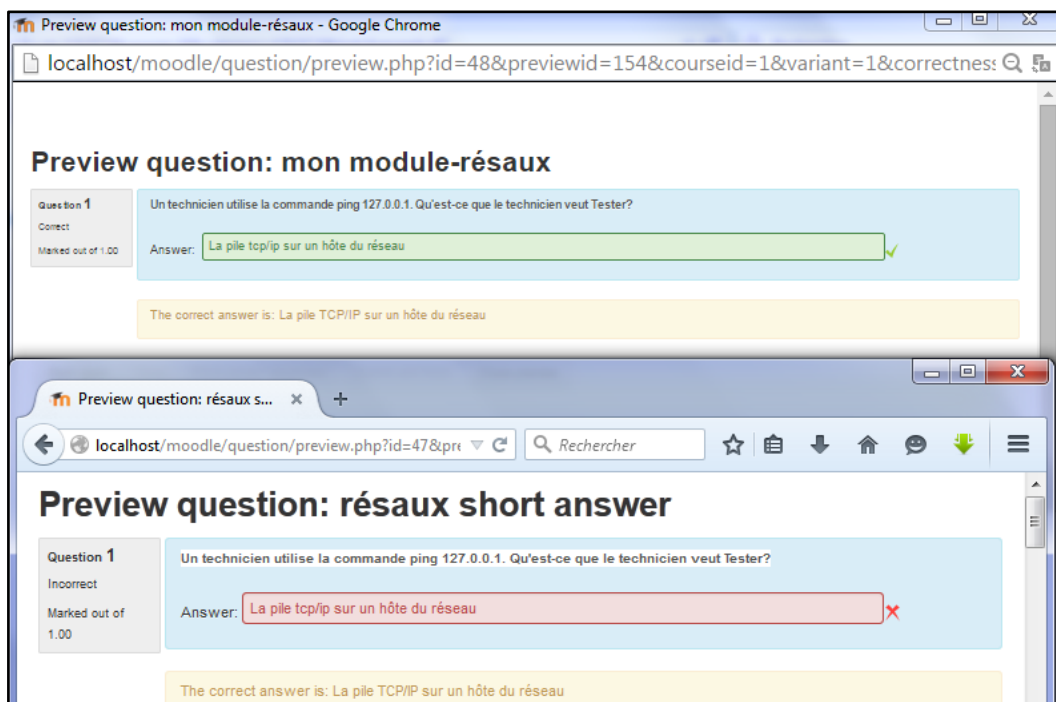


Figure 5.25: cas des majuscules et minuscules

## Résultats

1. Pour faire nos expérimentations, nous avons constitué un échantillon de réponses réelles des étudiants qui ont déjà passé l'examen avec le module architecture des ordinateurs ; on a récupéré toutes les réponses en format xml et on a analysé chaque réponse en les comparant avec la réponse type du professeur ; on a analysé les réponses de 37 étudiants dont chacune contient 11 réponses.

Etape 1 : Préparation des résultats de notre Module

Pour faire notre test de Module on a suivi les étapes suivants :

- ✓ Création de question à nouveau.
- ✓ Réintroduction des réponses des 37 étudiants qui ont été notés « 0 » par le Module de Moodle et les réponses sont partiellement justes.
- ✓ Récupération du fichier Résultats en Format XSL.

Etape 2 :

On a nettoyé nos données de telle sorte les étudiants qui ont répondu juste ou faux soient ignorés, il nous reste 34 étudiants ; on a totalisé les notes de chaque étudiant et on a obtenu les résultats suivants (tableaux 5.1):

Etudiant	Total des réponses récupérés	Note globale de Module développée	Note attribuée manuellement par le professeur
E1	2	0.84	1
E2	1	0.28	1
E3	2	0.56	1.25
E4	3	0.56	2
E5	4	0.56	1
E6	2	0.56	1
E7	4	0.56	2
E8	3	0.56	1.5
E9	2	0.56	1
E10	3	0.56	1.5
E11	2	0.56	1
E12	1	0.28	0.5
E13	3	0.84	1.5
E14	2	0.28	1

