

Republique Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE  
DEPARTEMENT D'INFORMATIQUE

## **Mémoire de Fin d'Etudes De MASTER PROFESSIONNEL**

Domaine : **Mathématiques et Informatique**

Filière : **Informatique**

Spécialité : **Ingénierie des systèmes  
d'information**

*Présenté par*

**Fatima Iamrache**

**Kheloudja Oukfif**

**Thème**

**Analyse et classification d'opinion sur les produits  
ENIEM**

*Mémoire soutenu publiquement le 29/09/ 2016.Devant le jury composé de :*

Président : M Arezki HAMMACHE

Encadreur : M<sup>me</sup> BOUARAB Farida

Co-Encadreur : M<sup>me</sup> CHOUAKI Samia

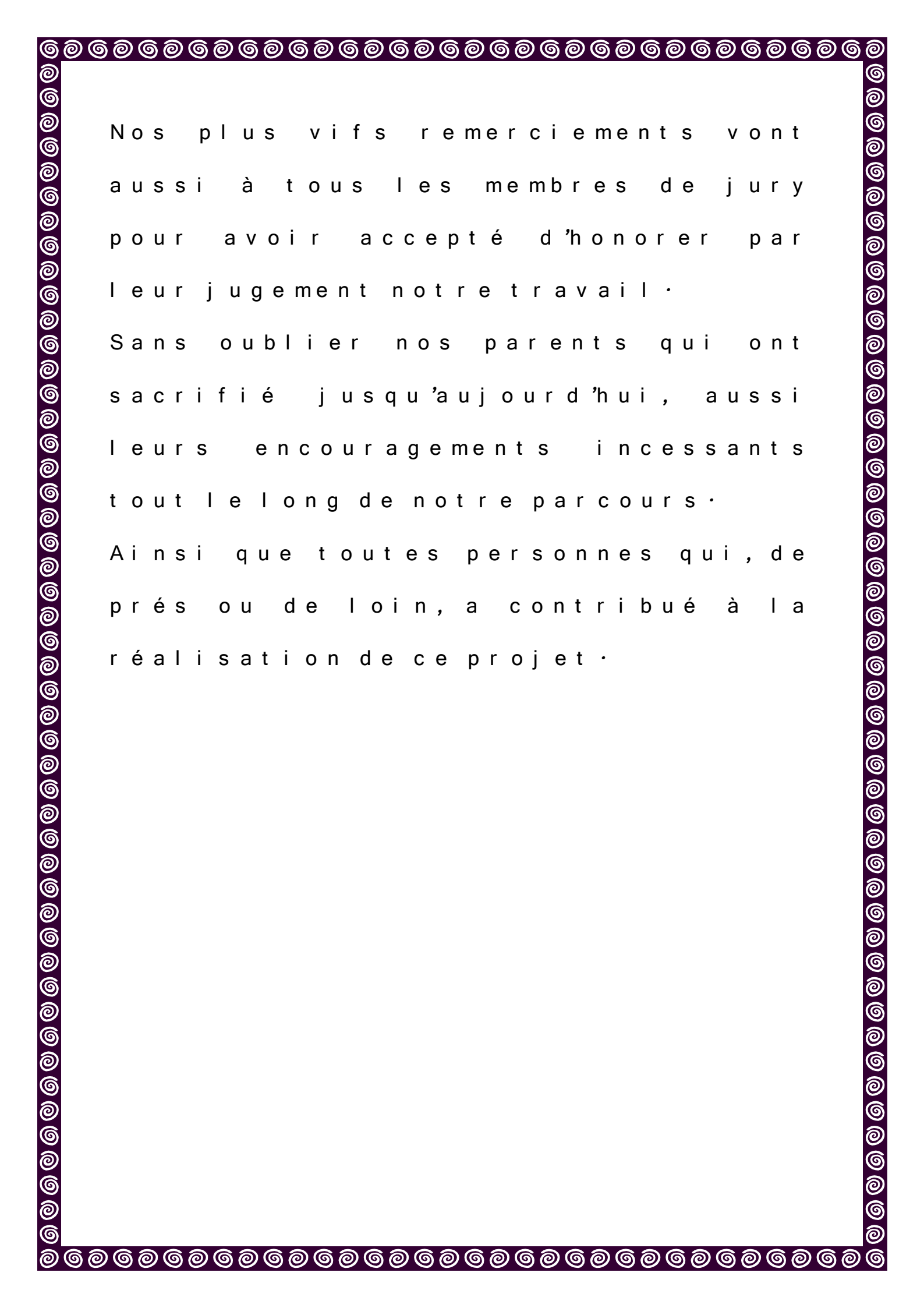
Examineur : M Karima OUKFIF

Examineur : M Rebiha HADAoui

## *Remerciements*

Quelle meilleure opportunité que cette thèse qui marque Achèvement des études universitaires, pour exprimer les plus vifs remerciements à notre promotrice madame BOUARAB pour nous avoir encadrées et guidées tout au long de notre projet.

Nos plus vifs remerciements vont aussi à tout le personnel de l'entreprise nationale des industries de l'électroménagère ENIEM, plus particulièrement Madame CHOUAKI qui nous a généreusement aidés durant notre stage au sein de l'unité commerciale, nous tenons à lui exprimer notre profonde reconnaissance pour le temps précieux qu'elle nous a consacré, ses conseil, ses encouragements et pour la documentation qu'elle a mis à notre disposition.



Nos plus vifs remerciements vont aussi à tous les membres de jury pour avoir accepté d'honorer par leur jugement notre travail.

Sans oublier nos parents qui ont sacrifié jusqu'aujourd'hui, aussi leurs encouragements incessants tout le long de notre parcours.

Ainsi que toutes personnes qui, de près ou de loin, a contribué à la réalisation de ce projet.

## Dédicaces

**Je dédie ce modeste travail à :**

*A celle qui a bercé mes rêves : ma mère ;*

*A celui qui a nourri mes ambitions : mon père ;*

*A mes anges gardiens : mes frères ;*

*A celle qui a soulevé bien des fardeaux avec moi : ma sœur ;*

*A toi Yacine ;*

*A toute mes ami(e)s sans exception ;*

*A mon binôme et sa famille ;*

FATIMA

## Dédicaces

**Je dédie ce modeste travail à :**

*A la mémoire de ma grand-mère & mon frère lounis ;*

*A celle qui a bercé mes rêves : ma mère ;*

*A celui qui a nourri mes ambitions : mon père ;*

*A mes anges gardiens : mes frères ;*

*A celles qui ont soulevé bien des fardeaux avec moi :  
mes sœurs ;*

*A mes nièces et mes neveux ;*

*A mon mari : Redouane et sa famille ;*

*A toute mes ami(e)s sans exception ;*

*A mon binôme et sa famille ;*

KHELOUDJA

# Liste des figures :

|                                                                                                                                 |    |
|---------------------------------------------------------------------------------------------------------------------------------|----|
| Figure1 : Organigramme générale de l'ENIEM.....                                                                                 | 04 |
| Figure2 : Organigramme générale de l'unité commerciale.....                                                                     | 06 |
| Figure 3 : Processus de fouille d'opinions.....                                                                                 | 13 |
| Figure 4 : Exemple d'hyperplan (H) séparant les individus appartenant à la classe (+) et ceux appartenant à la classe (–) ..... | 24 |
| Figure5 : Exemple de représentation d'un acteur .....                                                                           | 30 |
| Figure 6: Diagramme de cas d'utilisation générale .....                                                                         | 35 |
| Figure 7: Diagramme de séquence du cas d'utilisation « Authentification » .....                                                 | 38 |
| Figure 8: Diagramme de séquence du cas d'utilisation « Ajouter une famille de service » .....                                   | 39 |
| Figure 9: Diagramme de séquence du cas d'utilisation « Ajouter les polarités des mots ».....                                    | 40 |
| Figure 10: Diagramme de séquence du cas d'utilisation « Ajouter un fonctionnaire eniem ».....                                   | 41 |
| Figure 11: Diagramme de séquence du cas d'utilisation « Modifier un client ».....                                               | 42 |
| Figure 12: Diagramme de séquence du cas d'utilisation « Supprimer un fonctionnaire eniem ».....                                 | 43 |
| Figure 13: Diagramme de séquence du cas d'utilisation « Créer questionnaire ».....                                              | 44 |
| Figure 14: Diagramme de séquence du cas d'utilisation « Activer questionnaire ».....                                            | 45 |
| Figure 15: Diagrammes de classe globale.....                                                                                    | 47 |
| Figure 16 : Interface PHPMyAdmin.....                                                                                           | 55 |
| Figure 17 : Interface Eclipse.....                                                                                              | 59 |
| Figure 18 : Capture d'écran de l'interface « page d'accueil ».....                                                              | 59 |
| Figure 19 : Capture d'écran de l'interface « authentification administrateur ».....                                             | 60 |
| Figure 20: Capture d'écran de l'espace « administrateur ».....                                                                  | 61 |
| Figure 21: Capture d'écran de l'espace « client ».....                                                                          | 62 |
| Figure 22: Capture d'écran de l'espace « expert linguiste ».....                                                                | 63 |
| Figure 23: Capture d'écran de « questionnaire ».....                                                                            | 64 |
| Figure 24: Capture d'écran de l'interface « analyser ».....                                                                     | 65 |
| Figure 25: Capture d'écran de l'interface « synthétiser».....                                                                   | 66 |

# Sommaire :

---

## Sommaire

Liste des figures

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| Introduction générale.....                                                     | 1  |
| Chapitre I : Présentation de l'organisme d'accueil.....                        | 2  |
| 1. Introduction .....                                                          | 2  |
| 2. Entreprise ENIEM.....                                                       | 2  |
| 2.1. Historique .....                                                          | 2  |
| 2.2. Mission de l'entreprise ENIEM.....                                        | 3  |
| 2.3. Organigramme de l'entreprise ENIEM .....                                  | 4  |
| 2.4. Organisation générale.....                                                | 5  |
| 2.5. Présentation de l'unité commerciale.....                                  | 6  |
| 3. Problématique.....                                                          | 9  |
| Chapitre II: La fouille d'opinion et l'analyse des sentiments.....             | 2  |
| 1. Introduction .....                                                          | 10 |
| 2. Texte Mining .....                                                          | 10 |
| 3. Objectivité et subjectivité .....                                           | 10 |
| 4. La fouille d'opinion et l'analyse de sentiment .....                        | 11 |
| 4.1. La fouille d'opinion.....                                                 | 11 |
| 4.2. L'analyse des sentiments.....                                             | 12 |
| 4.3. Processus de la fouille d'opinions .....                                  | 13 |
| 4.4. Domaines d'application.....                                               | 14 |
| 4.5. Difficultés de la fouille d'opinions et de l'analyse des sentiments ..... | 15 |
| 5. Les approches de l'analyse d'opinion .....                                  | 17 |
| 5.1. Les approches symboliques.....                                            | 17 |
| 5.2. Les approches statistiques .....                                          | 20 |
| 5.3. Les approches hybrides .....                                              | 25 |
| 6. Conclusion.....                                                             | 27 |
| Chapitre III : Analyse et conception.....                                      | 28 |
| 1. Introduction .....                                                          | 28 |
| 2. L'objectif de travail .....                                                 | 28 |
| 3. La démarche d'élaboration de notre projet.....                              | 28 |

## Sommaire :

---

|                                                              |    |
|--------------------------------------------------------------|----|
| 4. Le processus de développement d'applications web .....    | 29 |
| 5. Les étapes d'élaboration de notre projet.....             | 29 |
| 5.1. Analyse.....                                            | 29 |
| 5.2. Vue dynamique du système.....                           | 36 |
| 5.3. Vue statique du système .....                           | 46 |
| 6. Conclusion.....                                           | 53 |
| Chapitre IV: Déploiement et implémentation .....             | 52 |
| 1. Introduction .....                                        | 54 |
| 2. Présentation du système .....                             | 54 |
| 2.1. Description de l'environnement de développement : ..... | 54 |
| 2.2. Les outils de développement .....                       | 54 |
| 2.3. Captures d'écran.....                                   | 59 |
| 3. Conclusion.....                                           | 67 |
| Conclusion générale .....                                    | 68 |
| Référence bibliographique                                    |    |

# Introduction générale

## Introduction générale

La Fouille d'opinion (Opinion Mining) et l'Analyse des sentiments (Sentiments Analysis), font partie d'un domaine émergent. Ce dernier s'occupe de traitement de la subjectivité : opinions, avis, sentiments, émotions, évaluations, croyances ou jugements personnel. Ensuite, il attribue une polarité (positive, négative ou neutre) à cette opinion.

Ces données d'opinion revêtent aujourd'hui une importance stratégique et économique évidente car leur analyse permet de connaître les points forts et les points faibles des produits, d'estimer la perception du produit par les clients, afin d'améliorer les produits. Il permet également aux clients de donner son opinion, de l'aider à la prise de décision, en s'inspirant des sentiments et d'opinions d'autres clients sur les produits.

Notre travail s'inscrit principalement dans le contexte de l'analyse d'opinion appliqué au domaine de l'entreprise et notre objectif est de développer une application WEB pour recueillir et synthétiser les opinions des clients de l'entreprise ENIEM. L'analyse de cette opinion permettra à l'entreprise d'améliorer la qualité de ses services et d'augmenter le taux de satisfaction de ses clients.

Pour ce faire, nous avons organisé notre mémoire en quatre chapitres, le premier présente l'organisme d'accueil qui n'est autre que l'entreprise nationale des industries de l'électroménagers ENIEM.

Le deuxième chapitre aborde les concepts de la fouille d'opinion et l'analyse de sentiment, les applications de cette discipline notamment dans le monde de l'entreprise, ainsi que la présentation des différentes approches de l'analyse d'opinion à savoir les méthodes symbolique, les méthodes statistiques et les méthodes hybrides.

Dans le troisième chapitre, nous présentons l'analyse et la conception de notre application, Nous décrivons le processus de développement d'application Web en utilisant le langage de modélisation très répandu UML.

Le Dernier chapitre est consacré à la description des étapes de réalisation et d'implémentation de notre application.

# **Chapitre I : Présentation de l'organisme d'accueil**

## 1. Introduction

Afin de vous permettre de familiariser avec notre domaine d'application (analyse et classification d'opinion), nous allons présenter dans ce chapitre l'entreprise ENIEM.

## 2. Entreprise ENIEM

L'ENIEM (Entreprise Nationale des Industries de l'Electroménager), est une entreprise publique économique de droit algérien (EPE). Spécialisée dans la production et la commercialisation des appareils électroménagers.

### 2.1. Historique

L'entreprise nationale des industries de l'électroménager ENIEM est issue de la restructuration organique de la SONELEC (société nationale de fabrication et montage du matériel électrique et électronique).

Elle a été créée en vertu du décret N°83-19 du 02 janvier 1983. L'entreprise a été chargée de la production et la commercialisation des produits électroménagers et disposait à sa création de :

- Complexe d'appareils électroménagers (CAM) de Tizi Ouzou, entré en production en juin 1977.
- Unité lampes de Mohammedia(ULM), entrée en production en juin 1979.

Le statut de l'ENIEM est passé d'une entreprise publique et économique (EPE) à celui d'une société par actions (SPA). Ainsi elle est passée à l'autonomie le 08 octobre 1989 avec un capital social de 40.000.00 DA. Depuis 1996, l'entreprise est organisée en unités, et filiale l'unité lampes de Mohammedia.

L'ENIEM est la première entreprise du Maghreb à être certifiée ISO 9002 depuis le premier juillet 1998 par les experts de l'association française de l'assurance de la qualité (AFAQ), puis gratifiée en 2003 de l'ISO 9001 version 2000. A noter que les produits ENIEM sont 0% cfc (chloro fluoro carbones), et ce depuis 1997.

Le siège social de l'entreprise se situe au chef lieu de la wilaya de Tizi Ouzou.

Les unités de production froids, cuisson, climatisation, et prestations technique, issue de la restructuration du complexe(CAM) et l'unité commerciale sont implantées à la zone industrielle AISSAT IDIR de OUAD AISSI, distance de 10 Km du chef lieu de la wilaya. L'unité sanitaire est située à Miliana, wilaya d'Ain Défila, et la filiale lampes de Mohammedia à la wilaya de Mascara.

# Chapitre I : Présentation de l'organisme d'accueil

---

## 2.2. Mission de l'entreprise ENIEM

Dans le cadre de développement économique et social, l'ENIEM assure les fonctions de production, montage, commercialisation et de la recherche dans les différentes branches de l'électroménagers notamment :

- Les équipements ménagers domestiques.
- Les équipements industriels.
- Le petit appareil ménager.

Elle assure également la production :

- Des appareils réfrigérateurs et congélateurs des différentes capacités (160L à 520L).
- Des cuisiniers à gaz 4 et 5 feux.
- Des climatiseurs types fenêtres et Split système (1CV à 2,5 CV)

Ainsi la politique de l'entreprise se manifeste par la ferme volonté de la direction générale à :

- Accroître la satisfaction des clients.
- Développer les compétences et la communication.
- Veiller à la réduction des coûts de non qualité par maîtrise de processus.
- Améliorer en contenu l'efficacité du système de management de la qualité.

# Chapitre I : Présentation de l'organisme d'accueil

## 2.3. Organigramme de l'entreprise ENIEM

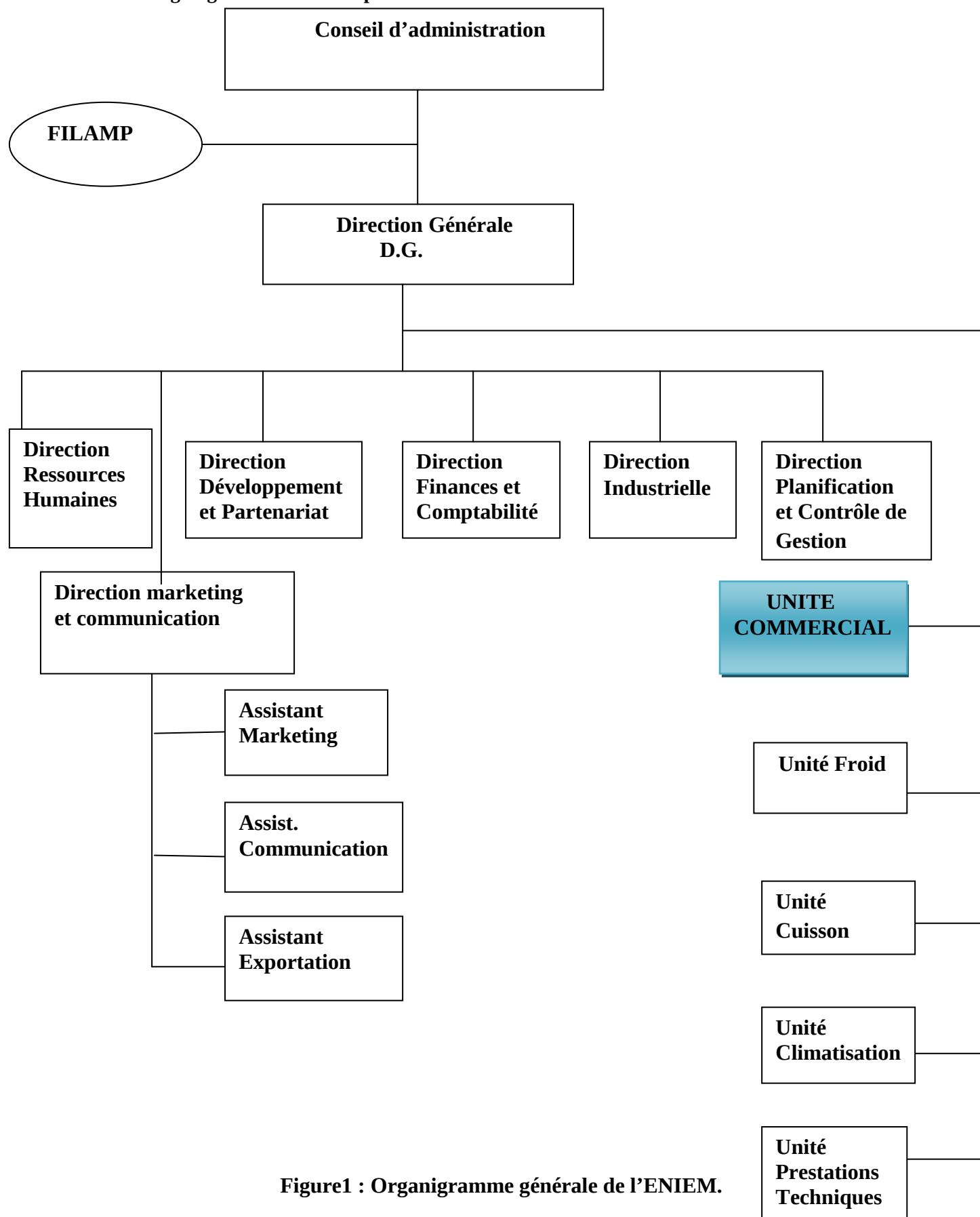


Figure1 : Organigramme générale de l'ENIEM.

