

UFR de Lettres et Langues

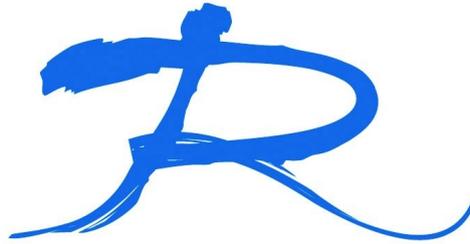
Laboratoire Ligérien de Linguistique de Tours (LLL)

3, Rue des Tanneurs BP 4103

37041 TOURS CEDEX 1

06.82.26.60.00

lll@univ-tours.fr



UNIVERSITÉ
FRANÇOIS - RABELAIS
TOURS

Cahier des charges

Réalisation d'une base de données
dictionnairique

Maître d'ouvrage
Jean-Michel Fournier

Maître d'oeuvre
Gwénael BOISSAY

Table des matières

1.	Présentation du document.....	1
1.1.	Demandeur.....	1
1.2.	Conventions et terminologies.....	1
2.	Présentation du produit.....	1
2.1.	Présentation générale du problème.....	1
2.2.	Utilisateurs.....	1
2.3.	Enoncé du besoin.....	2
2.4.	Mise en fonctionnement.....	2
3.	Description de l'environnement.....	2
3.1.	Entités.....	2
4.	Description des fonctions à satisfaire.....	3
4.1.	Fonctions principales.....	3
4.2.	Fonctions complémentaires.....	3
5.	Contraintes.....	3
5.1.	Contraintes techniques.....	3
5.2.	Contraintes d'interface.....	3
5.3.	Contraintes de temps.....	3
6.	Documentation.....	4
7.	Plan de certification.....	4
8.	Plan de développement.....	4

1. Présentation du document

1.1. Demandeur

Le demandeur est M. Jean-Michel Fournier, directeur du Laboratoire Ligérien de Linguistique de Tours (LLL). Les destinataires finaux seront les personnes faisant partie du LLL.

1.2. Conventions et terminologies

Le demandeur étant à la fois spécialisé dans le domaine de l'informatique et dans le domaine linguistique, certaines terminologies devront être définies explicitement pour plus de clarté si nécessaire.

2. Présentation du produit

2.1. Présentation générale du problème

Dans le cadre des travaux du groupe « morphophonologie de l'anglais » du Laboratoire Ligérien de Linguistique de Tours, l'équipe souhaite pouvoir vérifier des hypothèses, tester l'application automatique d'un modèle qu'elle a développé et explorer de nouvelles pistes.

Pour cela, l'équipe du LLL a décidé de lancer un projet de réalisation d'une base de données à jour sur la prononciation des unités lexicales en anglais britannique contemporain, comprenant l'ensemble des informations syntaxiques, lexicales et morphologiques, ainsi que les données de fréquence, d'usage et de variation. Cette base de données devrait comprendre environ 70000 entrées.

Dans cette optique, l'équipe a conclu des accords avec Pearson, Macquarie et Cambridge University Press pour l'obtention de dictionnaires. Ainsi, la base de données du projet sur la prononciation des unités lexicales en anglais contemporain sera fondée sur ces trois dictionnaires. (cf <http://www.jeanmichelfournier.com/French/Bienvenue.html>)

Les trois dictionnaires existants sont sous forme de fichiers textes au format XML dont le contenu diffère. Il sera donc nécessaire d'homogénéiser les données communes nécessaires à tous, et ensuite créer une base de données permettant d'exploiter ses données.

2.2. Utilisateurs

Les utilisateurs finaux seront les membres du LLL qui pourront effectuer des recherches de chaînes de caractères, obtenir des informations lexicales et statistiques.

2.3. Enoncé du besoin

- Homogénéisation des fichiers XML (EPD, LPD et Macquarie)
- Modélisation d'une base de données regroupant les informations des dictionnaires.
- Choix d'un SGBDR et création de la base de données
- Modélisation et développement de l'interface d'accès aux données (ajout, modification, recherche...)
- Moyens d'évolutivité : possibilité d'ajouter des dictionnaires, des champs supplémentaires...
- Documentation utilisateur / développeur

2.4. Mise en fonctionnement

La solution devra être disponible sur une architecture de type PC sous les systèmes d'exploitation de type Windows, MacOS X et Linux.

La solution devra être mobile et embarquée facilement sur chaque poste utilisateur.

Les modifications appliquées par les utilisateurs sur la base de données locale à leurs postes devront pouvoir être ajoutées à une base de données centrale tout en conservant la cohérence des données (synchronisation).

3. Description de l'environnement

3.1. Entités

- L'application est destinée aux membres du LLL
- Un échange de données se fera entre le poste utilisateur et le serveur central de base de données lors de la synchronisation des bases de données embarquées.

4. Description des fonctions à satisfaire

4.1. Fonctions principales

- Fonctions de recherche multicritères : fournir à l'utilisateur des critères de recherche pour qu'il n'ait pas à effectuer une requête SQL lui-même
- Tri en cascade : effectuer une nouvelle recherche sur le résultat d'une recherche précédente
- Edition de la base de données : ajout/suppression/modification
- Fonctions d'exportation des résultats de recherche (XML, CSV...)

4.2. Fonctions complémentaires

- Changement automatique de l'encodage du clavier lorsque l'on est dans un champ de type phonétique
- Recherche phonétique / Décomposition syllabique
- Génération des schémas accentuels numériques

5. Contraintes

5.1. Contraintes techniques

- Pas d'architecture distribuée client/serveur, l'application et la base de données se doit d'être embarquée
- Architecture PC, systèmes Windows/MacOS/Linux
- La solution doit être implémentée dans un langage de haut niveau du type Java ou PHP

5.2. Contraintes d'interface

- Ergonomique et complète (ex : sélection de cases à cocher lors de recherches multicritères) + exécution de requêtes SQL personnalisées
- Environnement graphique fenêtré (GDI, X11)

5.3. Contraintes de temps

La solution apportée doit être prête pour la fin du mois de mars 2010.

6. Documentation

- Documents de spécification, conception, code source...
- Manuel utilisateur

7. Plan de certification

- Réunions pilotes avec le client

8. Plan de développement

- Etude et compréhension de la problématique
- Rédaction du cahier des charges
- Analyse des fichiers XML
- Modélisation de la base de données
- Choix du SGBD et création de la base de données
- Modélisation de l'application
- Développement dans le langage choisi
- Tests de validation