Etudiant	Total des réponses	Note globale de Module	Note attribuée manuellement par
----------	--------------------	------------------------	---------------------------------

	récupérés	développée	le professeur
E15	1	0.28	0.5
E16	2	0.56	1
E17	2	0.84	1
E18	1	0.28	0.5
E19	2	0.84	1
E20	1	0.28	0.5
E21	1	0.28	0.5
E22	3	0.56	1.5
E23	3	0.84	1.5
E24	1	0.28	0.5
E25	2	0.28	1
E26	3	0.56	1.5
E27	4	0.84	1.75
E28	3	0.56	1.5
E29	1	0.28	0.5
E30	2	0.84	1
E31	2	0.28	1
E32	1	0.28	0.5
E33	2	0.84	1
E34	5	1.39	2.5
Total	76	18.75	37.5

Tableau 5.1 : Résultats récupéré

D'après ces résultats on a dressé ce graphe en comparant avec la note de professeur.

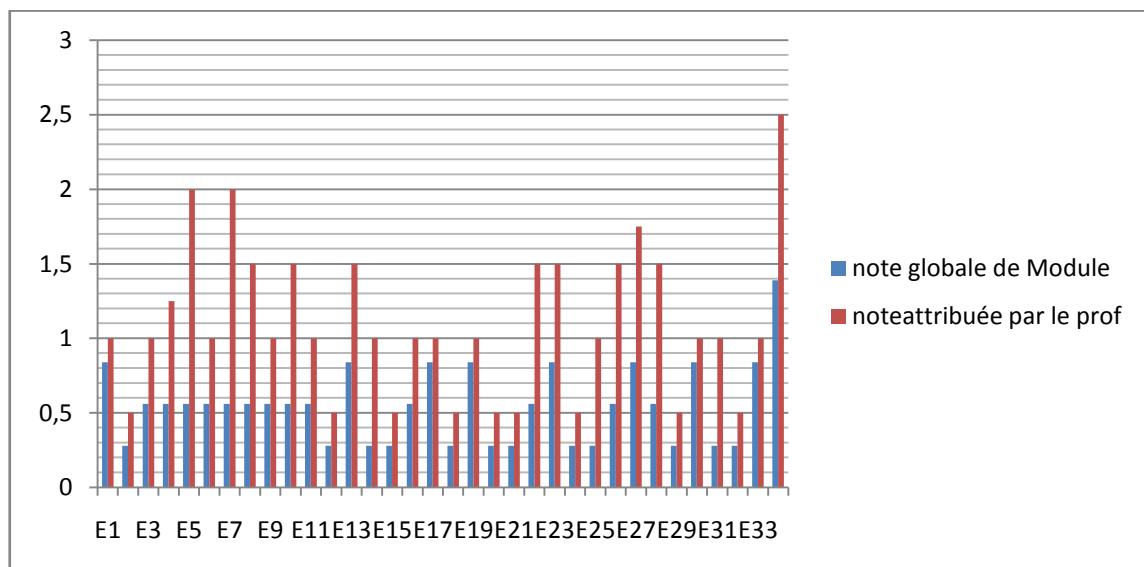


Figure 5.26 : Comparaison des résultats de module avec la correction humaine



Discussion :

D'après les résultats obtenus ; notre outils est améliorer par rapport à l'outil existants et le pourcentage d'approximation par rapport au professeur est d'environ50%

## 6. CONCLUSION GENERALE

L'objectif de notre projet est la conception et l'implémentation d'un nouveau module de test pour les réponses librement rédigées avec de nouvelles options pour améliorer la correction automatique.

Notre travail concerne l'étude des différentes approches de correction automatique et la mise en œuvre d'un système avec des fonctionnalités plus efficaces que les fonctionnalités existantes.

De ce fait, nous avons entamé notre projet par une étude sur la correction automatique dans différents logiciels et plates formes de Elearning, une analyse complète a été faite en plus d'une étude comparative des différentes approches.

En second lieu, nous avons consacré une étude détaillée sur les types de questions librement rédigées, des critiques sur les différentes méthodes de corrections automatiques ont été attribués, leurs points faibles ont formé la base de la conception de notre système, cette dernière est basée sur l'ajout des nouvelles fonctionnalités et options pour résoudre les problématiques actuelles telles que les faiblesses dans l'orthographe, la répétition erronée de lettre, l'oubli d'un accent, l'utilisation d'un mauvais déterminant etc.