# Chapitre I : Présentation de l'organisme d'accueil

---

## 2.4. Organisation générale

### 2.4.1. La direction générale

Elle est composée d'un staff directionnel (président directeur général) et d'assistants, à savoir ; juridique administration et de plusieurs direction qui sont :

- Direction marketing et communication.
- Direction développement et partenariat.
- Direction ressources humaines.
- Direction finance et comptabilité.
- Direction planification et contrôle de gestion.
- Direction Industrielle.

### 2.4.2 Les unités

L'entreprise s'est organisée par centre d'activités stratégique qui se composent de 05 unités qui sont :

- L'unité froid.
- L'unité de cuisson.
- L'unité de climatisation.
- L'unité prestation technique (UPT).
- L'unité commerciale (UC).
  
- **Unité froid** : spécialisée dans la fabrication de réfrigérateurs, congélateurs et armoires frigorifique.
- **Unité cuisson** : chargée de la production et du développement des produits de cuisson (cuisinières, réchauds plats...).
- **Unité climatisation** : chargée de la production de tout types de climatiseurs à savoir ; Split système et monoblocs.
- **Une unité commerciale** : s'occupe de la gestion de la commercialisation des produits de l'entreprise.
- **Une unité de prestations techniques** : elle fournit les services nécessaires aux unités de productions ainsi que des prestations techniques.

En plus de ses unités il y'a deux filiales dont le capital est à 100% ENIEM.

- Unité de lampe de Mohammedia (ULM) entrée en production en juin 1979.

## Chapitre I : Présentation de l'organisme d'accueil

- Unité produits sanitaires : installée à Miliana wilaya de Mascara.

Comme notre étude va se dérouler au niveau de l'unité commerciale nous détaillons dans ce qui suit cette unité.

### 2.5. Présentation de l'unité commerciale

#### 2.5.1. L'Organigramme

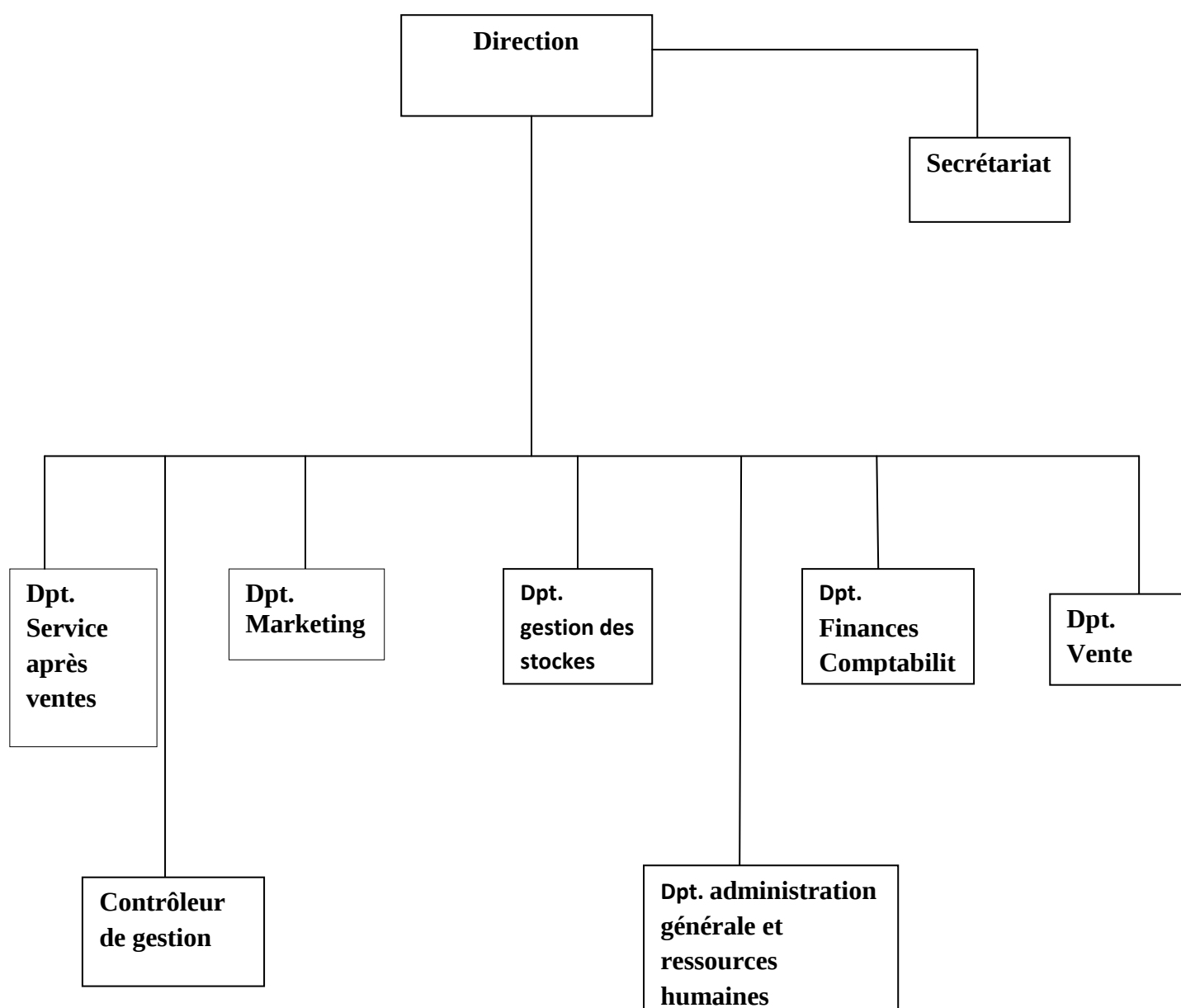


Figure2 : Organigramme générale de l'unité commerciale

### 2.5.2. Historique et fonction

L'unité commerciale créée par l'entreprise le 02 janvier 1989, assure la fonction de commercialisation des produits finis de l'entreprise, de la promotion des exportations et de la gestion du réseau SAV (Service Après Vente) et pour cela, un réseau d'agents agréée a été mise en place.

### 2.5.3. Organisation générale

L'unité commerciale est composée essentiellement, d'une direction commerciale au sommet de la hiérarchie à laquelle est rattaché le secrétariat. Cette direction a sous sa tutelle sept départements qui collaborent pour mettre en œuvre la stratégie commerciale de l'entreprise.

Ces départements sont :

#### 2.5.3.1. Département Vente

Trois services composent ce département ; un service vente (facturation), un service client et un service synthèse et recouvrement.

- **Service vente** : chargé principalement de :
  - La réception des commandes des clients.
  - La gestion des facturations.
  - La gestion des règlements.
  - Le contrôle des livraisons par rapport aux facturations.
- **Service client** : cette structure enregistre les clients et gère le réseau des agents agréés à travers le territoire national, et ce par :
  - La définition des critères de sélection des agents agréés.
  - La réception et l'étude des dossiers de soumission.
  - L'établissement des conventions vente.
  - La classification et le suivi des clients.
- **Service synthèse et recouvrement** : le travail de ce service consiste à recouvrir les créances qui sont générées par les ventes à terme, les ventes par facilité, les ventes par traite ou bien par chèque. Il est aussi chargé du traitement et du suivi des marchés.

#### 2.5.3.2. Département Marketing

Ce département met en œuvre la stratégie concurrentielle de l'entreprise. Il est composé de chargés des foires qui s'occupent de la préparation et de l'organisation des foires et salons nationaux et internationaux ainsi que la réalisation des maquettes, ainsi que de chargés de marchés qui font la prospection et l'étude du marché.

## Chapitre I : Présentation de l'organisme d'accueil

---

Ce département assure également l'écoute client afin d'en assurer sa pleine satisfaction à travers notamment l'étude des questionnaires retourné par les clients.

### **2.5.3.3. Département Service Après Vente**

Ce département est responsable de la gestion des circuits SAV à travers tout le territoire national. Plus de 200 agents agréés forment le réseau SAV de l'ENIEM, ils sont à la charge du département service après vente, qui doit les alimenter en pièces de rechange pour assurer les prestations service qui s'inscrivent dans le cadre de la garantie sur les produits de l'entreprise. Lors des retours de produits par les clients, le département SAV, à travers une commission technique spécialisée décide de la réception ou remplacement du (des) produit(s) retourné.

Le département SAV se charge de recueillir le questionnaire de satisfaction client qu'il transmet au département marketing pour traitement.

### **2.5.3.4. Département Finance et Comptabilité**

Ce département assure la comptabilisation de toutes les opérations permettant l'établissement des bilans et des comptes concernant la couverture des dépenses des autres départements en matière de prestations techniques à base desquelles sont définis les prix de vente et les taxes sur les produits finis.

Ce département est composé de deux services : un service comptabilité générale et un service finances.

### **2.5.3.5. Département Contrôle de Gestion**

Ce département est un organe de contrôle de toute la gestion de l'unité commerciale. Il réalise et présente tous les travaux permettant de produire une information complète et cohérente des activités de l'unité (production, commercialisation, approvisionnement et finance). Il a accès pratiquement à toutes les informations produites par les autres services ou départements. Alors qu'il exploite et analyse l'information relative aux agrégats de gestion afin de préconiser les actions correctives nécessaires avec toute l'anticipation attendue, il exploite les résultats des audits finance les interprète et fait les recommandations nécessaires.

### **2.5.3.6. Département gestion des stocks**

Le département gestion des stocks est chargé de la tenue des stocks des produits finis se trouvant dans les magasins de stockage du complexe, de la livraison de ces produits aux clients, de l'alimentation des points de ventes en produits finis à travers tout le territoire national. Il gère également le transport de ces produits finis.

### 2.5.3.7. Département ARGH

Le département administration générale et ressources humaines est formé du service gestion du personnel qui gère les paies, les emplois, la formation et les correspondants sociaux. Et du service, moyens généraux qui se charge de la gestion des affaires générales, les achats et des relations extérieures.

### 3. Problématique

Une évaluation de la satisfaction consommateur est un élément important pour garantir la pérennité de l'entreprise, il faut donc y prêter une grande attention.

Pour réussir cette mission « évaluation de la satisfaction client », la structure marketing de l'unité commerciale a mis en place, un système d'échange d'information ENIEM/CLIENTS sous forme d'un questionnaire consommateur soutenu par deux tombolas semestrielles pour encourager et motiver ses clients qu'elle invite à retourner dument remplis.

Dans notre projet, nous nous intéressons à ce volet qui consiste à extraire l'opinion des clients afin de pouvoir évaluer le degré de leur satisfaction.

## **Chapitre II: La fouille d'opinion et l'analyse des sentiments**

### 1. Introduction

La fouille d'opinion (Opinion Mining, Sentiment Analysis ou Subjectivity Analysis) est un sous domaine de la fouille de texte. Son but étant de ressortir les marques d'opinions et de sentiments des documents textuels. Une opinion peut être définie comme l'expression des sentiments d'une personne envers une entité.

Les marchands et les fabricants des produits permettent aux clients de donner leurs avis et opinions sur les produits qu'ils ont vendus.

### 2. Texte Mining

Le Texte Mining est une technique permettant d'automatiser le traitement de gros volumes de contenus texte pour en extraire les principales tendances et répertorier de manière statistique les différents sujets évoqués.

Les techniques de Texte Mining sont surtout utilisées pour des données déjà disponibles au format numérique et sont de plus en plus utilisées pour une analyse quasiment en temps réel.

Le Texte Mining est notablement utilisé dans la mesure de l'e-réputation ou pour détecter des tendances.

### 3. Objectivité et subjectivité

Il existe deux catégories principales pour classer l'information textuelle : Ceux qui présentent des faits et ceux qui présentent des opinions. Dans le premier cas, il s'agit de descriptions objectives. Dans l'autre, il s'agit d'expressions subjectives.

❖ **Objectivité** : L'objectivité d'un sujet ou d'un texte est liée aux expressions et aux vocabulaires employés. Elle s'exprime généralement en termes de neutralité, d'impartialité, de désintéressement, ou d'impersonnalité. Il s'agit d'une prise de distance du sujet vis-à-vis de lui-même. L'individu objectif est censé, au moment de porter un jugement, abandonner tout ce qui lui est propre (idées, croyances ou préférences personnelles) pour atteindre une espèce d'universalité. Certains indices textuels permettent de déterminer l'objectivité de son auteur :

- l'énoncé de faits réels et vérifiables
- un style, un ton et un vocabulaire neutres
- l'emploi de pronoms personnels à la troisième personne, comme «il» ou «on», sauf à l'intérieur des citations où ils ne sont pas obligatoires

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

- l'utilisation de citations, de références et de statistiques pour appuyer ou renforcer des affirmations
  - l'absence d'interpellation du lecteur ou de l'interlocuteur par l'auteur du texte
    - o ainsi, jamais l'emploi de la première et la deuxième personne n'est permis, sauf dans les citations
    - o on évite les phrases interrogatives directes et exclamatives
    - o on n'utilise pas l'impératif
  - L'emploi de la forme impersonnelle (il faut que, il y a, il existe, il paraît que, etc.)
- ❖ **Subjectivité** : La subjectivité s'oppose à l'objectivité en cela qu'elle permet à l'auteur d'exprimer ses idées personnelles. Est subjectif ce qui dépend de moi ou d'un point de vue particulier. Un jugement est subjectif s'il reflète les passions, les préjugés et les choix personnels d'un sujet. Synonyme de partialité. Voici quelques indices textuels...
- la formulation d'opinions personnelles, de jugements, de goûts, de sentiments, d'émotions
  - un style, un ton et un vocabulaire descriptifs, expressifs, imagés – l'emploi de la phrase exclamative est alors tout à fait justifié
  - l'emploi de pronoms personnels de la première et la deuxième personne à l'intérieur comme à l'extérieur des citations : «je», «tu», «nous» et «vous»
  - l'utilisation de citations pour appuyer ou renforcer des opinions ou des jugements
  - l'interpellation du lecteur par l'auteur peut aussi être utilisée le message par l'emploi des pronoms personnels (tu, nous, vous) ou autres (phrases impératives ou interrogatives par exemple).

### 4. La fouille d'opinion et l'analyse de sentiment

**4.1. La fouille d'opinion « Opinion Mining »** en anglais, est une discipline qui consiste à extraire et analyser pour un objet donné, les opinions exprimées sur chacun de ses aspects.

Une **opinion** est caractérisée par sa **polarité** (positive, négative et neutre), son **intensité** (par exemple, « excellent » est plus fort que « bon ») et son **émetteur**, c'est-à-dire celui qui émet l'opinion. De plus, elle porte sur un **objet** qui peut être un produit, un service, un événement, une personne, une entreprise...

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

Le domaine de la fouille d'opinion peut-être divisé en trois sous-domaines :

- l'identification des textes d'opinion : consiste à identifier et localiser les parties de textes porteuses d'opinion dans une collection textuelle, c.-à-d. classer les textes ou les parties de texte selon (objectifs ou subjectifs) ;
- le résumé d'opinion : consiste à rendre l'information rapidement et facilement accessible en mettant en avant les opinions exprimées et les cibles de ces opinions présentes dans un texte.
- la classification d'opinion : son objectif est d'attribuer une étiquette au texte selon l'opinion qu'il exprime (positive, négative et neutre).

### 4.2. L'analyse des sentiments

Est la partie du texte Mining qui essaye de définir les opinions, sentiments et attitudes présente dans un texte ou un ensemble de texte.

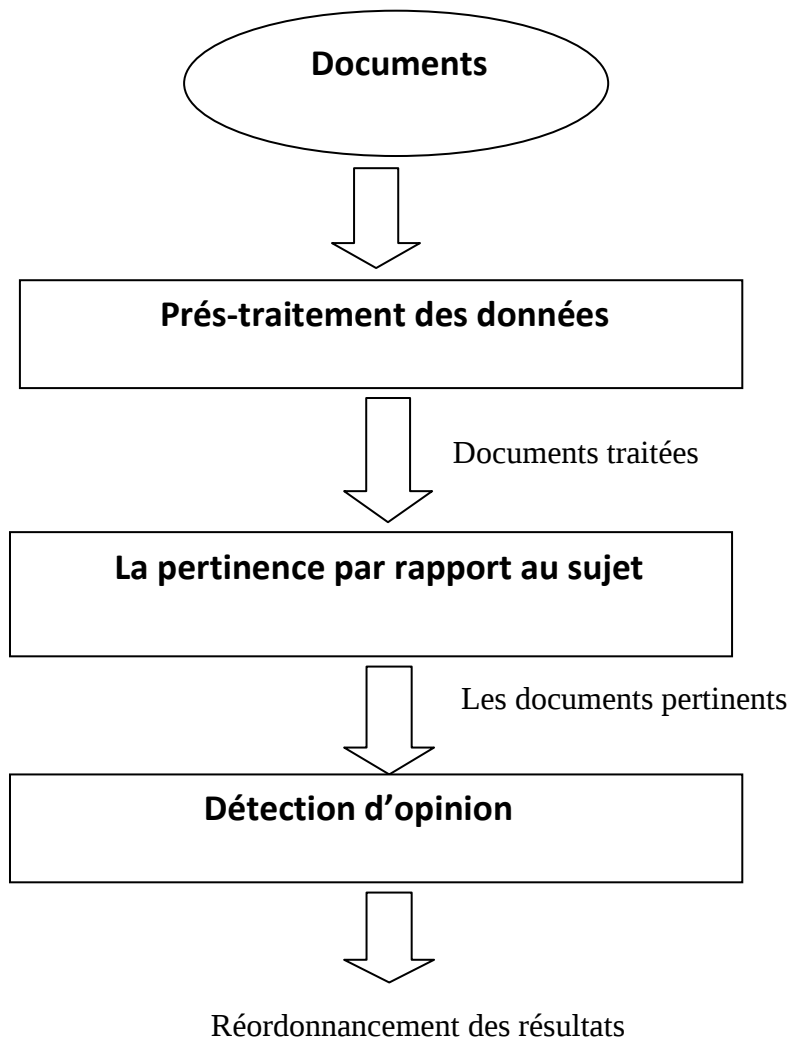
Développée essentiellement depuis les années 2000, elle est particulièrement utilisée en marketing pour analyser par exemple les commentaires des internautes ou les comparatifs et tests des blogueurs ou encore les réseaux sociaux : une grande part de la littérature sur le sujet concerne par exemple les tweets. Mais elle peut également être utilisée pour sonder l'opinion publique sur un sujet, pour chercher à caractériser les relations sociales dans les forums.

L'analyse de sentiment demande bien plus de compréhension de la langue que l'analyse de texte et la classification par sujet. En effet, si les algorithmes les plus simples considèrent uniquement les statistiques de fréquence d'apparition des mots, cela se révèle en général insuffisant pour définir l'opinion dominante dans un document, surtout lorsque le contenu est court comme des messages dans un forum.

Aujourd'hui, la fouille d'opinion et l'analyse des sentiments font partie du même domaine de recherche.

### 4.3. Processus de la fouille d'opinions

La figure suivante montre les étapes de la fouille d'opinions :



**Figure 3 : Processus de fouille d'opinions.**

#### **a. Acquisition et prétraitement des données**

Dans cette phase, les textes sont prétraités linguistiquement en éliminant les mots vides et les mots qui n'apportent aucune information importante, ainsi qu'une analyse lexicale pour enlever les mots qui ont le même sens. Dans cette étape, un étiquetage grammatical est fait (pour reconnaître l'adjectif, l'adverbe, le verbe, etc.), les grammaires de dépendances sont utilisées pour structurer la phrase de manière hiérarchique.

### **b. La pertinence par rapport au sujet**

Cette phase consiste à étudier la pertinence des documents par rapport à un sujet donné. Les documents sont classés, et généralement les 1000 premiers documents les plus pertinents sont extraits, et sont utilisés pour l'étape suivante.

### **c. La détection d'opinions**

La détection d'opinions utilise plusieurs méthodes pour le but de réordonner les documents pertinents selon un score d'opinion.

## **4.4. Domaines d'application**

L'importance de la détection d'opinion est présente dans plusieurs domaines, mais la plus grande application de l'Opinion Mining reste dans le monde du business, et du politique.

