

TER Master 1 (FMIN 200)  
Cahier des charges: Oracle Lexical

VEYSSIER Julien, BISQUERT Pierre  
PAIVA LIMA DA SILVA Bruno, BELMONTE Remy

-

Encadrant : Mathieu Lafourcade

6 février 2009

Université Montpellier 2

## Table des matières

|          |                                 |          |
|----------|---------------------------------|----------|
| <b>1</b> | <b>Introduction</b>             | <b>3</b> |
| 1.1      | Présentation du sujet . . . . . | 3        |
| 1.2      | Motivations . . . . .           | 3        |
| 1.3      | Besoins . . . . .               | 4        |
| 1.4      | Contraintes . . . . .           | 4        |
| <b>2</b> | <b>Solution envisagée</b>       | <b>5</b> |
| <b>3</b> | <b>Organisation du travail</b>  | <b>5</b> |
| 3.1      | Communication . . . . .         | 5        |
| 3.2      | Travail . . . . .               | 6        |
| <b>4</b> | <b>Liens et Références</b>      | <b>7</b> |
| <b>5</b> | <b>Diagramme de Gantt</b>       | <b>8</b> |

# 1 Introduction

## 1.1 Présentation du sujet

Ce TER a plusieurs objectifs qui tournent tous autour d'un but commun, obtenir des données lexicales de qualité et les utiliser dans une interaction entre un humain et un programme.

Le premier objectif est de développer un oracle lexical. Un oracle est vu dans la grèce antique comme la réponse d'une divinité à une personne venue la consulter. Par extension, le mot oracle désigne aussi le lieu ou la divinité qui donne la réponse. Dans notre cas, l'oracle est une application web qui permet de retrouver un mot que l'on aurait oublié en se servant d'indices fournis par l'utilisateur. Cet oracle se servira de données lexicales pour arriver à fournir une réponse correcte à son "fidèle".

Le deuxième objectif est de développer un jeu de devinette. Dans ce jeu, l'utilisateur devra deviner un mot choisi par l'ordinateur. Des mots ou indices lui seront proposés les uns après les autres jusqu'à ce qu'il trouve la bonne réponse. Toute réponse pourra aussi aider à créer des données lexicales.

Le troisième objectif est d'améliorer l'oracle en permettant à l'utilisateur de proposer un mot en cas d'échec de la part du programme.

Le quatrième objectif est de pouvoir créer ou obtenir des données lexicales en effectuant de la fouille de texte. Une fouille doit pouvoir fournir des données exploitables par le jeu de devinette et par l'oracle.

L'oracle et le jeu de devinette seront accessibles sur le web dans le but de faire participer un maximum de personnes à l'alimentation de nos données et d'obtenir une grande diversité dans les informations récoltées.

## 1.2 Motivations

Notre première et principale motivation est bien évidemment de se mettre en condition réelle de production de logiciel qui nous permettra, à n'en pas douter, de développer notre esprit d'analyse et notre capacité à nous plonger dans un sujet orienté sur la voie de la recherche.

De plus, nous souhaiterions réussir à produire un logiciel qui soit utilisable dans un cadre "réel", c'est à dire pouvant être réellement utile à une personne, notamment dans le fait d'aider à retrouver un mot ; permettant ainsi

de dépasser la simple conception répondant à un sujet.

Par ailleurs, nous aspirons à créer, par le biais de ce sujet, une base de données lexicales utilisable tout d'abord par notre programme mais étant également une source de données alternative pour des sujets de recherche de Monsieur Lafourcade, et pour le TALN (Traitement Automatique du Langage Naturel) en général, permettant, de la même façon que nous profitons de leurs données et méthodes, de leur faire profiter des notres.

### 1.3 Besoins

Le principal besoin en premier lieu est de pouvoir disposer d'une base de données lexicales pré-existante, nous offrant une fondation solide à la création de l'oracle proprement dit. Monsieur Lafourcade nous a, à ce titre, fourni les données de sa base de données lexicales issues du jeu "Jeux De Mots" (<http://www.jeuxdemots.org>). Cette base d'origine sera, à terme, remplacée par notre propre base, créée à partir de l'oracle lui-même.

Nous avons aussi besoin d'un serveur web et d'un système de gestion de bases de données. Nous avons choisi d'héberger nous même ces services sur le domaine aiur.fr qui nous appartient.

### 1.4 Contraintes

Les contraintes sont tout d'abord de l'ordre de l'accessibilité : il est en effet important de permettre un accès rapide et simple à l'oracle. Une application web est donc un choix très intéressant. Ceci nous permettra de plus de tenir à jour notre base de données lexicales plus aisément. Ce choix d'une application web implique donc d'être vigilant et de prendre en compte la sécurité du site, pour éviter par exemple la triche lorsque l'utilisateur s'adonnera à la partie "jeu" du site (modification des points) ou encore un accès illicite à la base de données lexicale.

Les résultats<sup>1</sup> devront être transmises à notre encadrant ou bien sous la forme de dump SQL<sup>2</sup>, ou bien via une page web avec un formulaire permettant des consultations de la base de données en tant réel.