De plus, une interface de communication homme machine (IHM) a été réalisée dans le cadre de ce travail. Cette dernière facilite la tâche à l'enseignant lors de la création de la question en attribuant à chaque option des boutons d'aide expliquant son fonctionnement.

Une dernière étape de ce travail est la vérification fonctionnelle de notre nouveau module a été conclue par un test en comparant le taux de réussite des étudiants avant et après la réalisation du projet.

Plusieurs perspectives s'ouvrent à ce travail. Ce qui nous semble le plus intéressant concerne l'ajout d'éléments spécifique à certaines matières. Ainsi à titre d'exemple, dans le contexte spécifique à un module de programmation, on pourrait ajouter l'impact de l'oubli de certains caractères qui sont très important dans la syntaxe d'un programme.

A titre d'exemple on pourrait autoriser l'oubli de la virgule dans l'instruction « `mov_ax,[bp+2]` » écrite en assembleur et considérer comme juste (100%) ou partiellement juste « `mov_ax[bp+2]` ».

Comme nous l'avons spécifié dans ce mémoire, ce type de règles ne peut être développé que par l'enseignant spécialiste dans la matière enseignée.

Il faut reconnaître que l'approche que nous avons proposée n'a pas été testée à grande échelle à cause des difficultés liées principalement à la non-utilisation des techniques du E-learning dans les examens réel d'étudiants.

En effet, le nombre d'enseignants pratiquant le test en ligne semble être nul. Les tests sur lesquels nous nous sommes reposés sont les tests réalisés dans le contexte des examens en ligne conduit par Mr Bennouar Djamel durant 4 années et durant 1 année à l'Université de Bouira. Ces tests visaient des modules particuliers d'informatique tel que la programmation orientée objet avec Java, l'architecture logicielle, l'architecture des ordinateurs, les réseaux informatiques et de télécommunications et l'algorithmiques.

Les questions de ces tests étaient préparées pour que les alternatives de réponses possibles ne soient pas énormes. Ainsi, à titre d'exemples plusieurs indications n'ont pas été testées en grandeur nature tels que l'acceptation de synonymes, et la-non-considération des stop word.

Les indications réellement mise en œuvre durant les test réalisé sont les mots clés, les mots-clés partiel, les accent et la redondance d'espace.

Perspectives ;

Pour démontrer que l'approche que nous avons proposée est réellement efficace plusieurs autres tests doivent être réalisés et des comparaisons avec d'autres outils doivent être conduites.

Actuellement les comparaisons n'ont été faites qu'avec les outils de plateforme libre telle que Moodle dans le contexte de modules particulier dans deux langues particulière, la langue française et le langage de programmation..

C'est ainsi que plusieurs perspectives s'ouvrent pour ce travail, notamment l'utilisation de la langue arabe et la conduite de tests dans des modules en sciences humaines ou le nombre d'alternatives à une réponse est élevé.

Cette dernière perspective semble très difficile à cause principalement du désintérêt porté aux outils d'évaluation automatique de l'étudiant observé chez une très grande majorité d'enseignants selon ce qui a été observé durant la première école sur le télé-enseignement organisé à l'Université de Blida en Avril 2012.