### ➤ **Politique**

Les documents portant des textes rapportant des débats politiques, (un corpus de réactions à des propositions de lois) sont de taille très importante. Ils contiennent environ dix fois plus de textes que celui des critiques d'un produit. Ce qui implique la variété énorme des jugements politiques dans ces articles. Ces derniers se retrouvent au cœur des débats avec l'avènement des médias sociaux (et plus spécialement sur le Web). Les acteurs politiques ont également suivi cette tendance, tel qu'avant de promulguer une nouvelle loi, les politiciens essayent de récolter l'avis des internautes sur cette loi. Il est intéressant de connaître aussi l'avis des internautes sur tel homme politique pour une élection présidentielle [1].

### ➤ **Marketing**

#### • **Coté entreprises**

L'opinion Mining permet au fournisseur d'un produit ou d'un service d'avoir plus de connaissances sur les consommateurs, pour anticiper leurs besoins et leurs attentes afin de tenter d'améliorer la qualité du produit/service et d'augmenter les profits.

Les professionnels du marketing digital, du service client et de la communication cherchent aujourd'hui l'outil idéal qui leur permettant d'analyser automatiquement la teneur et la tonalité des conversations publiées sur leurs marques sur le web [2].

Le domaine de la fouille d'opinion est devenu un enjeu majeur pour toute entreprise désireuse de mieux comprendre ce qui plait et déplaît à ses clients [3].

#### • **Coté clients**

Le client peut de son côté :

- ✓ Donner son opinion.

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

- ✓ S'inspirer des sentiments et opinions d'autres clients sur le produit auquel il s'intéresse et profiter ainsi d'une aide à la décision [4].
- ✓ Comparer les produits avant de les acquérir [3].
- ✓ Ne pas lire tous les commentaires concernant un produit donné, il suffit de voir le pourcentage positifs associé à ce produit.

Notre travail s'insère dans ce cas de figure.

### 4.5. Difficultés de la fouille d'opinions et de l'analyse des sentiments

Ici la classification de textes d'opinion (positive, négative ou neutre) a pour objectif l'analyse des sentiments exprimés. Cependant, cette classification se heurte à quelques difficultés :

- ✓ Difficulté due à l'ambiguïté de certains mots positifs ou négatifs selon les contextes et qui ne peut pas toujours être levée [5].
- ✓ Difficulté due au contexte : la nécessité d'une bonne analyse syntaxique du texte ; analyse qui peut se révéler particulièrement difficile dans des cas de coordination entre plusieurs parties d'une phrase [4].

Par exemple «les acteurs du film ont bien joué, la musique est bonne mais je n'ai pas aimée l'histoire», l'opinion de la dernière partie de la phrase est la plus importante.

- ✓ Difficulté due au langage naturel pour l'analyse automatique de sentiments selon les contextes intentionnels, pour lesquels l'expression d'opinion n'est pas un vrai sentiment. C'est le cas dans une phrase comme : « Je croyais que la C'est-à-dire était un beau pays. »
- ✓ Difficulté due aux structures syntaxiques et sémantiques d'une phrase et l'expression de l'opinion qu'elle véhicule [4].

Par exemple « l'histoire du film est intéressante mais les acteurs étaient mauvais ». Dans ce cas la polarité de la deuxième partie est opposée à la première [1].

- ✓ Difficulté due à l'analyse de la phrase par « paquets de mots ». Les deux phrases suivantes contiennent les mêmes paquets de mots sans pour autant exprimer les mêmes sentiments. La première phrase contient un sentiment positif alors que la deuxième est négative : « Je l'ai apprécié pas seulement à cause de ... », « Je l'ai pas apprécié seulement à cause de ... » [4].

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

- ✓ Difficulté due au vocabulaire qu'on utilise pour exprimer une opinion. Il diffère d'une personne à une autre, comme par exemple un anglo-saxon lorsqu'il exprime ses sentiments utilise des mots bien représentatifs de ce qu'il ressent contrairement aux personnes qui ne connaissent pas ou peu sa langue.
- ✓ Difficulté due à l'emploi d'une thématique. Une même thématique peut être utilisée dans classes et peut exprimer une toute autre signification, comme par exemple «un vieux vin», ou un «vieil habit».
- ✓ difficulté due au langage qu'utilisent les internautes pour s'exprimer. Les ponctuations ne sont pas forcément utilisées pour marquer les fins de phrases, des mots spécifiques sont utilisés tel que : «ha ha ha», «Goood», «super».
- ✓ difficulté de déterminer un lexique adapté à l'analyse de l'ensemble des textes d'opinion.
- ✓ difficulté à trouver une association entre l'opinion et la requête. En effet une opinion dans un document ne porte pas forcément sur la requête considérée.

### 5. Les approches de l'analyse d'opinion

En général, il existe trois types d'approches utilisées :

- Les approches symboliques
- Les approches statistiques
- Les approches hybrides.

#### 5.1. Les approches symboliques

La principale tâche dans cette approche est la conception de lexiques (ou dictionnaires) d'opinions. L'objectif de ces lexiques est de répertorier le plus de mots possible qui sont porteurs d'opinions. Ces mots permettent ensuite de classer les textes en deux (positif et négatif) ou trois catégories (positif, négatif et neutre). A titre d'exemple, dans [6], les auteurs décrivent un système appelé Opinion Observer, qui permet de comparer des produits concurrents en utilisant les commentaires écrits par les internautes. Pour cela, ils ont une liste prédéfinie de termes désignant des caractéristiques de produits. Lorsque l'une de ces caractéristiques est présente dans un texte, le système extrait les adjectifs proches dans la phrase. Ces adjectifs sont ensuite comparés aux adjectifs présents dans leur lexique d'opinions et ainsi, une polarité est attribuée à la caractéristique du produit.

##### a) Construction des lexiques d'opinion

Pour construire des lexiques d'opinions (sentiments), trois méthodes peuvent être appliquées [7], [8] :

- La méthode manuelle.
- La méthode basée sur les corpus.
- La méthode basée sur les dictionnaires.

La méthode manuelle, qui consiste à remplir le lexique de mots d'opinions sans l'aide d'outils particuliers, demande un effort important en terme du temps. La sélection des mots porteurs d'opinion et le choix de leur polarité se font donc uniquement par l'expertise humaine.

On peut supposer qu'une part non négligeable de subjectivité peut alors entrer en jeu et peut entraîner certaines erreurs de classification. Quelle que soit la méthode de construction des lexiques, une première étape de classification manuelle est nécessaire.

En effet, une première liste de mots et d'expressions doit être identifiée. Ces mots sont appelés les germes (ang. Seed words). Cet ensemble de mots est ensuite utilisé afin de découvrir, répertorier et classer d'autres mots et expressions porteurs d'opinions.

L'une des solutions permettant d'agréments cet ensemble de mots est donc l'utilisation de corpus de textes. Dans [9], l'auteur propose la méthode suivante :

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

Afin de déterminer la polarité de mots ou expressions non classés, il compte le nombre de fois où ces mots ou expressions apparaissent dans le corpus à côté de mots ou expressions présents dans la liste de germes. C'est ce qu'on appelle, chercher les "cooccurrences" de mots. Un mot apparaissant plus souvent à côté de mots positifs sera donc classé dans la catégorie positif et inversement.

Dans [10], les auteurs ont proposé une méthode similaire, mis à part qu'ils utilisent la probabilité qu'un mot non classé soit proche d'un mot classé afin de mesurer la force de l'orientation du premier nommé. D'autres méthodes [11], [12] utilisent également cette hypothèse dans le but de compléter les lexiques d'opinion : deux mots ou groupes de mots ayant un fort taux d'apparitions communes sont supposés posséder une forte proximité sémantique.

Une autre méthode basée sur le corpus permettant de compléter le lexique d'opinions, consiste à utiliser les conjonctions de coordination présentes entre un mot déjà classé et un mot non classé [13], [14]. Par exemple, si la conjonction AND sépare un mot classé positif dans le lexique d'opinion et un mot non classé, alors le mot non classé sera considéré comme étant positif. A l'inverse, si la conjonction BUT sépare un mot classé positif et un mot non classé, alors le mot non classé sera considéré comme étant négatif. Les conjonctions utilisées sont les suivantes : AND, OR, BUT, EITHER-OR, et NEITHER-NOR.

La dernière solution la plus utilisée est une méthode basée sur les dictionnaires. Elle consiste à utiliser des dictionnaires de synonymes et antonymes existants tels que WordNet [15], [16] afin de déterminer l'orientation sémantique de nouveaux mots. Par exemple, Liu [17], [18] utilisent ce dictionnaire afin de prédire l'orientation sémantique des adjectifs. Dans WordNet, les mots sont organisés sous forme d'arbres.

Afin de déterminer la polarité d'un mot, ils traversent les arbres de synonymes et d'antonymes du mot et, s'ils trouvent un mot déjà classé parmi les synonymes, ils affectent la même polarité au mot étudié, ou bien la polarité opposée s'ils trouvent un mot déjà classé parmi les antonymes. S'ils ne croisent aucun mot déjà classé, ils réitèrent l'expérience en partant de tous les synonymes et antonymes, et ce jusqu'à rencontrer un mot d'orientation sémantique connue. Cette méthode peut toutefois entraîner un certain nombre d'erreurs car un grand nombre de mots peuvent avoir différentes significations.

Des analyses plus poussées sont également faites. Afin de mesurer plus précisément la force de l'opinion exprimée dans une phrase, un moyen utilisé est l'extraction des adverbes associés aux adjectifs. Pour ce faire, Benamara et al. [19] proposent une classification des adverbes en cinq catégories : les adverbes d'affirmation, de doute, de forte intensité, de faible intensité et

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

les adverbes de négation et minimiseurs. Un système d'attribution de points en fonction de la catégorie de l'adverbe permet de calculer la force exprimée par le couple adverbe-adjectif.

Le tableau suivant présente quelques exemples de mots classés dans les catégories définies par Benamara et al. [19].

| Classe                       | Mots                                                                                         |
|------------------------------|----------------------------------------------------------------------------------------------|
| Adverbes d'affirmation       | Absolutely, entirely, fully, certainly, fairly, exactly, totally, enough...                  |
| Adverbes de doute            | Even, possibly, roughly, apparently, seemingly...                                            |
| Adverbes de forte intensité  | So, really, very, pretty, highly, extremely, much, well, too, quite...                       |
| Adverbes de faible intensité | Only, a little, almost, a bit, little, rather, nearly, barely, scarcely, weakly, slightly... |
| Négation et minimiseurs      | Not, never, less, no...                                                                      |

**Tableau : Exemple de catégories d'adverbes**

Toutes ces catégories d'adverbes ne sont pas toujours prises en compte car elles n'ont pas la même importance au niveau de la prédiction de note. Les négations paraissent logiquement être des termes importants à détecter, en plus des adjectifs et des verbes, car ils permettent d'inverser la polarité d'une phrase. Pour traiter cet aspect, Das et Chen [20] proposent par exemple d'ajouter des mots dans le dictionnaire d'opinion comme "like-NOT" qui sont utilisés lors de la détection d'une couple like-négation. Les négations peuvent alors être *not*, *don't*, *didn't* ou *never* pour le cas de l'anglais et *ne* pas pour le français. Mais le problème de la détection de la négation reste un problème très ouvert, les méthodes existantes n'étant pas réellement convaincantes.

En effet, l'expression de la négation peut être faite sans l'utilisation de ce type de mots. Les expressions telles que « *je suis contre* » ou encore « *je m'oppose* » peuvent également permettre d'inverser la polarité du reste de la phrase. La difficulté est également due aux différentes façons d'utiliser la négation comme le sarcasme ou l'ironie. L'interprétation de la négation nécessite alors une analyse syntaxique qui est un traitement très coûteux en temps de calcul et pas forcément très efficace suivant la qualité du texte analysé.

### **b. Classification des textes grâce aux lexiques**

Une fois que les mots porteurs d'opinion sont répertoriés dans les lexiques, la dernière étape consiste à déterminer la polarité d'une phrase à l'aide de ces mots. La solution la plus simple consiste à compter le nombre de mots positifs et le nombre de mots négatifs présents. S'il y a une majorité de termes positifs, la phrase est déclarée positive. A l'inverse, si les mots négatifs sont les plus nombreux, la phrase est déclarée négative. Les phrases possédant autant de mots négatifs que de mots positifs peuvent être déclarées neutres [10], ou encore, la polarité de la phrase peut dépendre du dernier mot d'opinion parcouru [17]. On peut encore extraire plusieurs opinions dans une même phrase et les associer aux caractéristiques discutées [21].

### **Limites de l'analyse de sentiment à partir des lexiques**

- les dictionnaires affectent une tonalité positive ou négative à un mot, sans tenir compte du contexte, c'est-à-dire du texte environnant, comme des paramètres de la communication située ;
- les dictionnaires ont tendance à éliminer les mots à valence ambiguë a priori ;
- le traitement des expressions ambiguës reste à faire et demande de faire appel à d'autres principes et à d'autres techniques ;
- lorsque la négation n'est pas prise en compte (ce qui peut paraître étonnant mais qui existe encore, par exemple dans les méthodes basées sur les « sacs de mots », qui calculent des fréquences d'occurrence dans un texte), le score de polarité peut être complètement faussé ;
- ces dictionnaires ne permettent pas de traiter des figures de rhétorique qui peuvent changer entièrement la valence des expressions.

### **5.2. Les approches statistiques**

Dans ce type d'approches, les mots sont généralement considérés comme des variables équivalentes. L'aspect sémantique n'est alors pas pris en compte. Les méthodes les plus utilisées pour la classification d'opinions sont les méthodes de classification supervisée. Ce type de méthodes consiste à construire un modèle de classification à l'aide d'exemples. Des exemples sont des données dont on connaît déjà la classe. On parle dans ce cas de données classées ou étiquetées.

Voici une définition plus formelle du problème de la classification supervisée :

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

Soit  $X$  un ensemble de données,  $Y$  un ensemble d'étiquettes (ou classes) et  $D$  un ensemble des représentations des données. Soit  $d : X \rightarrow D$ , une fonction qui associe à chaque donnée  $x \in X$  une représentation  $d(x) \in D$  et  $S \subset D \times Y$  un ensemble de données étiquetées  $(d(x), y)$  avec  $y \in Y$ . La classification supervisée consiste à construire un classifieur en s'appuyant sur l'ensemble  $S$ , qui permette de prédire la classe de toute nouvelle donnée  $x \in X$ , représentée par  $d(x) \in D$ .

D'après cette définition, quatre éléments distincts entrent en jeu dans la classification supervisée :

- La représentation des données (l'ensemble  $D$ ) ;
- Les étiquettes ou classes de prédiction (l'ensemble  $Y$ ) ;
- Les données étiquetées, qui constituent le corpus d'apprentissage (l'ensemble  $S$ ) ;
- Le classificateur ou le prédicteur.

### a) Le Corpus d'apprentissage

Tout d'abord, il est nécessaire de posséder des exemples (données étiquetées) afin de construire le "corpus d'apprentissage". Ce corpus ayant un impact direct sur l'apprentissage du modèle et par conséquent, sur les résultats de la classification, il est nécessaire que les exemples soient les plus représentatifs possibles de l'ensemble des données. En classification d'opinions, ou plus généralement en classification de textes, les corpus étudiés sont assez restreints car l'étiquetage des exemples est souvent effectué à la main. Ceci entraîne un coût élevé et ne permet donc pas l'obtention d'un gros corpus d'apprentissage. Cependant, les données présentes sur les sites Web 2.0, qui peuvent souvent être étiquetées par les auteurs eux-mêmes ou par d'autres internautes, permettent aujourd'hui de traiter des corpus beaucoup plus conséquents.

### b) Les classes de prédiction

Concernant le choix des classes, il est généralement imposé par le corpus d'apprentissage utilisé et la tâche visée. On parle de classification binaire lorsque le nombre de classes  $|Y|$  est égal à 2. Il peut naturellement être supérieur, mais il ne faut pas oublier qu'un nombre de classe élevé augmente le taux d'erreurs. En classification d'opinions, le nombre de classes de prédiction choisi est généralement de deux (aime, n'aime pas) ou trois (aime, avis neutre, n'aime pas) [21], [22].

### c) La représentation

Les documents textuels sont des suites de caractères. Afin d'en permettre l'exploitation, il est nécessaire de représenter les textes numériquement. En fouille de textes, les documents sont généralement représentés sous leur forme vectorielle dite en "sac de mots". Les mots sont alors considérés comme des objets distincts non ordonnés. Cette représentation consiste tout d'abord à sélectionner des variables représentant les dimensions d'un vecteur puis à représenter chaque document sur ce vecteur. En classification de texte, la représentation implique donc une étape préliminaire appelée segmentation, qui consiste à découper le texte afin de sélectionner les variables. Une fois les variables acquises, plusieurs documents peuvent être représentés sur le même espace vectoriel. On obtient alors une matrice dite de Salton [24]. Nous présentons tout d'abord en quoi consiste exactement la segmentation, puis nous énumérons les types de représentations les plus utilisées en recherche d'information qui sont également celles utilisées en classification d'opinions.

#### ➤ La segmentation

Un document est un ensemble de caractères. La segmentation consiste à découper cet espace de caractères afin d'obtenir un espace de variables. Cet espace de variables est appelé vocabulaire et différents choix concernant sa construction sont possibles suivant le délimiteur choisi. On peut considérer les marques de ponctuation, afin de découper le document en phrases, ou encore les espaces associés à la ponctuation afin d'obtenir des mots. Ce sont d'ailleurs les mots qui sont les variables les plus utilisées en classification d'opinion, mais d'autres choix existent. On peut par exemple limiter la taille des variables à un nombre  $n$  de caractères, on parle alors de  $n$ -grammes de lettres. Les  $n$ -grammes peuvent également être formés de mots. Ces  $n$ -grammes de mots peuvent être construits selon leur ordre dans la phrase, afin de conserver un sens sémantique. Par exemple, dans la phrase « je n'aime pas ce film », le bigrammes (ou 2-grammes) « aime pas » sera considéré. On peut également construire les  $n$ -grammes de mots selon qu'ils apparaissent dans la même phrase par exemple. Une fois le vocabulaire sélectionné, différents choix sont possibles concernant la représentation du document sur le vecteur.

#### ➤ La représentation binaire

Cette représentation est la moins coûteuse en temps de calcul. Elle consiste à indiquer, pour un document, quels mots du vocabulaire sont présents (valeur égale à 1) et quels mots sont absents (valeur égale à 0) [25], [26], [27].

### ➤ La représentation fréquentielle

Cette représentation est une extension de la représentation binaire qui prend en compte le nombre d'occurrences des variables dans chaque document. Un texte est donc représenté par un vecteur dont chaque composante correspond à la fréquence des variables dans le texte [28].

### ➤ La représentation fréquentielle normalisée

Cette représentation consiste à normaliser les vecteurs de représentation des textes par la longueur des textes. C'est-à-dire que les fréquences des variables obtenues avec la représentation fréquentielle sont remplacées par la proportion des variables dans chaque document. La proportion s'obtient en divisant la fréquence de la variable par la taille du document.

### ➤ La représentation TF-IDF

La mesure statistique TF-IDF permet l'évaluation d'une variable dans un document à la fois par sa fréquence dans le document concerné, mais également par sa présence dans tous les autres documents du corpus.

La valeur TF (Term Frequency) correspond à la fréquence de la variable  $v$  dans le document  $d$  normalisée par la taille de  $d$ . La valeur IDF (Inverse Document Frequency) mesure l'importance de la variable  $v$  dans l'ensemble du corpus en calculant le logarithme de l'inverse du nombre de documents contenant  $v$  [29], [30].

### d) Le classificateur

Un classificateur est une procédure qui, à l'aide d'un ensemble d'exemples, produit une prédiction de la classe de toute donnée. Beaucoup de méthodes de classification supervisée existent et beaucoup d'entre elles ont été testées pour la classification d'opinions.

On peut citer les arbres de décision, les réseaux de neurones, la régression logistique, les règles de décision ainsi que des méthodes combinant différents classificateurs comme les systèmes de votes ou les algorithmes de Boosting. Toutefois, les méthodes les plus présentes dans la littérature, et qui semblent également être les plus performantes sur les textes, sont les Machine à Vecteurs de Support (SVM) [25], [26], [27], [28], [29], [31] et les classificateurs Naïfs Bayésiens (NB) [10], [27], [28].