---

<sup>1</sup>les données obtenues

<sup>2</sup>fichier de sauvegarde des données présentes dans un SGBD

## 2 Solution envisagée

En ce qui concerne l'application web, nous utiliserons Apache2 et MySQL ainsi que le langage PHP pour la programmation. Comme alternative au PHP, se présente le langage Python, langage lequel nous pourrions utiliser au cas où nous ayons besoin de développer des programmes au-delà des limites du PHP.

Le jeu que nous allons réaliser devra être attirant et facile à utiliser. Il comportera un système de points incitant le joueur à jouer correctement et fréquemment. Ces points pourront être comparés à ceux des amis du joueur. De plus, un travail sera certainement effectué quant à l'apparence graphique du jeu; elle devra être personnalisable tout en restant sobre. Evidemment, nous envisageons de partir dans une optique "communautaire" où chaque utilisateur peut avoir un compte qui maintient ses informations d'une session à l'autre. Ceci permettra à une personne inscrite d'établir une liste d'amis dont elle pourra consulter les scores. Un classement sera éventuellement affiché dans l'espace de chaque utilisateur, l'ensemble de ces idées étant mises en place de façon à rendre le jeu plaisant et addictif, ce qui permettrait, en retour, de pouvoir obtenir une bonne quantité de données lexicales à partir des parties d'utilisateurs.

N'ayant pas terminé l'analyse en profondeur de chaque point à traiter, nous nous laissons le temps et la possibilité d'ajuster ou d'introduire des nouvelles tâches et fonctionnalités au cours du projet.

## 3 Organisation du travail

### 3.1 Communication

Le début de l'organisation du travail a commencé bien avant le début de la rédaction de ce cahier des charges. Pour assurer qu'il n'y ait aucun problème de communication au sein de l'équipe de travail (étudiants + encadrant), nous nous sommes tout d'abord réunis à l'université entre étudiants pour trouver un accord commun quant aux différentes modalités de communication et échanges qui arriveront au cours du projet.

Lors de cette réunion, il a été décidé l'utilisation de mail, logiciels de messagerie instantanée, un dépôt en ligne sous un logiciel de gestion des sources, ainsi qu'une interface de gestion de projet en ligne avec plusieurs options

comme une vision de l'avancement du projet ainsi qu'un système d'échange de messages entre les participants du projet.

Entre les étudiants et l'encadrant du projet, des échanges de coordonnées pour contact ont eu lieu nous permettant d'être en synchronie avec l'encadrant et d'avoir différents de moyen de contact étant donné ses autres nombreuses occupations. Quelques rencontres (réunions) de discussion autour du projet ont eu lieu avant la rédaction de ce cahier.

### 3.2 Travail

Avant un début de développement, une phase d'analyse sera réalisée avec la participation de l'ensemble du groupe. Des questions comme comment stocker nos données, quel nom donner à nos différentes structures, et comment mettre en relation les plusieurs parties de notre projet seront donc posées dès la première semaine de travail, et seront résolues avec l'accord et surtout le travail de tous les intégrants du groupe afin qu'il y ait une compréhension collective et sans aucun problème de la base du projet par chacun.

Suite à cela, une stratégie devra être choisie pour permettre l'emmagasinement des diverses données lexicales dont nous souhaitons vérifier à tout moment et récupérer à la fin du projet pour une analyse postérieure. Après un choix de stratégie par le groupe, celle-ci sera présentée à l'encadrant, qui vérifiera et certifiera au début sa cohérence et sa qualité, et qui, dans un second temps, nous indiquera si cette stratégie est bien pertinente pour permettre une récupération de données lexicales réellement exploitables à la fin de ce projet.

Ces deux points traités, nous passerons donc à la suite du projet, où des tâches bien spécifiques seront attribués à chacun et dont le travail sera mis en commun au fur et à mesure, suivant l'avancement du projet.

Les tâches seront divisées selon les goûts et préférences de chacun, et seront au total au nombre de 4, chaque partie étant assignée à un intégrant du groupe. Ces parties englobent :

- Développement du "Devineur", l'application utilisant l'oracle pour retrouver un mot selon des indices
- Développement du jeu et du système de vérification des données lexicales récupérées par celui-ci

- Développement et mise en place du "fouilleur" de données pour ajouter des données actuelles à la base
- Ainsi qu'une partie de création et d'ammélioration d'une interface graphique de qualité pour les utilisateurs

Même si la division de ces parties entre les intégrants du groupe est pratiquement faite, nous attendrons notre première réunion après les parties communes initiales de travail prêtes et définies en tant que définitives pour choisir et valider une séparation effective du projet selon les choix, volontés et affinités de chacun avec chaque partie.

## 4 Liens et Références

Voici une liste de quelques pages web en relation directe ou indirecte avec notre projet :

- Jeux De Mots (<http://www.jeuxdemots.org>)
- Akinator, le génie du Web ([http://www.akinator.com/aki\\_fr/](http://www.akinator.com/aki_fr/))

## 5 Diagramme de Gantt