## 7. BIBLIOGRAPHIE

- [1] M. Tanana and P. Vidal, "Évaluation formative du savoir-faire des apprenants à l' aide d' algorithmes de classification," 2009.
- [2] G. Scallon, "L'évaluation des apprentissages dans une approche par compétences," *De Boeck Supérieur*, 2004.
- [3] MADDALENA TARAS, "ASSESSMENT – SUMMATIVE AND FORMATIVE –," *Br. J. Educ. Stud. ISSN 0007-1005*, vol. 53, no. 4, pp. 466–478, 2005.
- [4] M. K. Burns, "What is Formative Evaluation?," 2008.
- [5] P. Rodrigues, F., & Oliveira, "A System for Formative Assessment and Monitoring of Students' Progress," *Comput. Educ.*, vol. 76, pp. 30–41, 2014.
- [6] S. Jordan, "Student engagement with assessment and feedback: Some lessons from short-answer free-text e-assessment questions," *Comput. Educ.*, vol. 58, no. 2, pp. 818–834, Feb. 2012.
- [7] A. Klobucar, N. Elliot, P. Deess, O. Rudniy, and K. Joshi, "Assessing Writing Automated scoring in context : Rapid assessment for placed students," vol. 18, pp. 62–84, 2013.
- [8] C. R. Hancock, "Alternative Assessment and Second Language," pp. 1–7, 1994.
- [9] P. Diana, "Automatic evaluation of users' short essays by using statistical and shallow natural language processing techniques," 2004.
- [10] E. Negre, "Comparaison de textes: quelques approches...," 2013.
- [11] Q. A. Y. Cai, *Semantic Parsing Freebase: Towards Open-domain Semantic Parsing*, vol. 1. 2013.
- [12] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," vol. 26, no. 3, pp. 297–302, 2013.
- [13] U. D. Avignon, E. T. Des, and P. D. E. Vaucluse, "Traduction automatique statistique et adaptation à un domaine spécialisé," 2011.
- [14] W. Gomaa and A. Fahmy, "Short Answer Grading Using String Similarity And Corpus-Based Similarity," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, pp. 115–121, 2012.
- [15] R.W.HAMMING, "Error Detecting and Error Correcting Code," *Bell Syst. Tech. J.*, vol. xx, no. 2, 1950.
- [16] C. Lin, E. Hovy, and M. Rey, "Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics," no. June, pp. 71–78, 2003.
- [17] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "B LEU : a Method for Automatic Evaluation of Machine Translation," no. July, pp. 311–318, 2002.
- [18] P. Diana and E. Alfonseca, "Application of the Bleu algorithm for recognising textual entailments," pp. 1–4, 2001.