### ➤ Les Machines à Vecteurs de Support

Les Machines à Vecteurs de Support, appelées encore Séparateurs à Vaste Marge, sont des classificateurs binaires très populaires en classification de textes. Considérons tout d'abord le cas où les données sont linéairement séparables. Nommons positif et négatif les deux classes  $y \in Y$ . Si le problème est linéairement séparable, les individus positifs sont séparables des

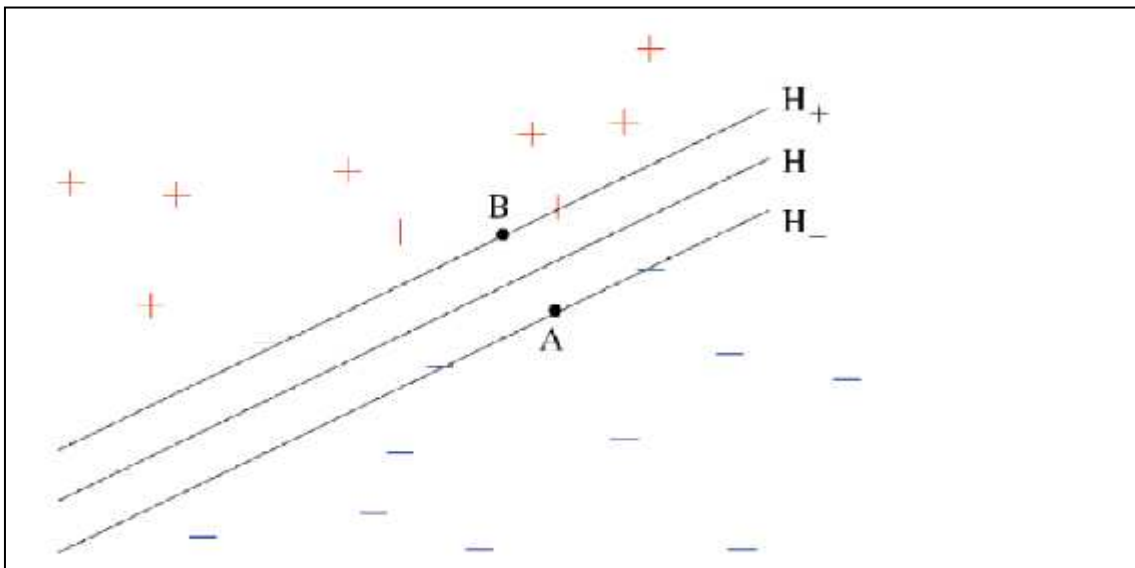
## Chapitre II : La fouille d'opinion et d'analyse des sentiments

individus négatifs par un hyperplan  $H$ . Notons  $H^+$  l'hyperplan parallèle à  $H$  qui contient l'individu positif le plus proche de  $H$ , respectivement  $H^-$  pour l'individu négatif. Une machine à vecteurs de support linéaire recherche alors l'hyperplan qui sépare les données de manière à ce que la distance entre  $H^+$  et  $H^-$  soit la plus grande possible.

Cet écart entre les deux hyperplans  $H^+$  et  $H^-$  est appelé la « marge ». Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple.

Dans le cas où les données ne sont pas linéairement séparables, l'idée des SVM est de changer l'espace de représentation. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace, généralement plus grand, appelé « espace de re-description ».

Dans le cas d'une classification multi-classes, plusieurs méthodes sont possibles, la plus connue étant le principe de l'Un-contre-tous. Il consiste à apprendre tout d'abord un modèle à l'aide de la première classe face à toutes les autres, puis de la deuxième face à toutes les autres et ainsi de suite. On obtient ainsi plusieurs classificateurs que l'on applique tous lors de la classification d'une nouvelle donnée. La classe attribuée à cette donnée est alors celle obtenant le meilleur score.



**Figure 4 : Exemple d'hyperplan ( $H$ ) séparant les individus appartenant à la classe (+) et ceux appartenant à la classe (-)**

### ➤ Les classifieurs Naïfs Bayésiens

Le principe d'un classificateur naïf bayésien consiste à maximiser la probabilité  $\Pr(y|d)$ , soit la probabilité d'occurrence de la classe de prédiction  $y$  connaissant la représentation de la nouvelle donnée  $x$  (on suppose donc ici  $d = d(x) = (d_1, d_2, \dots, d_n)$ ), et ce pour toutes les classes  $y \in Y$  et toutes les composantes qui interviennent dans la définition de l'espace de représentation  $D$ . Pour cela, on fait appel la règle de Bayes.

- **Règle de Bayes.** Soient  $A$  et  $B$  deux évènements. La règle de Bayes dit alors que la probabilité de l'évènement  $A$  sachant l'évènement  $B$  ( $\Pr(A|B)$ ) peut se calculer à l'aide des probabilités des évènements  $A$  et  $B$  ( $\Pr(A)$  et  $\Pr(B)$ ) et connaissant la probabilité de l'évènement  $B$  sachant l'évènement  $A$  ( $\Pr(B|A)$ ) par la formule suivante :

$$\Pr(A|B) = \Pr(B|A) \Pr(A) / \Pr(B)$$

**Application à la classification.** En appliquant la règle de Bayes à la problématique de la classification, on obtient l'équation suivante :

$$\Pr(y|d) = \Pr(d|y) \Pr(y) / \Pr(d)$$

Les probabilités de l'expression de droite doivent être estimées, à l'aide du corpus d'apprentissage  $S$ , afin de calculer la quantité qui nous intéresse, soit  $\Pr(y|d)$  :

- $\Pr(y)$  est la probabilité d'observer la classe  $y$  ;
- $\Pr(d)$  est la probabilité d'observer la représentation  $d$  ;
- $\Pr(d|y)$ , la vraisemblance de l'évènement « observer la représentation  $d$  » si  $s \in S$

est de classe  $y$ . Ce terme est plus difficile à estimer que le précédent.

En pratique, on ne s'intéresse qu'au numérateur, le numérateur ne dépendant pas de  $y$ . Concernant  $\Pr(d|y)$ , l'hypothèse habituellement faite dans ce type de classifieurs est que toutes les composantes  $d_i$  sont indépendantes, ce qui permet de calculer facilement la probabilité globale d'une classe connaissant une donnée. Cette non-dépendance des composantes correspond à « l'hypothèse de Bayes naïve ». On considère donc que  $\Pr(d|y) = \prod_i \Pr(d_i|y)$ . Maximiser  $\Pr(y|d)$  revient donc à maximiser  $(\prod_i \Pr(d_i|y))\Pr(y)$ . Les  $\Pr(d_i|y)$  sont évalués par les fréquences observées dans les exemples de l'ensemble  $S$ .

### 5.3. Les approches hybrides

Ces approches combinent les points forts des deux approches précédentes, cette combinaison a donné des résultats plus précis que chacune des approches employées séparément.

Plusieurs types d'hybridation sont présents dans la littérature. Dans tous les cas, elles utilisent des éléments décrits dans les deux approches précédentes. On peut distinguer parmi ces approches trois méthodes distinctes :

- La linguistique au service de l'apprentissage automatique.
- L'apprentissage automatique au service de la linguistique.
- Une fusion a posteriori des résultats des deux approches.

### **a) La linguistique au service de l'apprentissage automatique**

Une approche hybride consiste à utiliser les outils linguistiques afin de préparer le corpus avant de classer les textes à l'aide de l'apprentissage supervisé. Wilson et al. [31] préparent les données à l'aide d'outils de TALN afin de sélectionner un vocabulaire d'opinions. Ces mots présélectionnés sont ensuite utilisés comme vecteurs de représentation des textes pour les outils d'apprentissage supervisé. Trois algorithmes d'apprentissage sont comparés : BoostTexter [32], Ripper [33] et *SVMlight* [34]. Nigam et Hurst [35] utilisent des techniques provenant du Traitement Automatique des Langues afin de détecter dans les textes les mots et expressions porteurs d'opinions et ajoutent des marques dans le texte (traits grammaticaux et + ou – pour opinion positive et opinion négative). Ils utilisent ensuite l'apprentissage automatique pour classer les textes selon leur opinion générale.

### **b) L'apprentissage automatique au service de la linguistique**

Une autre façon de combiner les méthodes est d'utiliser les techniques d'apprentissage automatique dans le but de construire les dictionnaires d'opinions nécessaires à l'approche linguistique.

Hatzivassiloglou et McKeown [13] présentent une méthode ayant pour objectif de définir l'orientation sémantique des adjectifs pour la construction du dictionnaire d'opinions. Ils extraient tout d'abord tous les adjectifs du corpus à l'aide d'un analyseur syntaxique, puis utilisent un algorithme de clustering afin de classer les adjectifs selon leur polarité. Riloff et Wiebe [36] combinent les deux approches afin de répertorier les expressions porteuses d'opinion qui, selon eux, sont plus riches que des mots pris individuellement. Turney et Littman [37] utilisent une approche statistique pour classer un plus grand nombre de types de mots selon leur polarité : adjectifs, verbes, noms...

### **c) Une fusion a posteriori des résultats des deux approches**

Une dernière façon d'utiliser conjointement les approches basées sur la linguistique et celles basées sur l'apprentissage automatique est de construire plusieurs types de

## Chapitre II : La fouille d'opinion et d'analyse des sentiments

---

classificateurs afin de combiner leurs résultats, soit par des systèmes de vote, soit par un algorithme d'apprentissage [38]. Dans le cas d'un système de vote, on attribue généralement des poids suivant les performances des classificateurs utilisés. Si les résultats diffèrent, on conserve alors le résultat du classificateur ayant le poids le plus fort. Dans le cas de l'utilisation d'un algorithme d'apprentissage, les poids sont calculés automatiquement selon l'indice de confiance attribué par chaque classificateur à ses propres résultats.

### 6. Conclusion

Au cours de ce chapitre, nous avons donné un aperçu de ce qui est la fouille d'opinion et l'analyse des sentiments, ainsi que leur domaine d'application et leurs approches. Nous dédions dans le chapitre suivant à l'analyse et conception.

# **Chapitre III :**

## **Analyse et conception**

## 1. Introduction

Avant toute réalisation d'une application informatique, il convient de suivre une démarche méthodologique et rigoureuse pour la planifier et la concevoir, en mettant en évidence tous les objectifs tracés pour la bonne élaboration du projet souhaité.

Pour bien concevoir notre application, on a pensé, en premier lieu, à préciser bien nos objectifs. En second lieu, on décrira et en suivra un processus de développement d'application Web en utilisant le langage de modélisation très répandu UML.

## 2. L'objectif de travail

L'objectif de notre travail est de développer un nouveau système permettant d'analyser et classer les opinions des clients de l'entreprise ENIEM. Ainsi notre application devrait assurer les points suivants :

- Offrir une interface conviviale et simple à utiliser.
- L'accessibilité, à tout moment, au questionnaire.
- Permettre à l'administrateur de gérer toute l'application.

## 3. La démarche d'élaboration de notre projet

La création d'applications informatiques n'est pas chose aisée. Il faut un Processus rigoureux qui permet de mener à terme des projets d'une façon fiable.

La démarche d'élaboration de notre projet s'appuie sur la modélisation UML (Unified Modeling Language), ses avantages sont :

- Langage simple à comprendre et à utiliser.
- Langage très expressif qui couvre toutes les perspectives nécessaires au développement de tels systèmes.
- Langage adapté à la modélisation des systèmes, depuis les systèmes informatiques d'entreprises jusqu'à les applications distribuées basées sur le web, en passant par les systèmes temps réel embarqués.
- Il permet d'avoir une meilleure façon d'exprimer les besoins, en offrant des modèles qui facilitent le dialogue entre le concepteur et le futur utilisateur de système.

### 4. Le processus de développement d'applications web

Pour une meilleure organisation de travail, le processus de développement adopté est le suivant :

Spécification des acteurs, cas d'utilisation et les scenarios.

- **Analyse** : C'est au travers des activités d'analyse que l'on peut définir à partir de l'examen des besoins un modèle conceptuel du système à développer. C'est-à-dire d'un modèle capable d'être concrétisé sous forme logiciel. Les artefacts de cette étape incluent les classes et les diagrammes de séquences.
- **Conception** : Cette phase traite les artefacts produits durant l'étape d'analyse et elle est consacrée à la réalisation des cas d'utilisation et la définition de la vue logique du système. Les activités de la conception se focalisent sur les diagrammes de classe. Enfin, le modèle de conception peut être directement traduit en code source.
- **Implémentation** : A ce stade de processus, les cas d'utilisation sont terminés. Le problème a été analysé en profondeur et nous avons défini une conception appropriée à l'architecture. Nous pouvons alors entreprendre l'étape d'implémentation qui inclut les activités suivantes :
  - La traduction de la conception en code source et en composants.
  - La réalisation des tests unitaires.

### 5. Les étapes d'élaboration de notre projet

#### 5.1. Analyse

Nous avons cité dans le chapitre précédent les différentes approches d'analyse d'opinion. Pour notre solution, nous avons opté pour une approche symbolique basée sur le principe de SentiWordNet[39]. Cette approche consiste à mettre en place un dictionnaire de mot polarisés qui est alimenté au fur et à mesure avec des mots relatifs au domaine commercial de l'électroménager. Ce dictionnaire est tenu à jour par un expert du domaine, il se charge d'affecter des polarités positives, négatives et objectives à chaque nouveau mot.

##### 5.1.1. Acteurs

- **Un acteur** : Un acteur représente un rôle joué par une entité externe (utilisateur humain, dispositif matériel ou autre système) qui interagit directement avec le système étudié.

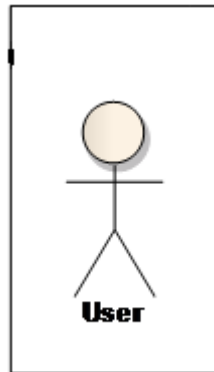
Un acteur peut consulter et/ou modifier directement l'état du système, en émettant et/ou en recevant des messages susceptibles d'être porteurs de données.

- **Représentation graphique d'un acteur**

## Chapitre III : Analyse et conception

---

La représentation graphique d'un acteur en UML est l'icône appelée stickman, avec le nom de l'acteur sous le dessin.



**Figure5 : Exemple de représentation d'un acteur**

- **Détermination des acteurs de notre système**

Les principaux acteurs que nous avons distingués et qui peuvent interagir avec notre système sont :

**Administrateur** : C'est la personne chargée du maintien et de la mise à jour de l'application.

**Fonctionnaire eniem** : c'est la personne chargée d'analyser et classer les opinions des clients.

**Expert linguiste** : c'est la personne chargée d'effectuer les polarités des nouveaux mots.

**Client** : Toute personne chargée de répondre aux questionnaires.

### 5.1.2. Les cas d'utilisation

Un cas d'utilisation constitue un moyen de recueillir et de décrire les besoins des acteurs du système.

Dans notre cas on distingue les cas suivants :

➤ **Cas d'utilisation relatif à l'administrateur :**

- S'authentifier.
- Gérer les fonctionnaires eniem.
- Gérer les clients.
- Gérer l'expert linguiste.

### ➤ Cas d'utilisation relatif au fonctionnaire eniem :

- S'authentifier.
- Gérer les familles de services.
- Gérer les services.
- Gérer les types de questions.
- Gérer les questions.
- Gérer les questionnaires.
- Paramétrage des coefficients d'opinion.
- Analyser l'opinion.
- Synthétiser l'opinion.

### ➤ Cas d'utilisation relatif a l'expert linguiste :

- S'authentifier.
- Gérer le dictionnaire.

### ➤ Cas d'utilisation relatif au client :

- S'authentifier.
- Remplir le questionnaire.

#### 5.1.3. Spécification des scenarios

Chaque résultat possible de l'exécution d'un cas d'utilisation et appelé scénario.

Un scénario est un chemin logique traversant le cas d'utilisation, un scénario est une instance de cas d'utilisation.

Les scénarios décrivant chacune des tâches définies auparavant sont récapitulés dans le tableau suivant :

| Acteurs | Taches               | Scénarios                                                                               |
|---------|----------------------|-----------------------------------------------------------------------------------------|
| Client  | T1 : s'authentifier. | S1 : cliquer sur le lien « client »<br>S2 : saisir le nom et le prénom.<br>S3 : valider |

## Chapitre III : Analyse et conception

|                            |                                    |                                                                                                                                                                             |
|----------------------------|------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                            | T2 : remplir le questionnaire      | S4 : cliquer sur le lien « questionnaire »<br>S5 : remplir le questionnaire<br>S6 : valider                                                                                 |
| <b>Fonctionnaire eniem</b> | T3 : s'authentifier.               | S7 : cliquer sur le lien «Fonctionnaire eniem »<br>S8 : saisir le login et le mot de passe.<br>S9 : valider                                                                 |
|                            | T4 : gérer les familles de service | S10 : visualiser la liste des Famille de service<br>S11 : modifier une famille de service<br>S12 : ajouter une famille de service<br>S13 : supprimer une famille de service |
|                            | T5 : gérer les services            | S14 : visualiser la liste des services<br>S15 : modifier un service<br>S16 : ajouter un service<br>S17 : supprimer un service                                               |
|                            | T6 : gérer les types de question   | S18 : visualiser la liste des types de question<br>S19 : ajouter un type de question<br>S20 : supprimer un type de question                                                 |

## Chapitre III : Analyse et conception

|                       |                                              |                                                                                                                                      |
|-----------------------|----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
|                       | T7 : gérer les questions                     | S21 : visualiser la liste des Questions<br>S22 : modifier une question<br>S23 : ajouter une question<br>S24 : supprimer une question |
|                       | T8 : gérer le questionnaire                  | S25 : créer un Questionnaire<br>S26 : activer un Questionnaire                                                                       |
|                       | T9 : Paramétrage des coefficients d'opinion. | S27 : Visualiser la liste des coefficients.<br>S28 : modifier un coefficient.                                                        |
|                       | T10 : synthétiser l'opinion                  | S29 : récupérer la valeur de l'opinion formatée<br>S30 : récupérer la valeur de l'opinion libre<br>S31 : calculer la valeur finale   |
|                       | T11 : analyser les opinions                  | S32 : calculer l'opinion formatée<br>S33 : calculer l'opinion libre                                                                  |
| <b>Administrateur</b> | T12 : s'authentifier.                        | S34 : cliquer sur le lien « Administrateur »<br>S35 : saisir le login et le mot de passe.<br>S36 : valider                           |
|                       | T13 : gérer le fonctionnaire eniem.          | S37 : visualiser la liste des fonctionnaires eniem<br>S38 : ajouter un                                                               |

|  |                                |                                                                            |
|--|--------------------------------|----------------------------------------------------------------------------|
|  |                                | fonctionnaire eniem<br>S39 : supprimer un fonctionnaire eniem              |
|  | T14 : gérer client             | S40 : visualiser la liste des clients<br>S41 : ajouter un client           |
|  | T15 : gérer l'expert linguiste | S42 : visualiser la liste des experts linguiste<br>S43 : ajouter un expert |

### 5.1.4. Les diagrammes

#### ➤ Diagramme de cas d'utilisation générale

Il décrit le comportement du système du point de vue utilisateur sous la forme d'actions et de réactions.

Il existe deux concepts fondamentaux dans la modélisation par cas d'utilisation :

- ✓ Les acteurs qui utilisent le système
- ✓ Les cas d'utilisation qui représentent l'utilisation du système par les acteurs.

Chaque cas d'utilisation indique une fonctionnalité du système déclenchée par un acteur externe du système. Ce genre de diagramme permet de mettre en place et de comprendre les besoins des utilisateurs.

Les cas d'utilisation peuvent être structurés. En plus de la relation de communication, qui consiste au déclenchement d'un cas d'utilisation par un acteur, nous pouvons citer deux types de liens ou relations qui sont les plus utilisés : le lien d'utilisation et le lien d'extension.

- ✓ **Le lien d'utilisation** : ce lien nommé « utilise » (ou « include » en anglais) indique que le cas d'utilisation source contient le comportement décrit dans le cas d'utilisation destination.
- ✓ **Le lien d'extension** : Ce lien indique que le cas d'utilisation source « étend » (en anglais « Extend ») ou précise le cas d'utilisation destination.