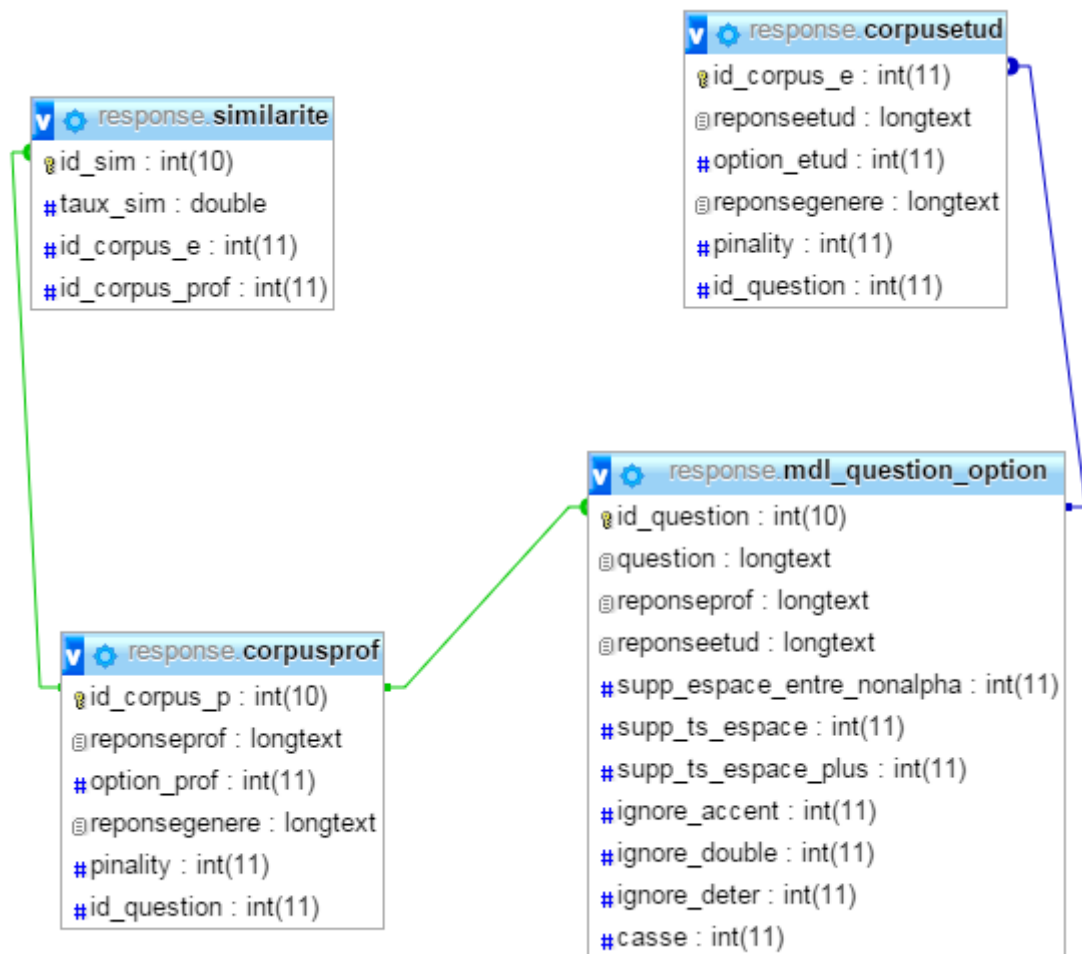
- [19] S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, A. Lagarda, H. Ney, and E. Vidal, "Statistical Approaches to Computer-Assisted Translation," no. December 2007, 2008.
- [20] W. Hassan and A. Aly, "Automatic scoring for answers to Arabic test questions &," *Comput. Speech Lang.*, 2013.
- [21] F. Noorbehbahani and a. a. Kardan, "The automatic assessment of free text answers using a modified BLEU algorithm," *Comput. Educ.*, vol. 56, no. 2, pp. 337–345, Feb. 2011.
- [22] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Process.*, vol. 25, no. 2–3, pp. 259–284, Jan. 1998.
- [23] Y. He, S. C. Hui, and T. T. Quan, "Automatic summary assessment for intelligent tutoring systems," *Comput. Educ.*, vol. 53, no. 3, pp. 890–899, Nov. 2009.
- [24] S. Valenti, F. Neri, and A. Cucchiarelli, "An Overview of Current Research on Automated Essay Grading," *J. Inf. Technol. Educ.*, vol. 2, pp. 3–118, 2003.
- [25] R. Ziai, N. Ott, and D. Meurers, "Short Answer Assessment : Establishing Links Between Research Strands," pp. 190–200, 2012.
- [26] P. Wiemer-hastings, A. C. Graesser, D. Harter, B. Klettke, K. Link, B. Olde, V. Pomeroy, K. Wiemer-, H. Yetman, S. Craig, P. Chipman, M. Ring, and C. Web-, "The Foundations and Architecture of Autotutor."
- [27] P. W. Foltz and D. Laham, "Automated Essay Scoring : Applications to Educational Technology Automated scoring with LSA," pp. 939–944.
- [28] R. Through, T. H. E. Use, O. F. For, N. By, N. O. F. Of, T. Submitted, T. O. The, O. F. The, A. O. F. The, N. Of, O. Fulfillment, R. For, T. H. E. Degree, O. Of, and E. Of, "B.s., u," 2001.
- [29] W. Hou, J. Tsao, S. Li, and L. Chen, "Automatic Assessment of Students ' Free-Text Answers," pp. 235–243, 2010.
- [30] C. Banea, S. Hassan, M. Mohler, and R. Mihalcea, "UNT : A Supervised Synergistic Approach to Semantic Text Similarity," pp. 635–642, 2012.
- [31] S. Hassan and R. Mihalcea, "Semantic Relatedness Using Salient Semantic Analysis," pp. 884–889, 2011.
- [32] E. Mathieu-dupas, "Algorithme des K plus proches voisins pondérés (WKNN) et Application en diagnostic," 2010.
- [33] L. Bin, L. Jun, Y. Jian-Min, and Z. Qiao-Ming, "Automated Essay Scoring Using the KNN Algorithm," in *2008 International Conference on Computer Science and Software Engineering*, 2008, vol. 1, pp. 735–738.
- [34] A. Fazal, F. K. Hussain, and T. S. Dillon, "An innovative approach for automatically grading spelling in essays using rubric-based scoring," *J. Comput. Syst. Sci.*, vol. 79, no. 7, pp. 1040–1056, Nov. 2013.