## ❖ Diagramme du cas d'utilisation générale

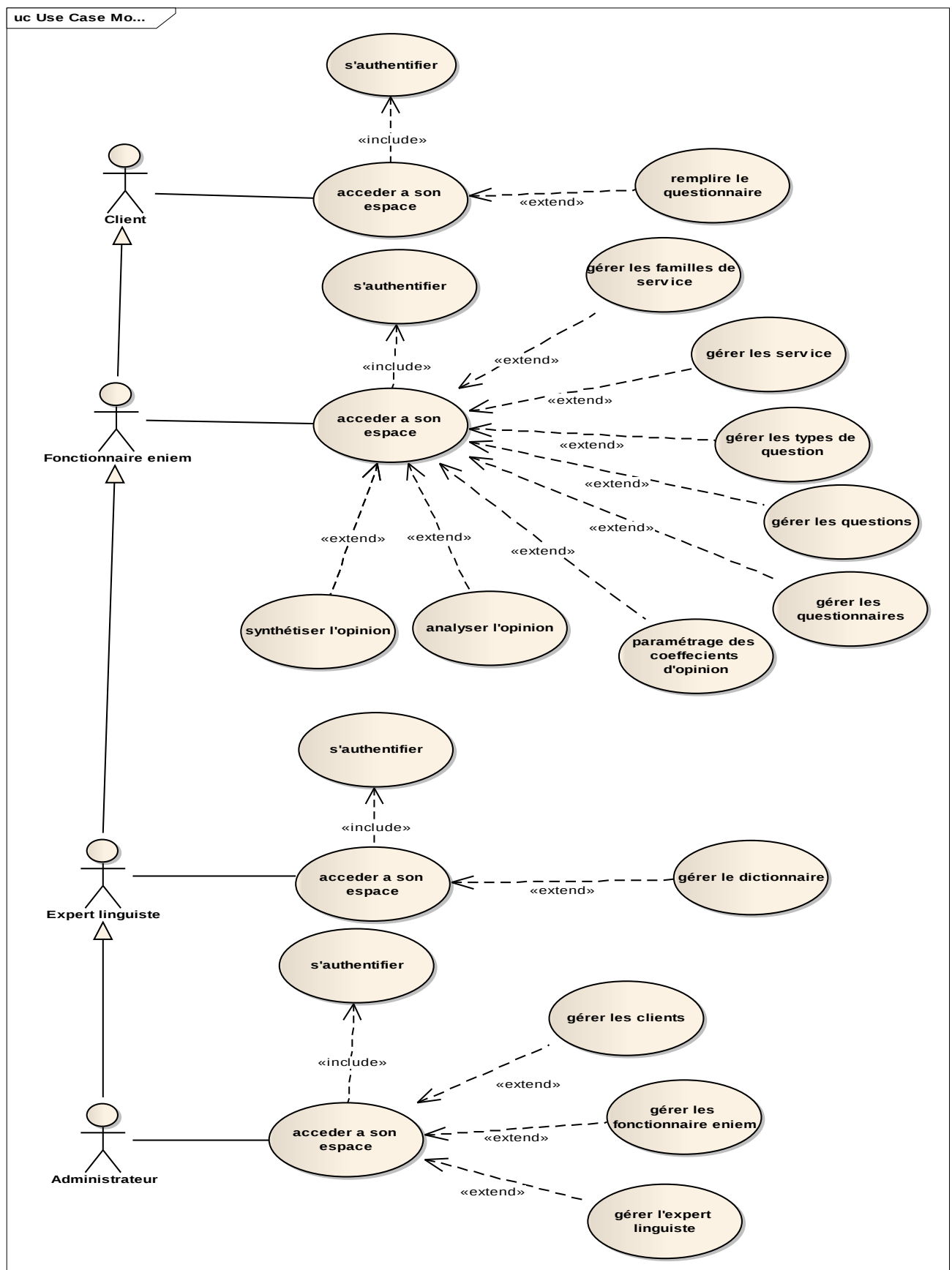


Figure 6 : Diagramme de cas d'utilisation générale

Dans cette première étape nous avons présenté les besoins que doit remplir notre système.

Cette étude a permis de synthétiser le cas d'utilisation modélisant les besoins des utilisateurs du système.

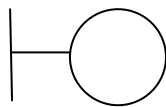
### 5.2. Vue dynamique du système

#### 5.2.1. Introduction

Dans cette partie nous fournissons une spécification complète des besoins issus des cas d'utilisations, ceci nous permet d'établir un modèle d'analyse détaillant chaque cas d'utilisation et montrant l'aspect dynamique du système grâce a des diagrammes de séquences qui décrivent les interactions entre les acteurs du système et l'application qu'on doit développer.

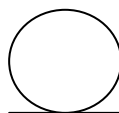
Les classes d'analyse peuvent être répertoriées dans trois catégories d'objets :

- ✓ **Les objets de type interface** : Ils représentent l'interface entre l'acteur et les systèmes tels des écrans de saisie ou des pages Web complètes. Ils sont modélisés par l'icône suivante :



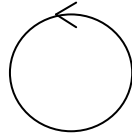
- ✓ **Les objets de type entités**

Ce sont des objets décrits dans le cas d'utilisation mais qui lui survivront les commandes, les produits sont des objets dont les instances peuvent apparaitre dans de nombreux cas d'utilisation, ils sont modélisés par l'icône suivante :



### ✓ Les objets de type contrôle

Ils représentent des processus, des activités du système tel un calcul ou une recherche, ils dirigent les activités des objets entités et interface, ils sont modélisés par l'icône suivante :



### 5.2.2. Diagrammes de séquence

Comme nous avons un nombre important de cas d'utilisation nous nous sommes contentés de représenter que quelques cas :

- ✓ Authentification.
- ✓ Ajouter une famille de service
- ✓ Supprimer un fonctionnaire eniem
- ✓ Ajouter les polarités des mots
- ✓ Ajouter un fonctionnaire eniem
- ✓ Modifier un client
- ✓ Créer un questionnaire
- ✓ Activer un questionnaire

### ❖ Diagramme de séquence du cas d'utilisation « Authentification »

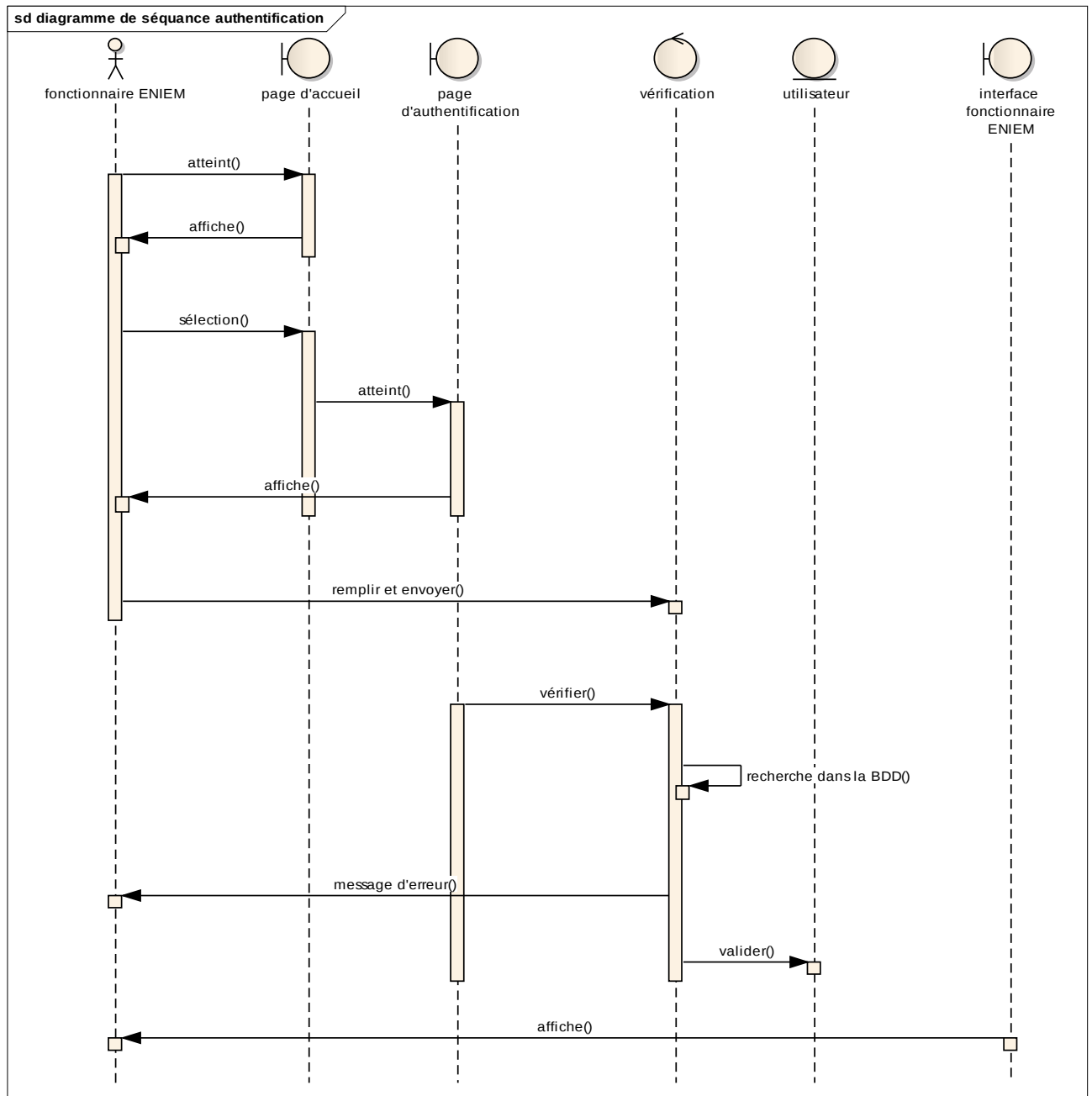


Figure 7 : Diagramme de séquence du cas d'utilisation « Authentification »

### Description :

- Le fonctionnaire eniem atteint le site, le système lui affiche la page d'accueil.
- Il clique sur le lien fonctionnaire eniem, la page d'identification s'affiche.
- Le fonctionnaire eniem saisie son identifiant et son mot de passe il clique sur « connexion».
- Affichage de l'interface fonctionnaire eniem sinon affichage d'un message d'erreur.

### ❖ Diagramme de séquence du cas d'utilisation « Ajouter une famille de service »

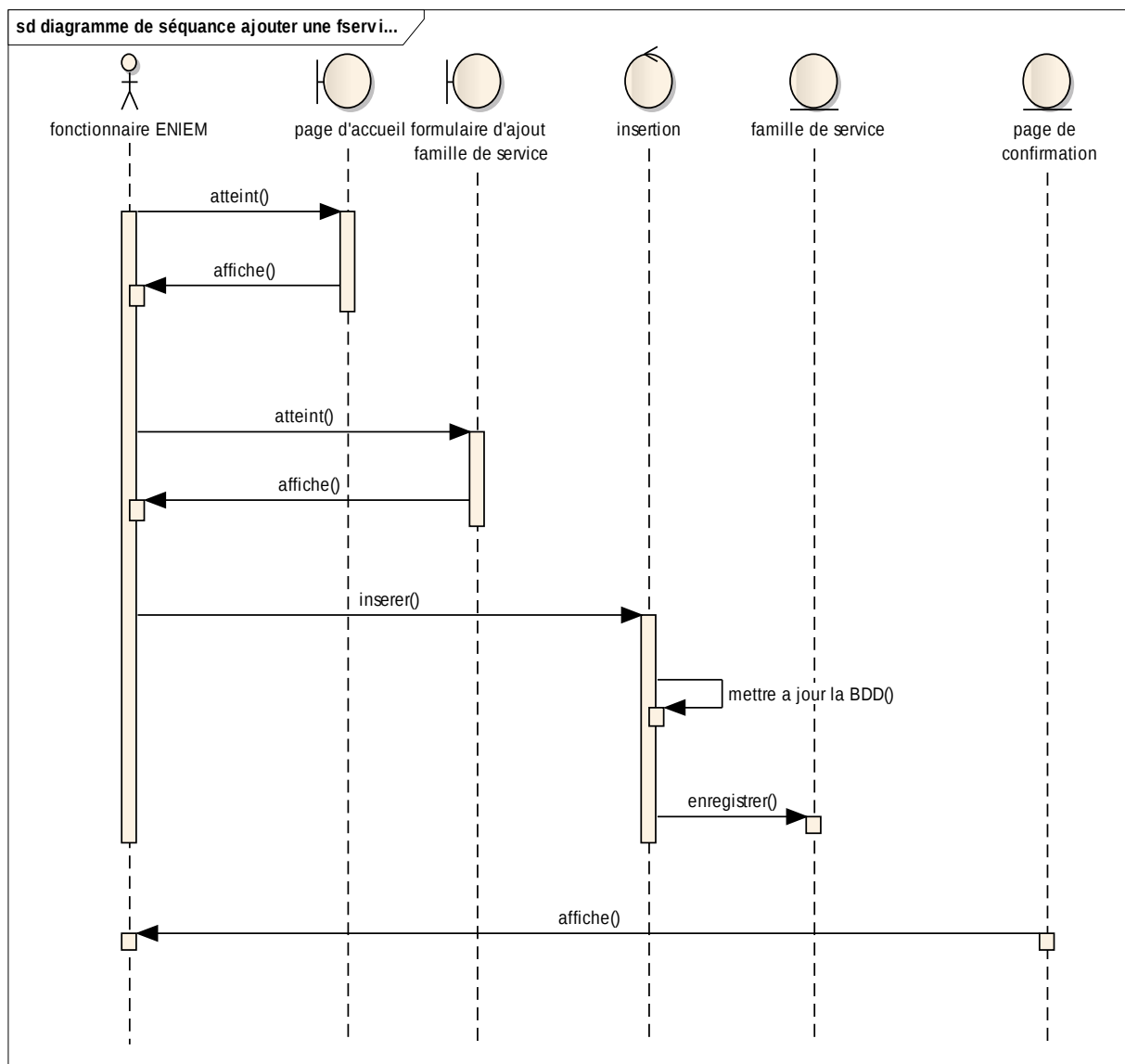


Figure 8 : Diagramme de séquence du cas d'utilisation « Ajouter une famille de service »

### Description :

- Après l'authentification le fonctionnaire eniem atteint son espace.
- Il clique sur le lien famille de service.
- Le système lui affiche la liste des familles de service et un lien pour une nouvelle famille.
- Le fonctionnaire eniem clique sur le lien.
- Le système affiche un formulaire d'ajout.
- Le fonctionnaire eniem saisie les informations nécessaire et clique sur le lien « ajouter ».
- Affichage d'un message de confirmation.

### ❖ Diagramme de séquence du cas d'utilisation « Ajouter les polarités des mots »

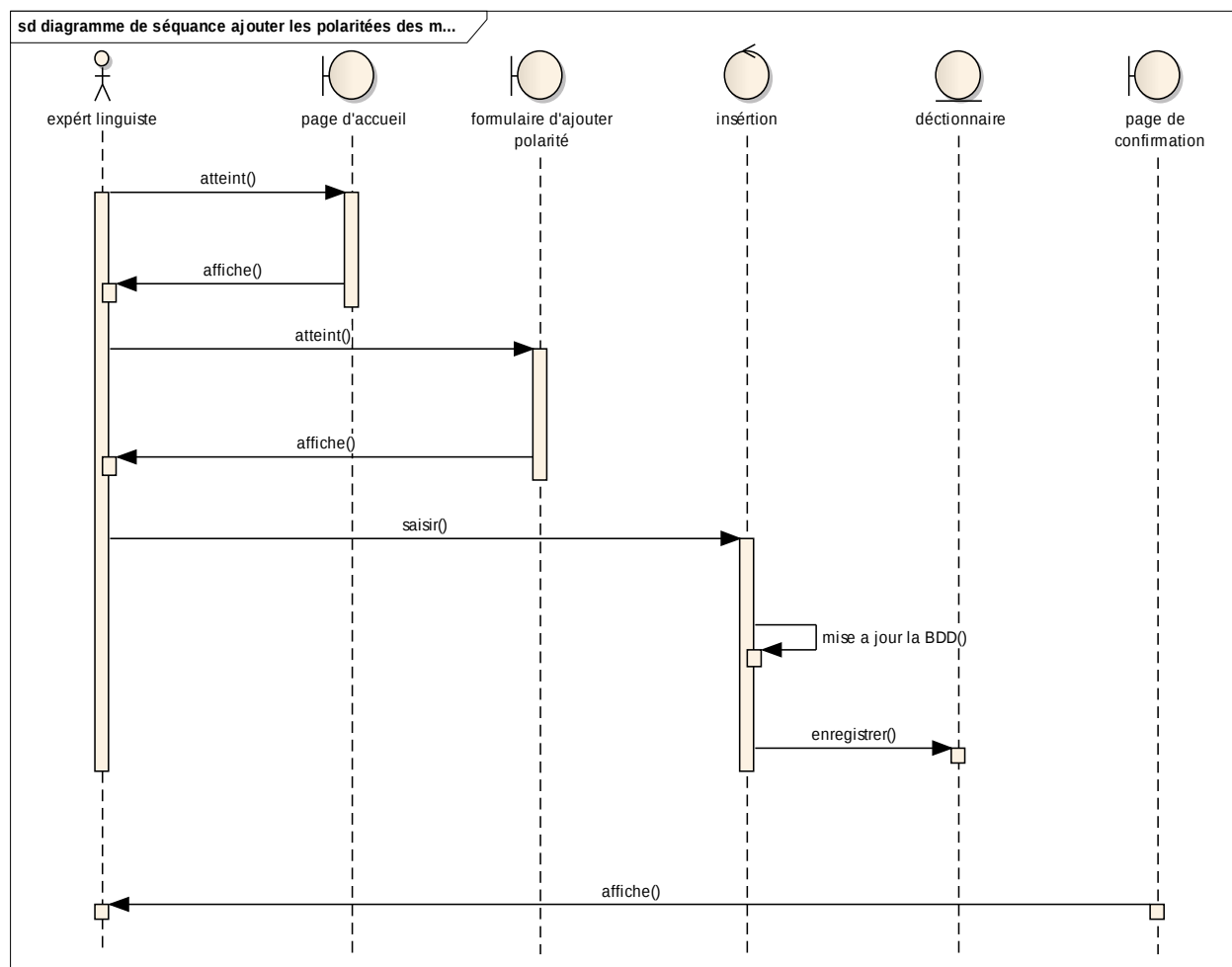


Figure 9 : Diagramme de séquence du cas d'utilisation « Ajouter les polarités des mots »

### Description :

- Après l'authentification l'expert linguiste atteint son espace.
- Le système lui affiche la liste des nouveaux mots auxquels il va effectuer les polarités.
- L'expert linguiste saisie les polarités puis clique sur « envoyer ».
- Affichage d'un message de confirmation.