- [35] R. Swartz, "AUTOMATED EVALUATION OF ESSAYS AND SHORT ANSWERS Jill Burstein."
- [36] S. Abney, "Part-of-Speech Tagging and Partial Parsing," pp. 1–23, 1996.
- [37] J. Z. Sukkarieh and E. Bolge, "Building a Textual Entailment Suite for the Evaluation of Automatic Content Scoring Technologies," pp. 3149–3156.
- [38] E. B. Page and L. Emploi, "The use of the computer in analyzing student essays," pp. 210–225.
- [39] D. Whittington and H. Hunt, "Approaches to the Computerized Assessment of Free Text Responses."
- [40] "Moodle.org: Moodle Statistics." [Online]. Available: <http://moodle.net/stats/>. [Accessed: 05-Nov-2014].
- [41] M. D. Shermis and B. Hamner, "Contrasting State-of-the-Art Automated Scoring of Essays: Analysis," in *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, 2013, pp. 313–346.
- [42] P. Dessus and B. Lemaire, "Assistance informatique à la correction de copies," pp. 205–220, 2004.
- [43] T. K. Landauer, "Automatic Essay Assessment," *Assess. Educ. Princ. Policy Pract.*, vol. 10, no. 3, pp. 295–308, Nov. 2003.
- [44] Y. Attali and J. Burstein, "With e-rater V . 2," vol. 4, no. 3, 2006.
- [45] D. Callear, D. Callearportacuk, J. Jerrams-smith, and V. Soh, "Bridging Gaps in Computerised Assessment of Texts," pp. 139–140, 2001.
- [46] C. Leacock and M. Chodorow, "C-rater : Automated Scoring of Short-Answer Questions," *Comput. Hum.*, vol. 37, pp. 389–405, 2003.
- [47] R. Siddiqi and C. J. Harrison, "On the Automated Assessment of Short Free-Text Responses," pp. 1–11.
- [48] R. Siddiqi, C. J. Harrison, and R. Siddiqi, "Improving Teaching and Learning through Automated Short-Answer Marking," *IEEE Trans. Learn. Technol.*, vol. 3, no. 3, pp. 237–249, Jul. 2010.
- [49] M. Mohler, R. Bunescu, and R. Mihalcea, "Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 752–762.
- [50] S. Dikli, "An Overview of Automated Scoring of Essays," vol. 5, no. 1, 2006.
- [51] J. Wang and M. S. Brown, "Automated Essay Scoring Versus Human Scoring : A Comparative Study," vol. 6, no. 2, 2007.
- [52] A. Ben-Simon, "Toward more substantively meaningful automated essay scoring," *J. Technol. Learn.*, vol. 6, no. 1, pp. 1–47, 2007.

- [53] D. Kanejiya, A. Kumar, and S. Prasad, "Automatic evaluation of students' answers using syntactically enhanced LSA," *Proc. HLT-NAACL 03 Work. Build. Educ. Appl. using Nat. Lang. Process.*, vol. 2, pp. 53–60, 2003.
- [54] E. Alfonseca, R. M. Carro, M. Freire, A. Ortigosa, D. Pérez, and P. Rodríguez, "Authoring of Adaptive Computer Assisted Assessment of Free-text Answers," *Educ. Technol. Soc.*, vol. 8, pp. 53–65, 2005.
- [55] T. K. Landauer, "Automatic Essay Assessment," *Assess. Educ. Princ. Policy Pract.*, vol. 10, no. 3, pp. 295–308, Nov. 2003.
- [56] J. Burstein, "Automated Essay Scoring With E-rater v.2.0," no. November, 2005.
- [57] S. Pecialite, "- S CIENCES et T ECHNIQUES DU L ANGUEDOC - M émoire de S tage de M aster améliorer LSA," 2006.
- [58] P. Diana, E. Alfonseca, and P. Rodr, "Application of the BLEU method for evaluating free-text answers in an e-learning environment," pp. 1351–1354.
- [59] W. H. Gomaa and A. A. Fahmy, "Automatic scoring for answers to Arabic test questions," *Comput. Speech Lang.*, vol. 28, no. 4, pp. 833–857, Jul. 2014.
- [60] D. Perez-Marin, "Adaptive Computer Assisted Assessment of free-text students' answers: An approach to automatically generate students' conceptual models," 2009.
- [61] P. Selvi and a. K. Bnerjee, "Automatic Short -Answer Grading System (ASAGS)," 2010.
- [62] M. Mohler and R. Mihalcea, "Text-to-text Semantic Similarity for Automatic Short Answer Grading," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL '09)*, 2009, pp. 567–575.



## Annexe



**Les tables de notre Module ajoutée à Moodle**