### ❖ Diagramme de séquence du cas d'utilisation « Ajouter un fonctionnaire eniem »

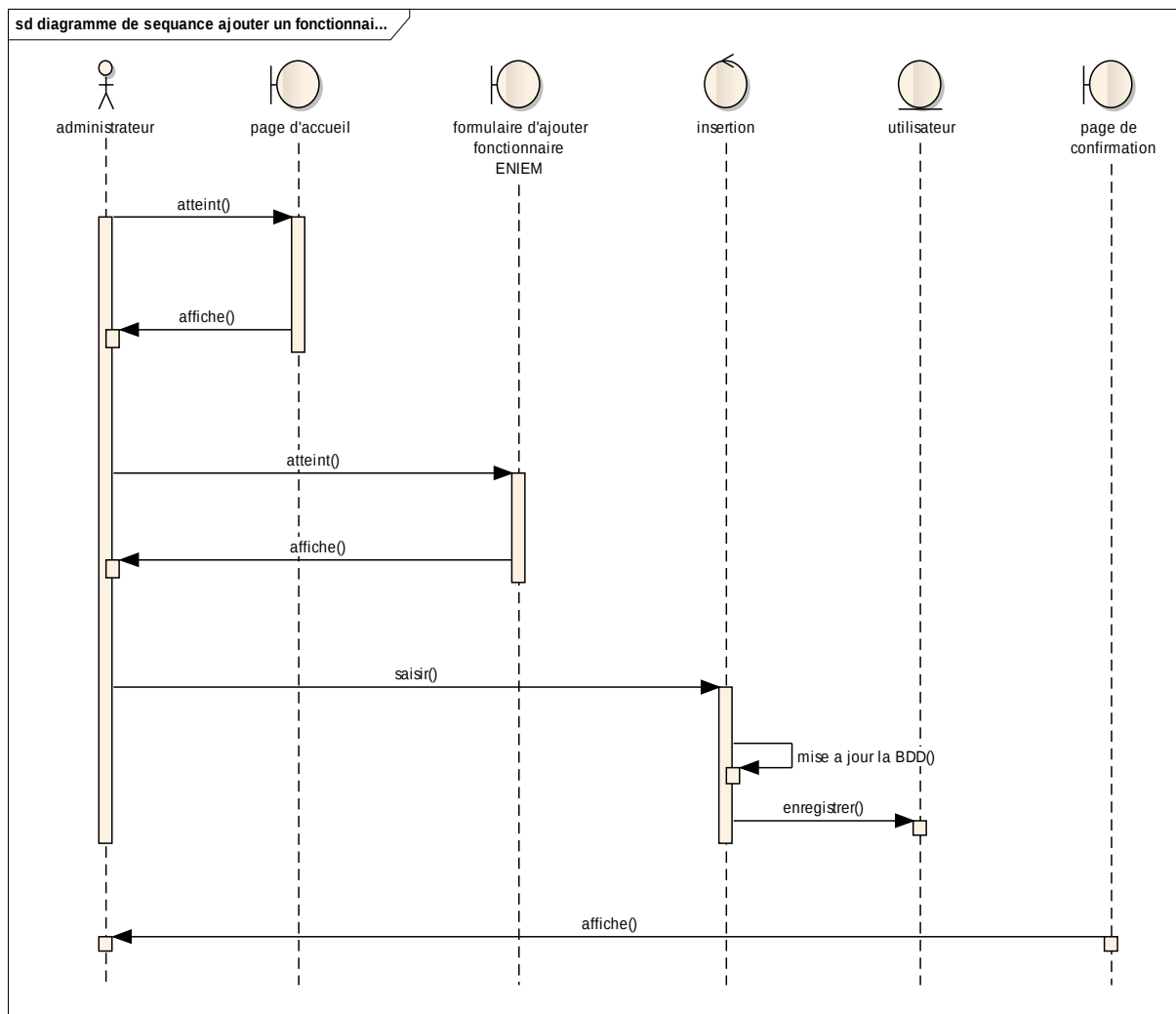
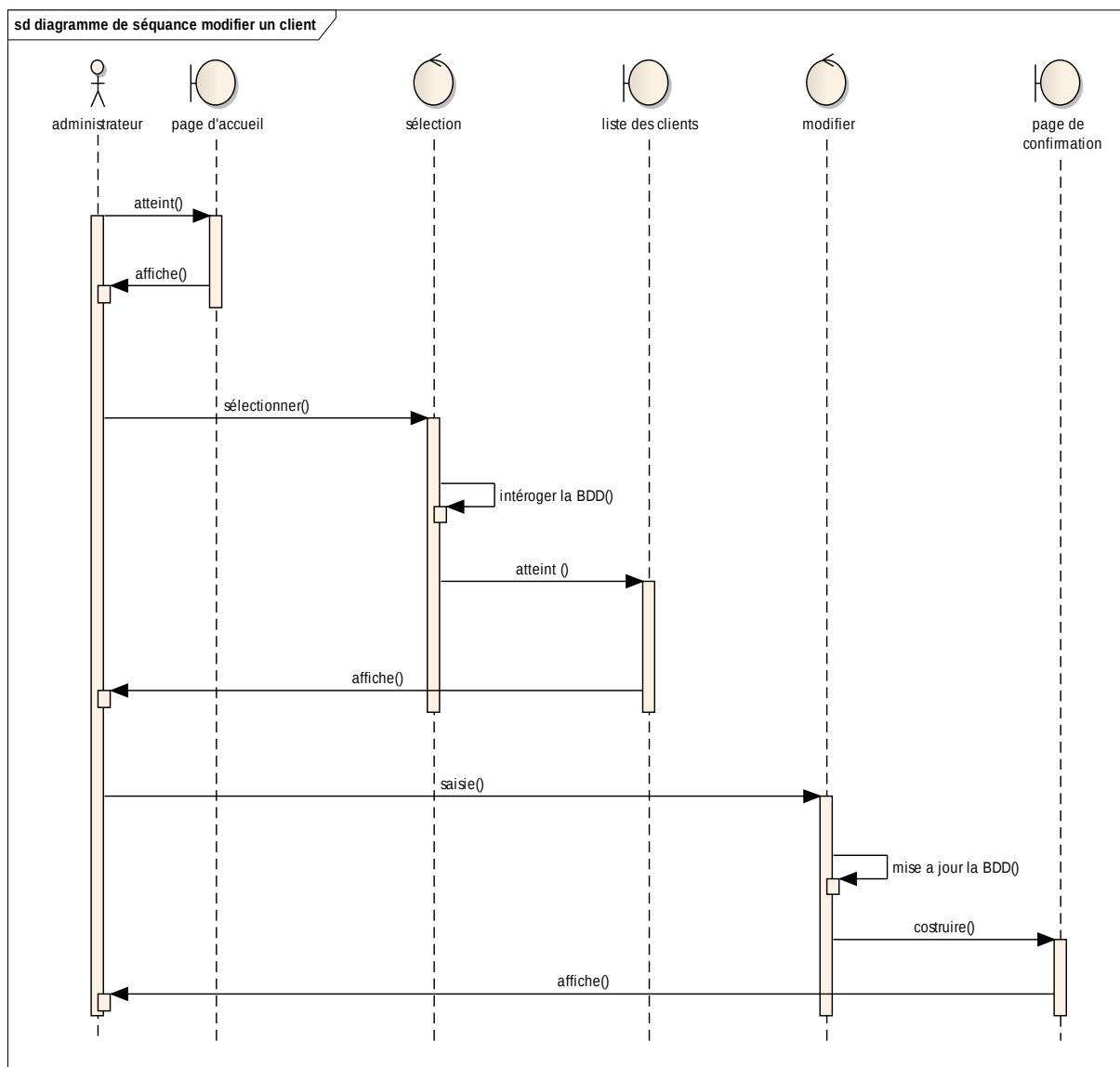


Figure 10 : Diagramme de séquence du cas d'utilisation « Ajouter un Fonctionnaire eniem »

### Description :

- Après l'authentification l'administrateur atteint son espace.
- Il clique sur le lien fonctionnaire eniem.
- Le système lui affiche la liste des fonctionnaires eniem et un lien pour ajouter un fonctionnaire eniem.
- l'administrateur clique sur le lien.
- Le système affiche un formulaire d'ajout fonctionnaire eniem.
- l'administrateur saisie les informations nécessaire et clique sur le lien « envoyer ».
- Affichage d'un message de confirmation.

### ❖ Diagramme de séquence du cas d'utilisation « Modifier un client »



**Figure 11 : Diagramme de séquence du cas d'utilisation « Modifier un client »**

### Description :

- Après l'authentification l'administrateur atteint la son espace.
- Il clique sur le lien gérer client.
- Le système lui affiche la liste des clients.
- L'administrateur visualise un client et clique sur le lien modifier.
- L'administrateur modifier ce qu'il faut modifier, puis il clique sur « envoyer ».
- Affichage d'un message de confirmation.

### ❖ Diagramme de séquence du cas d'utilisation « Supprimer un fonctionnaire eniem »

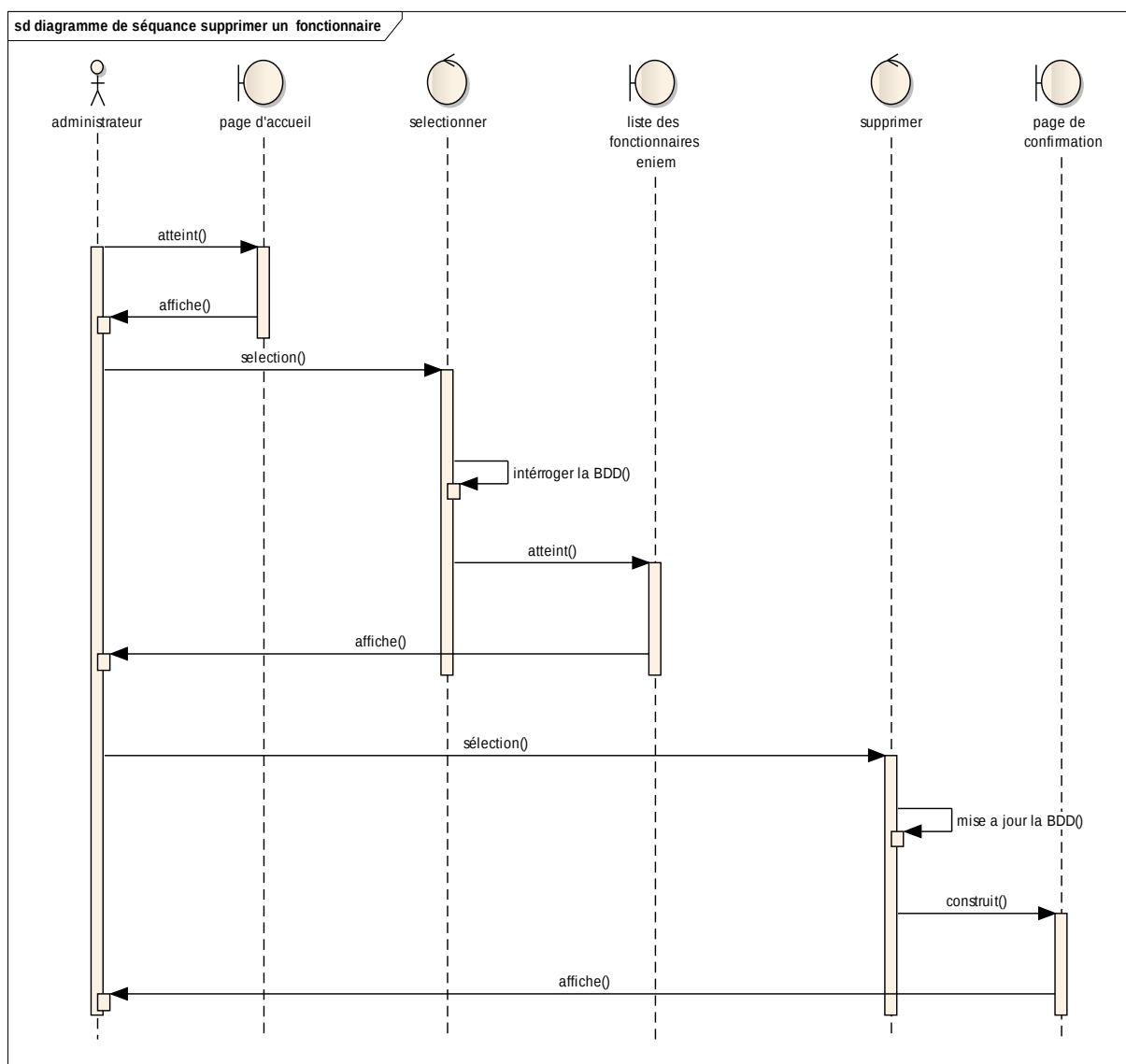


Figure 12: Diagramme de séquence du cas d'utilisation « Supprimer un fonctionnaire eniem »

### Description :

- Après l'authentification l'administrateur atteint son espace.
- Il clique sur le lien gérer fonctionnaire eniem.
- Le système lui affiche la liste des fonctionnaires eniem.
- L'administrateur visualise un fonctionnaire eniem et clique sur le lien « supprimer ».
- Affichage d'un message de confirmation.

### ❖ Diagramme de séquence du cas d'utilisation « Créer questionnaire »

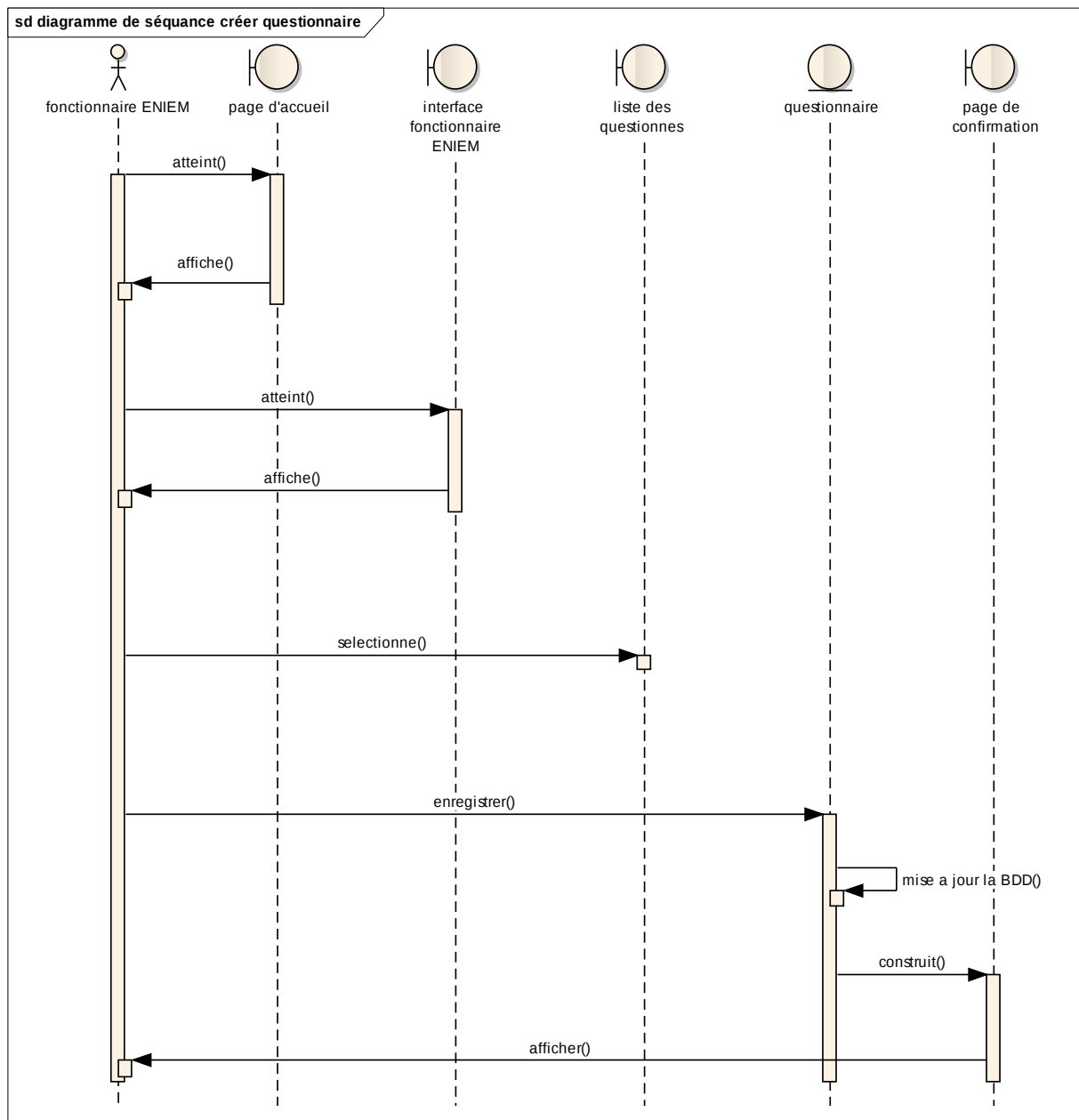


Figure 13: Diagramme de séquence du cas d'utilisation « Créer questionnaire »

### Description :

- Après l'authentification le fonctionnaire eniem atteint son espace.
- Il clique sur le lien créer questionnaire.
- Le système lui affiche la liste des questionnes.
- Le fonctionnaire eniem sélectionne les questionnes et clique sur le lien « créer ».
- Affichage d'un message de confirmation.

### ❖ Diagramme de séquence du cas d'utilisation « Activer questionnaire »

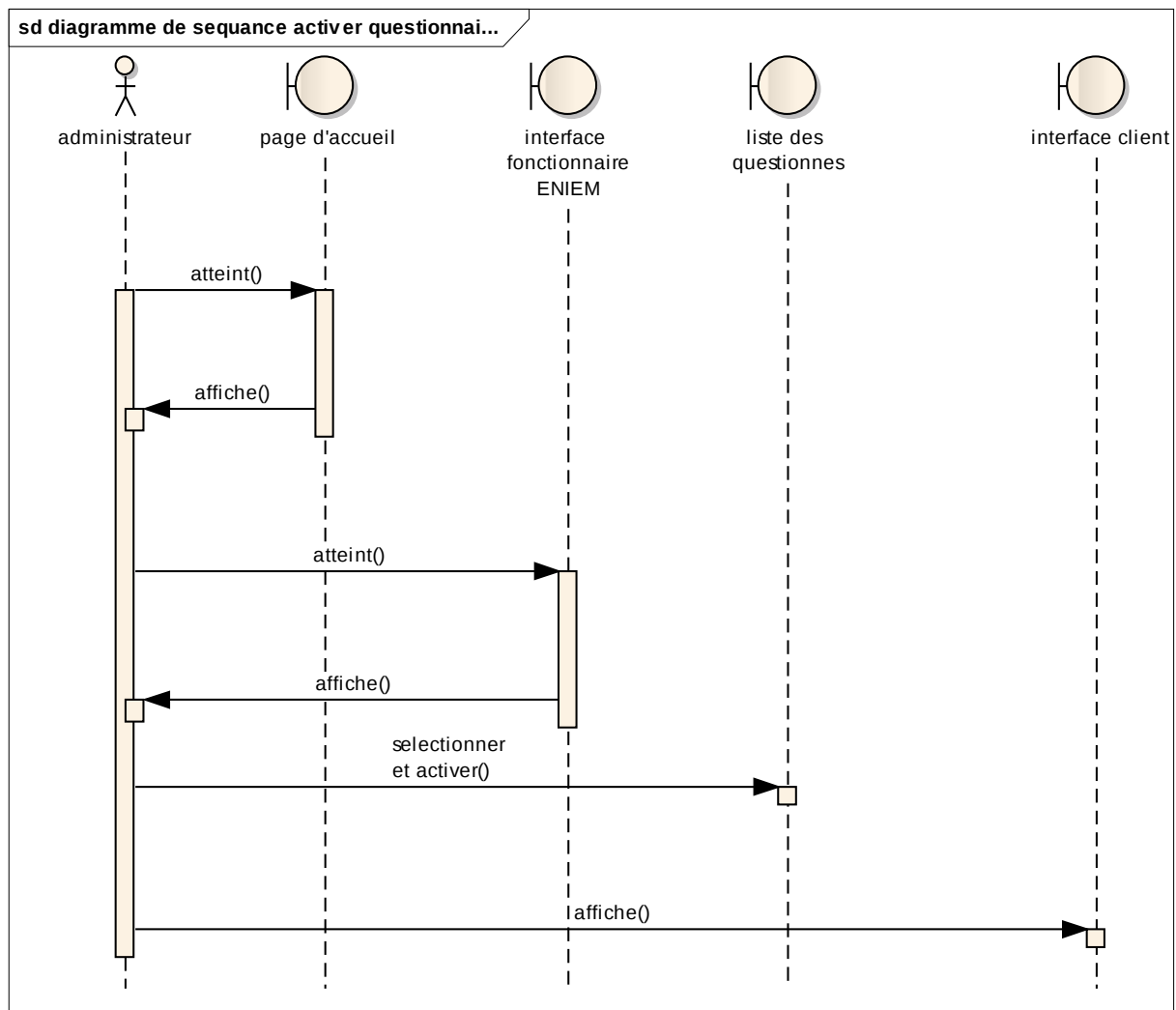


Figure 14: Diagramme de séquence du cas d'utilisation « Activer questionnaire »

### Description :

- Après l'authentification le fonctionnaire eniem atteint son espace.
- Il clique sur le lien activer questionnaire.
- Le système lui affiche la liste des questionnaires.
- Le fonctionnaire eniem sélectionne un questionnaire et clique sur le lien « activer ».
- Affichage d'un message de confirmation

### 5.3. Vue statique du système

#### 5.3.1. Introduction

Nous allons nous concentrer ici sur le diagramme de classe, ce dernier va nous servir pour construire notre base de données.

#### 5.3.2. Diagrammes de classe

Un diagramme de classe montre la structure statique d'un système. Il permet la visualisation des classes et des relations entre elles. Son but est d'expliquer ce qu'il faut réaliser dans l'application.

### 5.3.3. Diagrammes de classe globale

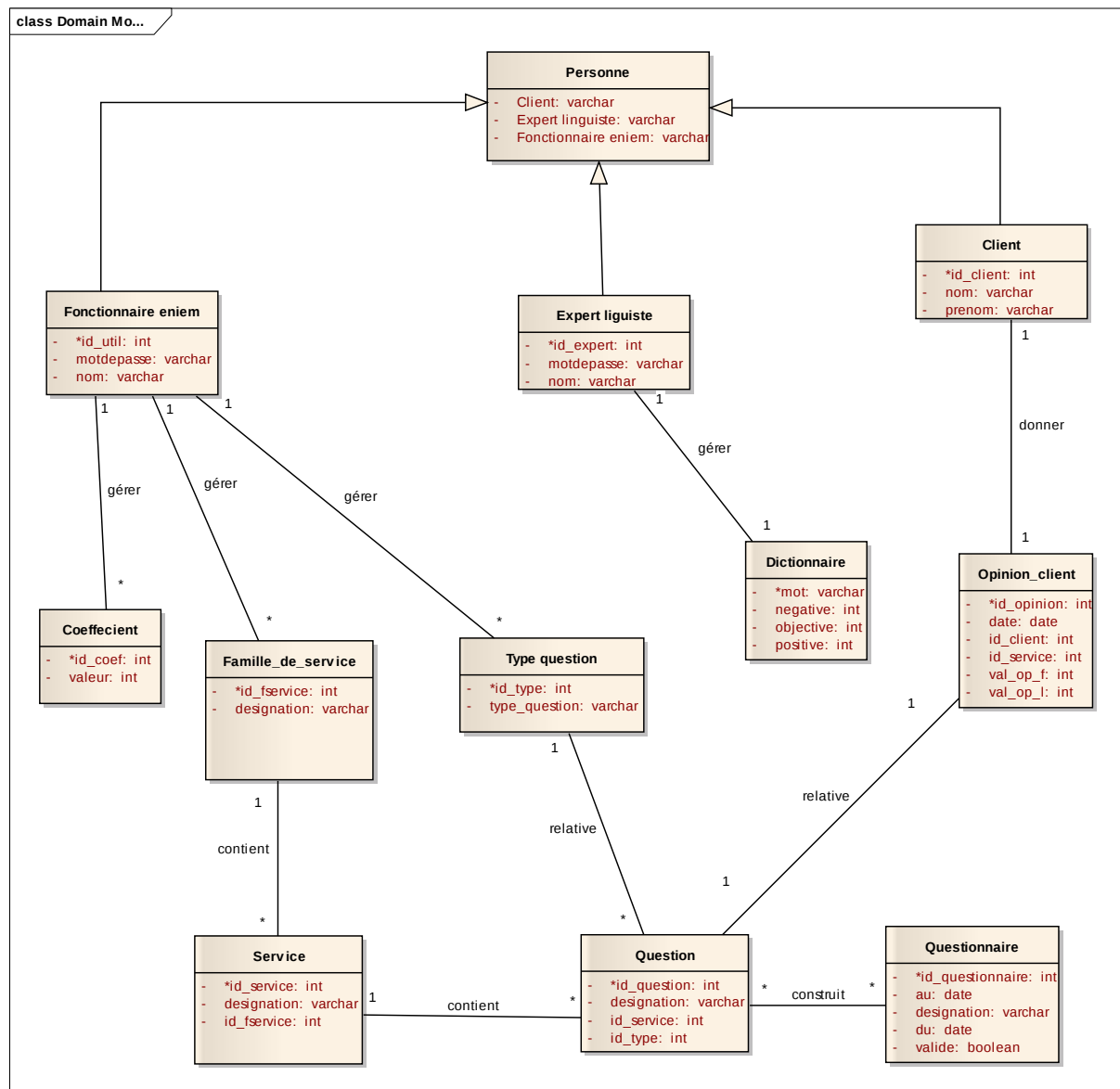


Figure 15: Diagrammes de classe globale

### 5.3.4. Conception de la base de donnée

Dans les étapes précédentes nous avons modélisés notre système avec les différents diagrammes (cas d'utilisation, séquence, classe), nous intéressons aux données nécessaires pour le fonctionnement de l'application.

Grace aux diagrammes de classes globales nous pouvons obtenir la liste des tables suivantes :

## Chapitre III : Analyse et conception

- La base de données « eniem » contient les tables suivantes :

admin,Utilisateur,client,expert,coef,dictionnaire,famille-service,ligne\_questionnaire,opinion\_client,question,questionnaire,reponse\_comm,service,temporaire,type\_de\_question.

**Table «admin»:**

| Nom du Champ      | Type de données | description                     | clé      |
|-------------------|-----------------|---------------------------------|----------|
| <b>id_admin</b>   | int(100)        | Identifiant de l'administrateur | Primaire |
| <b>Nom</b>        | Varchar(100)    | Nom de l'administrateur         |          |
| <b>Motdepasse</b> | varchar(100)    | Mot de passe                    |          |

**Table «utilisateur»:**

| Nom du Champ          | Type de données | description                        | clé      |
|-----------------------|-----------------|------------------------------------|----------|
| <b><u>id_util</u></b> | int(100)        | Identifiant de fonctionnaire eniem | Primaire |
| <b>Nom</b>            | Varchar(100)    | Nom de fonctionnaire eniem         |          |
| <b>Motdepasse</b>     | Varchar(100)    | Mot de passe                       |          |

## Chapitre III : Analyse et conception

Table «client»:

| Nom du Champ     | Type de données | description           | clé      |
|------------------|-----------------|-----------------------|----------|
| <b>id_client</b> | int(100)        | Identifiant de client | Primaire |
| <b>Nom</b>       | Varchar(100)    | Nom de client         |          |
| <b>Prenom</b>    | Varchar(100)    | Prénom de client      |          |

Table « coef »:

| Nom du Champ   | Type de données | description                | clé      |
|----------------|-----------------|----------------------------|----------|
| <b>id_coef</b> | int(100)        | Identifiant de coefficient | primaire |
| <b>Valeur</b>  | Float(100)      | Valeur de coefficient      |          |

Table «dictionnaire»:

| Nom du Champ     | Type de données | description                    | clé      |
|------------------|-----------------|--------------------------------|----------|
| <b>Mot</b>       | Varchar(100)    | Le mot                         | primaire |
| <b>positive</b>  | float (100)     | La valeur de polarité positive |          |
| <b>négative</b>  | float (100)     | La valeur de polarité négative |          |
| <b>objective</b> | float (100)     | La valeur de l'objectivité     |          |

## Chapitre III : Analyse et conception

Table «expert»:

| Nom du Champ      | Type de données | description                        | clé      |
|-------------------|-----------------|------------------------------------|----------|
| <b>id_expert</b>  | int(100)        | Identifiant de l'expert linguiste  | primaire |
| <b>Nom</b>        | Varchar(100)    | Le nom de l'expert linguiste       |          |
| <b>Motdepasse</b> | Varchar(100)    | Mot de passe de l'expert linguiste |          |

Table «famille\_service»:

| Nom du champ       | Type de données | description                          | Clé      |
|--------------------|-----------------|--------------------------------------|----------|
| <b>id_fservice</b> | int(100)        | Identifiant de la famille de service | Primaire |
| <b>désignation</b> | Varchar(100)    | La désignation de famille de service |          |

Table «ligne\_questionnaire»:

| Nom du champ            | Type de données | description                  | Clé       |
|-------------------------|-----------------|------------------------------|-----------|
| <b>id_questionnaire</b> | int(100)        | Identifiant du questionnaire | Primaire  |
| <b>id_question</b>      | int(100)        | Identifiant du questionnaire | Etrangère |

## Chapitre III : Analyse et conception

Table «opinion\_client»:

| Nom du Champ       | Type de données | description                     | Clé       |
|--------------------|-----------------|---------------------------------|-----------|
| <b>id_rep</b>      | int(100)        | Identifiant de la réponse       | Primaire  |
| <b>id_question</b> | int(100)        | Identifiant de question         | Etrangère |
| <b>val_op_f</b>    | float(100)      | La valeur de l'opinion formatée |           |
| <b>Commentaire</b> | varchar(55)     | Le commentaire                  |           |
| <b>Val_op_l</b>    | float(100)      | La valeur de l'opinion libre    |           |
| <b>date</b>        | date            | Date de réponse                 |           |

Table «question»:

| Nom du Champ              | Type de données | description             | Clé       |
|---------------------------|-----------------|-------------------------|-----------|
| <b><u>id_question</u></b> | int(100)        | Identifiant de question | Primaire  |
| <b>id_service</b>         | int(100)        | Identifiant de question | Etrangère |
| <b>Désignation</b>        | varchar(55)     | La désignation          |           |
| <b>id_type</b>            | int(100)        | Identifiant de question |           |

## Chapitre III : Analyse et conception

Table «questionnaire»:

| Nom du champ     | Type de données | description             | Clé      |
|------------------|-----------------|-------------------------|----------|
| id_questionnaire | int(100)        | Identifiant de question | Primaire |
| désignation      | int(100)        | désignation             |          |
| valide           | Booléen         | valide                  |          |
| Du               | Date            | du                      |          |
| Au               | Date            | au                      |          |

Table «reponse\_comm»:

| Nom du champ     | Type de données | description                         | Clé       |
|------------------|-----------------|-------------------------------------|-----------|
| id_rep           | int(100)        | Identifiant de reponse              | Primaire  |
| id_questionnaire | int(100)        | Identifiant de <u>questionnaire</u> | Etrangère |
| id_client        | int(100)        | Identifiant de <u>client</u>        | Etrangère |
| date             | date            | date                                |           |

Table «service»:

| Nom du champ | Type de données | description                       | Clé       |
|--------------|-----------------|-----------------------------------|-----------|
| id_service   | int(100)        | Identifiant de service            | Primaire  |
| id_fservice  | int(100)        | Identifiant de famille de service | Etrangère |
| Désignation  | varchar(55)     | désignation                       |           |

## Chapitre III : Analyse et conception

Table «temporaire»:

| Nom du Champ | Type de données | description | Clé      |
|--------------|-----------------|-------------|----------|
| <b>Id</b>    | int(100)        | Identifiant | Primaire |
| <b>Mot</b>   | varchar(55)     | Le mot      |          |

Table «type\_de\_question»:

| Nom du Champ          | Type de données | description         | Clé      |
|-----------------------|-----------------|---------------------|----------|
| <b><u>id_type</u></b> | int(100)        | Identifiant de type | Primaire |
| <b>type_question</b>  | varchar(55)     | Le type de question |          |

### 6. Conclusion

Dans ce chapitre, nous avons suivi une démarche de modélisation pour développer notre application, en basant sur la méthode Uml, au départ nous avons spécifier les besoins et les divers cas d'utilisation, puis la conception de diagramme de cas d'utilisation et de séquence, et nous avons élaboré le diagramme de classe, et enfin, nous avons défini les tables de notre base de données.

Reste à définir les outils et les langages de programmation qui vont nous aider à Mettre en œuvre notre application, ce qui sera l'objectif du chapitre suivant.

# **Chapitre**

## **IV: Déploiement et implémentation**



## 1. Introduction

L'étape d'implémentation consiste en la traduction de la conception en code source. Notre activité principale est de traduire les interfaces de la conception en code exécutable.

Dans ce chapitre nous décrivons l'environnement et les outils de développement de notre application Web ainsi que des exemples d'interfaces développés illustrer ses principales fonctionnalités.

## 2. Présentation du système

### 2.1. Description de l'environnement de développement :

L'environnement de développement Windows 7 professionnel avec un serveur Web Apache et une base de données MYSQL et un langage de script python.

#### Pour quoi WAMP (Windows Apache MySQL)

Utiliser **Windows 7** professionnel en raison de sa simplicité d'utilisation, sa popularité, sa convivialité et surtout pour sa puissance ainsi pour toutes les ressources et options qu'il propose.

**Apache** est l'un des serveurs le plus utilisé sur Internet. Apache est multi plate forme et gratuit, son installation est facile, rapide et son utilisation pas très compliqué.

**Mysql** est un gestionnaire de base de données SQL. Il est rapide et multiutilisateur.

### 2.2. Les outils de développement

#### 2.2.1. L'interface PHPMyAdmin

PHPMyAdmin est une application web qui permet de gérer un serveur de bases de données MySQL. Dans un environnement multiutilisateur, cette interface écrite en PHP permet également de donner à un utilisateur un accès à ses propres bases de données. La figure suivante montre une copie d'écran de la page d'accueil de PHPMyAdmin. L'écran est divisé en deux parties. sur la gauche on peut afficher toutes les bases de données gérés par le serveur. la partie droite présente l'ensemble des opérations disponibles en fonction du contexte.

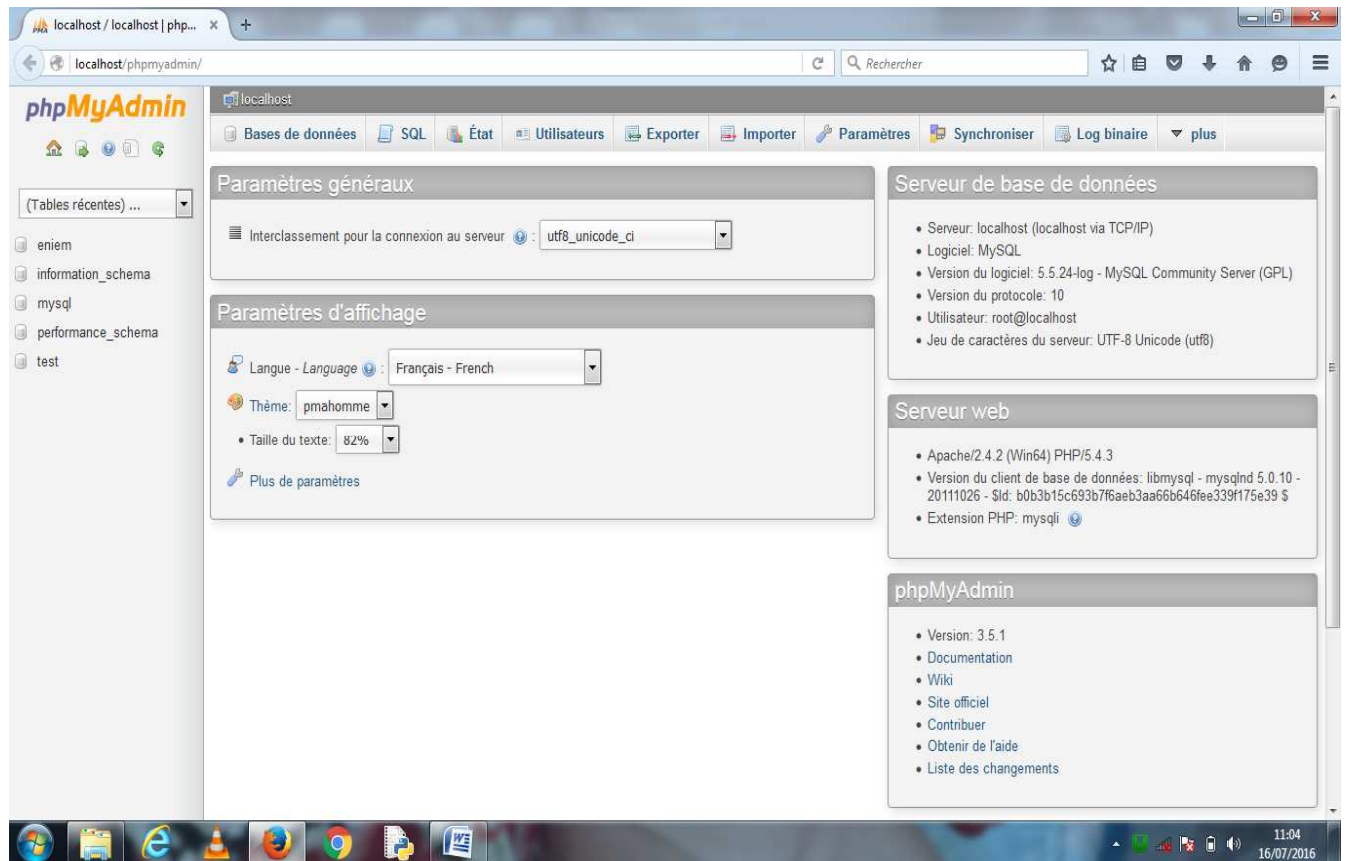


Figure 16 : Interface PHPMyAdmin

### 2.2.2. Les serveurs

#### 2.2.2.1. Serveur Apache

C'est le serveur le plus répandu sur Internet, permettant la configuration de l'environnement d'exécution de pages web. Il s'agit d'une application fonctionnant à la base sur les systèmes d'exploitation de type Unix, mais il a désormais été porté sur de nombreux systèmes, dont Microsoft Windows grâce à sa conception modulaire (morceaux de code) qui correspond à différents aspects ou fonctions du serveur.

Cette conception autorise le développeur à choisir quelles fonctionnalités seront incluses dans le serveur en sélectionnant les modules à charger soit à la compilation, soit à l'exécution. Elle lui permet aussi d'écrire son propre morceau de code qui pourra ensuite être facilement intégré dans le serveur Web Apache.

#### 2.2.2.2. Le serveur de base de données MySQL

MySQL est un système de base de données relationnelles robuste et rapide, basé sur le langage de requête SQL (Structured Query Language) qui est un langage standard pour le traitement des bases de données. Le serveur MySQL peut fonctionner en mode client/serveur. Il contrôle l'accès aux données pour assurer que plusieurs utilisateurs peuvent se servir simultanément d'une même base de données, pour y accéder rapidement et pour garantir que seuls les utilisateurs autorisés peuvent accéder aux données. Le serveur MySQL offre des

## Chapitre IV : Déploiement et implémentation

---

fonctions nombreuses et puissantes. Ses possibilités de connexions, sa rapidité et sa sécurité font de lui un serveur hautement adapté à Internet.

MySQL est disponible depuis 1996, mais son développement remonte à 1979.

### 2.2.3. Les langages de programmation

#### 2.2.3.1. HTML (Hyper Text Markup Language)

C'est un langage informatique permettant de décrire le contenu d'un document (titre, paragraphe, disposition d'images,...). Et d'y inclure des hyperliens, ces derniers peuvent mener à l'autre endroit dans le projet, mais également à sur n'importe quelles autres adresses sur le World Wide Web et même à des adresses Internet, qui ne font pas partie du WWW.

Exemple d'un fichier HTML :

```
<HTML>

<HEAD>

<TITLE>EXEMPLE</TITLE>

</HEAD>

<BODY>

BONJOUR MONSIEUR

</BODY>

</HTML>
```

Les balises <HTML> et </HTML> stipulent que ce fichier texte est formaté selon le langage

HTML et délimitent le contenu à interpréter.

Les balises <HEAD> et </HEAD> viennent du mot HEADER (entête) et délimitent l'en tête du document contenant son titre et des informations sur son contenu. Les balises <BODY> et </BODY> délimitent le corps du document contenant le texte, son formatage les objets et les liens qu'il inclut.

#### 2.2.3.2. SQL (Structured Query Language)

C'est un langage de manipulation de base de données (LMD), un langage de définition de données et un langage de contrôle de données (LCO) pour les bases de données relationnelles.

**SQL** est un LDD c'est-à-dire qu'il permet de créer, supprimer et modifier la structure des tables dans une base de données.

## Chapitre IV : Déploiement et implémentation

---

**SQL** est un LMD, c'est-à-dire qu'il permet de sélectionner, insérer ou supprimer des données dans la table d'une base de données relationnelles.

**SQL** est un LCD, c'est-à-dire qu'il permet de définir des autorisations pour la gestion des droits d'accès aux informations.

### 2.2.3.3. Java Script

C'est un langage de script incorporé dans un document HTML. C'est un langage de programmation qui permet d'apporter des améliorations au langage HTML en permettant d'exécuter des commandes du côté client, c'est-à-dire au niveau du navigateur et non du serveur Web. Il permet d'effectuer des contrôles de saisie pour valider les champs d'un formulaire, d'ouvrir ou fermer des nouvelles fenêtres ou encore de gérer des éléments graphiques. Les balises annonçant un code JavaScript sont les suivantes :

```
<SCRIPT language= JavaScript >
```

Placez ici votre code </SCRIPT>

### 2.2.3.4. Python

Python est un langage développé depuis 1989 par Guido van Rossum et de nombreux contributeurs bénévoles. Il permet -sans l'imposer- une approche modulaire et orientée objet de la programmation [40].

Le langage Python est un langage de programmation de haut niveau, facile à apprendre, orienté objet, totalement libre et terriblement efficace. Il est conçu pour produire du code de qualité, portable et facile à intégrer. Ainsi la conception d'un programme Python est très rapide et offre au développeur une bonne productivité. En tant que langage dynamique, il est très souple d'utilisation et constitue un complément idéal à des langages compilés. Contrairement à des langages spécifiques comme PHP qui se focalise sur un domaine précis, Python est universel. Il peut être utilisé dans un grand nombre de contextes. Un autre avantage de Python est la richesse de ses bibliothèques. C'est toutes ces nombreuses et importantes caractéristiques qui font la force de ce langage. Contrairement à certains langages comme Java, Python est facile à manipuler et peut s'apprendre sans formation. Il est aussi portable et plus général que Java ; il est utilisé dans une grande variété de domaines. Il possède aussi des bibliothèques beaucoup plus riches que celle de Java.

En plus de toutes ces caractéristiques nous l'avons choisi par rapport à sa force dans l'analyse d'opinion grâce à son module NLTK.

#### ❖ NLTK (Natural language toolkit) :

Le langage Toolkit Natural est une bibliothèque open source pour le langage de programmation Python écrit à l'origine par Steven Bird, Edward Loper et Ewan Klein pour une utilisation dans le développement et l'éducation. Il est livré avec un guide pratique qui présente des sujets en linguistique informatique,

## Chapitre IV : Déploiement et implémentation

---

NLTK comprend plus de 50 corpus et sources lexicales comme le Penn Treebank Corpus, Open Wordnet multilingue, Signaler un Problème Corpus et dépendances Thesaurus Lin.

Le langage Toolkit Natural (NLTK) est une plate-forme utilisée pour la construction de programmes Python qui travaille avec des données du langage humain pour l'application dans le traitement statique du langage naturelle (NLP). Il contient des bibliothèques de traitement de texte pour tokenization, l'analyse, la classification, découplant, le marquage et le raisonnement sémantique. Il comprend également des démonstrations graphiques et des ensembles de données de l'échantillon. [41]

Python a plusieurs Framework tel que Django et flask nous on s'intéresse au Framework flask.

- ❖ **Flask** : est un Framework open-source de développement web en Python. Créé initialement par Armin Ronacher. Son but principal est d'être léger, afin de garder la souplesse de la programmation Python, associé à un système de template. Le souhait de Ronacher était de réaliser un Framework web contenu dans un seul fichier Python mais pouvant maintenir des applications très demandées. [42]

### 2.2.3.5. Css

Le terme **CSS** est l'acronyme anglais de Cascading Style Sheets qui peut se traduire par "feuilles de style en cascade". Le CSS est un langage informatique utilisé sur l'internet pour mettre en forme les fichiers HTML ou XML. Ainsi, les feuilles de style, aussi appelé les fichiers CSS, comprennent du code qui permet de gérer le design d'une page en HTML.

### 2.2.4. Autres logiciel

#### 2.2.4.1. Eclipse

Eclipse IDE est un environnement de développement intégré (le terme Eclipse désigne également le projet correspondant, lancé par IBM) extensible, universel et polyvalent, permettant potentiellement de créer des projets de développement mettant en œuvre n'importe quel langage de programmation.



Figure 17 : Interface Eclipse

### 2.3. Captures d'écran

- ❖ **La page d'accueil** : c'est la première page téléchargée et visualisée par l'administrateur, le client, l'expert linguiste et le fonctionnaire eniem, c'est la page principale de notre application.



Figure 18 : Capture d'écran de l'interface « page d'accueil »

## Chapitre IV : Déploiement et implémentation

- ❖ **Page administrateur** : Pour accéder à son espace l'administrateur doit saisir son identifiant et mot de passe dans un emplacement spécial dans la page d'accueil comme suit :

The screenshot shows the ENIEM website interface. At the top left is the ENIEM logo with the tagline "ENIEM, LE MEILLEUR POUR VOUS". A navigation bar contains five links: "Accueil", "Admin", "Utilisateur", "Client", and "Expert". Below the navigation bar, a breadcrumb trail reads "Vous êtes dans : Accueil > Authentification Admin". The main content area is titled "Connexion" and contains two input fields: "Identifiant:" with a placeholder "nom" and "Mot de passe:" with a placeholder "mot de passe". A "connexion" button is positioned below the password field. At the bottom of the page, a footer provides contact information: "Pour toute information, appelez le: 026 22 54 11" and "Adresse e-mail: commercial@eniem.com.dz".

Figure 19 : Capture d'écran de l'interface « authentification administrateur »

- ❖ **Espace administrateur** : Après l'authentification l'administrateur accède à son espace.



Figure 20: Capture d'écran de l'espace « administrateur »

- ❖ **Espace client** : Après l'authentification le client accède à son espace .

The screenshot shows the ENIEM web application interface. At the top left is the ENIEM logo, which includes a blue and red flame icon and the text "ENIEM" in blue, with the tagline "ENIEM, LE MEILLEUR POUR VOUS" below it. A horizontal navigation bar contains five black buttons with white text: "Accueil", "Admin", "Utilisateur", "Client", and "Expert". Below the navigation bar, a breadcrumb trail reads "Vous êtes dans : [Accueil](#) > [client](#)". The main content area is divided into two sections. The first section contains the text "Nom Client: [nom client1](#)" and "Prenom Client: [prenom client1](#)". The second section is titled "Questionnaire:" and contains a link "Bulletin de participation 'Questionnaire Consommateur'". At the bottom of the page, a black footer bar contains the text "Pour toute information, appelez le: 026 22 54 11" and "Adresse e-mail: [commercial@eniem.com.dz](mailto:commercial@eniem.com.dz)".

Figure 21: Capture d'écran de l'espace « client ».

- ❖ **Espace expert linguiste** : après l'authentification l'expert linguiste accède à son espace



Figure 22: Capture d'écran de l'espace « expert linguiste».

### ❖ Questionnaire

Non Client: [non-client](#)  
Pseudo Client: [pseudo-client](#)

### Bulletin de participation "Questionnaire Consommateur"

Contact : contact@telephonique

Évaluez votre satisfaction de la qualité de la relation avec Bilibili ?

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

Contact : contact@telephonique

Votre opinion sur l'aspect téléphonique :

Contact : contact@telephonique

Évaluez votre satisfaction de l'aspect téléphonique ?

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

Produit : performance et fiabilité du produit

Votre opinion sur les produits Bilibili :

Produit : performance et fiabilité du produit

Évaluez votre satisfaction de la qualité des produits Bilibili ?

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

Produit : performance et fiabilité du produit

Évaluez votre satisfaction de la documentation technique ?

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

Service après-vente : délai d'intervention

Évaluez votre satisfaction du délai d'intervention ?

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

Service après-vente : qualité de l'intervention

Évaluez votre satisfaction de la qualité de la prestation technique ?

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

Service après-vente : délai d'intervention

Votre opinion sur le service après-vente :

☐ Satisfait ☐ Moyennement satisfait ☐ Non satisfait

**Valider**

Figure 23: Capture d'écran de « questionnaire ».

### ❖ Page Analyse d'opinion

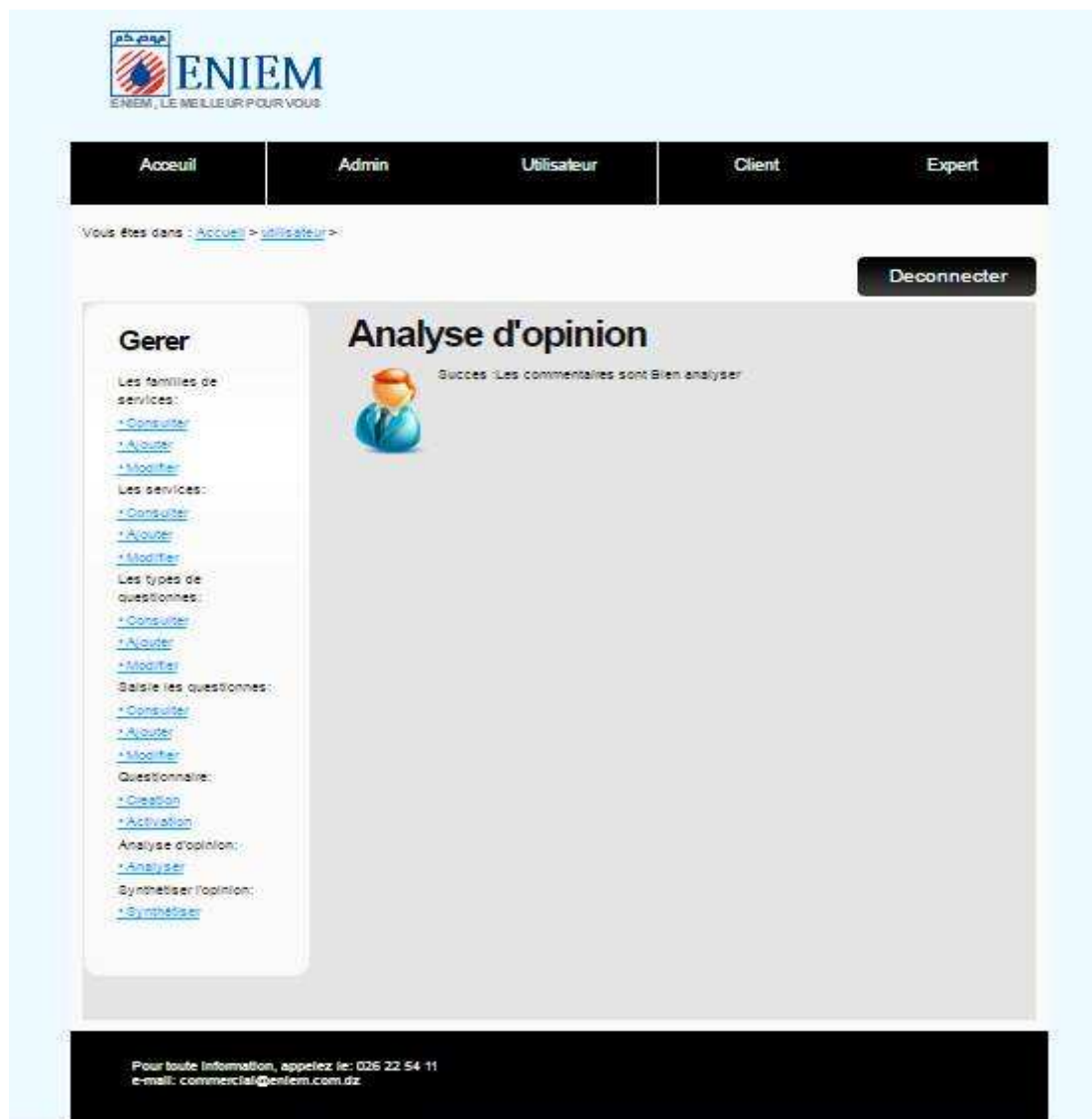


Figure 24: Capture d'écran de l'interface « analyser ».

### ❖ Page synthétiser l'opinion

The screenshot shows the ENIEM web application interface. At the top, there is a navigation bar with links: Accueil, Admin, Utilisateur, Client, and Expert. Below this, a breadcrumb trail indicates the current location: Vous êtes dans : Accueil > utilisateur >. A 'Deconnecter' button is visible in the top right corner.

The main content area is titled 'Synthétiser l'opinion'. On the left, there is a sidebar with a 'Gerer' section containing various management options for services, questionnaires, and analysis.

The central part of the page displays a table with the following data:

| Id service | Id Client | Message               |
|------------|-----------|-----------------------|
| 1          | 1         | Moyennement satisfait |
| 1          | 1         | satisfait             |
| 1          | 1         | Moyennement satisfait |
| 1          | 1         | satisfait             |
| 1          | 2         | satisfait             |
| 1          | 3         | Moyennement satisfait |
| 1          | 1         | satisfait             |
| 1          | 2         | satisfait             |
| 1          | 3         | satisfait             |
| 1          | 1         | satisfait             |
| 1          | 1         | Moyennement satisfait |

At the bottom of the page, there is a footer with contact information: Pour toute information, appelez le: 026 22 54 11, e-mail: commercial@eniem.com.dz.

Figure 25: Capture d'écran de l'interface « synthétiser».

### 3. Conclusion

Dans ce chapitre, nous avons présenté l'environnement d'implémentation et de développement de notre application, en se focalisant sur les techniques de programmation utilisées pour implémenter les pages. La description de notre application s'est faite en présentant les interfaces essentielles.

# **Conclusion générale**



### Conclusion générale

Le travail présenté dans ce mémoire s'inscrit dans le cadre de l'analyse d'opinion appliqué au domaine de l'entreprise. Il consiste à développer une application WEB qui permet de recueillir et de synthétiser les opinions des clients de l'entreprise ENIEM.

L'objectif de ce travail est d'aider l'entreprise ENIEM à améliorer la qualité de ses produits et services d'une part et d'augmenter le taux de satisfaction de ses clients d'autre part.

Le travail que nous avons réalisé nous a permis de nous initier au domaine de l'analyse d'opinion qui est très en vogue par ces temps.

Il nous a permis également d'approfondir nos connaissances théoriques et pratiques en rapport avec le web et les bases de données et les techniques de programmation web. Nous avons eu l'occasion aussi de renforcer nos connaissances techniques notamment les langages HTML, JAVA Script, Python et aussi d'utiliser plusieurs logiciels tels que eclipse et entreprise architecte ainsi que l'environnement WAMPS et le système de gestion de base de données PHPMYAdmin.

En perspective à notre travail, le traitement de la négation et de l'ambigüité dans l'analyse de l'opinion libre peuvent être envisagé.

# **Référence bibliographique**

### Référence bibliographique

- [1]. Faiza Belbachir, (2010). *Expérimentation de fonctions pour la détection d'opinions dans les blogs*. Mémoire de master. Université de Paul Sabatier, Toulouse.
- [2]. URL: <http://synthesio.com/corporate/wp-content/uploads/2010/11/SYNTHESIO-TALN.pdf>
- [3]. Damien Poirier et al, (2010). *Approches Statistique et Linguistique Pour la Classification de Textes d'Opinion Portant sur les Films*. URL: [http://hal.archives-ouvertes.fr/docs/00/46/64/12/PDF/rnti09-poirier\\_et\\_al.pdf](http://hal.archives-ouvertes.fr/docs/00/46/64/12/PDF/rnti09-poirier_et_al.pdf)
- [4]. Sigrid Maurel, Paolo Curtoni et Luca Dini, (). *L'analyse des sentiments dans les forums*. URL: <http://www2.lirmm.fr/~mroche/FODOP08/ArticlesFODOP08/Article2.pdf>
- [5]. Dominique Boullier et Audrey Lohard, (2012). « *Chapitre 5. Détecter les tonalités : opinion mining et sentiment analysis* », de *Opinion mining et Sentiment analysis*. URL : <http://books.openedition.org/oep/214>
- [6] :Liu B., (2008), « Opinion Mining & Summarization - Sentiment Analysis », Tutoriel donné à la conference WWW-2008, Avril 21, Pékin, Chine.
- [7] : Andreevskaia A., Bergler S., (2006), "Mining wordnet for fuzzy sentiment: Sentimenttag extraction from wordnet glosses", Proceedings EACL-06, the 11rd Conference of the European Chapter of the Association for Computational Linguistics.
- [8] : Liu B., (2009), "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data Centric Systems and Applications)". Springer-Verlag New York, Inc. Secaucus, NJ, (USA), ISBN: 3540378812, 1st edition.
- [9] : Turney. P. D. (2002), "Thumbs up or thumbs down ?: semantic orientation applied to unsupervised classification of reviews". In ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417-424, Morristown, NJ, (USA).
- [10] : Yu H., Hatzivassiloglou V., (2003), "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences". In Proceedings of the 2003 conference on Empirical methods in natural language processing, pp.129-136, Morristown, NJ, (USA).
- [11] : Pereira F., Tishby N., Lee L., (1993), "Distributional clustering of english words". In Proceedings of the 31st annual meeting on Association for Computational Linguistics, pp. 183-190, Morristown, NJ, (USA).

## Référence bibliographique

---

- [12] : Lin D., (1998), “Automatic retrieval and clustering of similar words”. In Proceedings of the 17th international conference on Computational linguistics, pp. 768-774, Morristown, NJ, (USA)
- [13] Hatzivassiloglou V., McKeown K. R., (1997), “Predicting the semantic orientation of adjectives”. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, pp. 174–181, Morristown, NJ, (USA).
- [14]. Kanayama H., Nasukawa T., (2006), “Fully automatic lexicon expansion for domain oriented sentiment analysis”, EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 355-363
- [15] Martin P., (1995), “Using the WordNet Concept Catalogue and a Relation Hierarchy for Knowledge Acquisition”. Proc. of Peirce'95, 4th, International Workshop on Peirce, University of California, Santa Cruz, USA, pp. 36-47.
- [16] Miller G. A., Beckwith R., Fellbaum C., Gross D., Miller K. J., (1990), “Introduction to wordnet : an on-line lexical database”. Int J Lexicography, Volume 3, issue 4, pp. 235-244.
- [17] Hu M., Liu B., (2004), “Mining and summarizing customer reviews”, Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pp. 168-177.
- [18] Liu B., (2006), “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data”, Database Management & Information Retrieval, 2ème edition, 2011, Springer Publisher.
- [19] Benamara F., Cesarano C., Picariello A., Reforgiato D., Subrahmanian V., (2007), “Sentiment analysis : Adjectives and adverbs are better than adjectives alone”, In International Conference on Weblogs and Social Media (ICWSM), Boulder, Colorado, (U.S.A), pp. 203–206, AAAI Press.
- [20] Das S., Chen M., (2001), “Yahoo! for amazon: Extracting market sentiment from stock message boards”. Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).
- [21] Hu M., Liu B., (2004), “Mining opinion features in customer reviews”, In Deborah L. Mcguinness, George Ferguson, Deborah L. Mcguinness, and George Ferguson, editors, AAAI, AAAI Press, pp. 755–760

## Référence bibliographique

---

- [22] Ounis, I. M., Macdonald C., Mishne G., Soboroff I., (2006), “Overview of the Trec-2006 Blog Track”. In Proceedings of the Fifteenth Text Retrieval Conference (TREC 2006), pp. 17-31.
- [23]. Ounis I., Macdonald C., Soboroff I., (2008), “Overview of the trec-2008 blog track”. In Proceedings of The seventeenth Text REtrieval Conference (TREC 2008). NIST.
- [24] Salton G., Lesk M. E., (1995), “The smart automatic document retrieval systems—an illustration”. Commun. ACM, Volume8, issue 6, New York (USA).
- [25] Nigam K., Hurst M., (2006) “Towards a robust metric of polarity. In Computing Attitude and Affect in Text: Theory and Applications”, The Information Retrieval Series Volume 20, 2006, pp. 265-279 Dordrecht, the Netherlands. Springer.
- [26] Crestan E., Gigandet S., Vinot R., (2007), « Approches naïves à l’analyse d’opinion ». In Actes de l’atelier de clôture du 3ème DEfi Fouille de Textes, pp. 45–56, Grenoble, (France).
- [27] Pang B., Lee L., (2004), “A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts”. In ACL’04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Article N° 271, Morristown, NJ, (USA).
- [28] Plantié M., Roche M., Dray G., Poncelet P., (2008), “Is a voting approach accurate for opinion mining ?”, In DaWaK ’08 : Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery, pp. 413-422, Berlin, Heidelberg. Springer Verlag.
- [29] Génereux M., Santini M., (2007), « Défi : Classification de textes français subjectifs ». In Actes de l’atelier de clôture du 3ème DEfi Fouille de Textes, pp. 83–93, Grenoble (France).
- [30] Trinh A., (2007), « Classification de texte et estimation probabiliste par machine à vecteurs de support ». In Actes de l’atelier de clôture du 3ème DEfi Fouille de Textes, pages 69–82, Grenoble (France).
- [31] Wilson T., Wiebe J., Hwa R., (2004), “Just how mad are you ? Finding strong and weak opinion clauses”. In In Proceedings of AAAI, pp. 761-769
- [32] Schapire R. E. Singer Y., (2000), “Boostexter: A boosting-based system for text categorization. Machine Learning”, Volume 39, Issue (2/3), pp.135–168.
- [33] Cohen W., (1996), “Learning trees and rules with set-valued features”. In Proceedings

## Référence bibliographique

---

of the 13th National Conference on Artificial Intelligence, pp. 709-716. AAAI Press.

[34] Joachims T., (1999), “Making large-scale support vector machine learning practical”. Advances in kernel methods: support vector learning, pp. 169–184. MIT Press Cambridge, MA, (USA).

[35] Nigam K., Hurst M., (2006), “Towards a robust metric of polarity”. In Computing Attitude and Affect in Text: Theory and Applications, pp. 265–279.

[36] Riloff E., Wiebe J., (2003), “Learning extraction patterns for subjective expressions”, In Proceedings of the 2003 conference on Empirical methods in natural language processing, pages 105–112, Morristown, NJ, (USA).

[37] Turney P. D., Littman M. L., (2003), “Measuring praise and criticism: Inference of semantic orientation from association”. ACM Trans. Inf. Syst., Volume 21, issue 4, pp.315-346.

[38] Dzikowski G., Wegrzyn-Wolska K, (2008), “An autonomous system designed for automatic detection and rating of film reviews”. Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on, Volume 1, pp. 847-850.

[39] Kreutzer, J., & Witte, N. (2013). Opinion Mining Using SentiWordNet. Uppsala University

[40] Gérard swinnen, (2008). Apprendre à programmer avec Python. URL: [http://www.framasoft.net/IMG/pdf/python\\_notes-2.pdf](http://www.framasoft.net/IMG/pdf/python_notes-2.pdf)

[41] [https://en.wikipedia.org/wiki/Natural\\_Language\\_Toolkit](https://en.wikipedia.org/wiki/Natural_Language_Toolkit)

[42] [https://fr.wikipedia.org/wiki/Flask\\_\(framework\)](https://fr.wikipedia.org/wiki/Flask_(framework))