

UNIVERSITE DE MONTPELLIER
RAPPORT DE PROJET

Moteur de requêtes SQL simples

Belkassim BOUZIDI
Chakib ELHOUITI
Massili KEZZOUL
Ramzi ZEROUAL
Fei YANG

Encadrante :
M^{me} Anne-Muriel CHIFOLLEAU

9 mai 2020

Introduction

Dans le cadre du cursus de la Licence 2 informatique, durant notre second semestre, il nous a été proposé un projet qui devra mettre en pratique nos connaissances et nos compétences au travers d'un cahier des charges. L'objet de cette démarche sera d'envisager la conception et le développement d'un moteur d'évaluation de requêtes SQL en mémoire vive.

Les requêtes considérées seront des requêtes simples de la forme : « SELECT ... FROM ... WHERE ... » sans imbrication.

À partir d'un ou plusieurs fichiers CSV, Comma-separated Values ¹, il est demandé de construire une représentation en mémoire des données et d'implémenter les procédures de projection, de sélection et de jointure, découlant de l'interrogation SQL.

Il s'agit principalement de reproduire les fonctionnalités de bases d'un SGBD ², tel que MySQL ou bien Oracle Database.

Notre groupe, composé de cinq personnes : Belkassim Bouzidi, Chakib Elhouiti, Massili Kezzoul, Ramzi Zeroual et Yang Feï, et encadré par M^{me} Anne-Muriel Chifolleau, a saisie l'opportunité de réaliser ce projet.

1. Comma-separated values, format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules. Voir page

2. Système de Gestion de Base de Données, voir page

Remerciements

Tout d'abord nous souhaitons adresser nos remerciements au corps professoral et administratif de la faculté des sciences de Montpellier qui déploient des efforts pour assurer à leurs étudiants une formation actualisée.

En second lieu, nous tenons à remercier notre encadrante M^{me} Anne-Muriel Chifolleau pour ses précieux conseils et son aide durant toute la période du travail.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre projet en acceptant d'examiner notre travail.

Nous remercions M^r Yahia Zeroual pour sa relecture attentive de ce rapport.

Table des matières

1	Organisation du projet	4
1.1	Méthodes d'organisation	4
1.2	Decoupage du projet	4
1.2.1	Phase de modélisation	4
1.2.2	Phase de développement	4
1.2.3	Finalisation du projet	4
1.3	Outils de collaboration	5
2	Introduction au sujet	6
2.1	Presentation du problème	6
2.1.1	Interet à resoudre le problèmes	6
2.1.2	Présenter l'approches choisie pour résoudre le problème	6
2.1.3	Cachier des charges détaillé	6
2.2	Technologies utilisées	6
3	Conception Modélisation Implementation	7
3.1	Conception et Modélisation	7
3.2	Implémentation	7
4	Analyse des résultats	8
4.1	résultats	8
4.2	Problèmes rencotrés	8
5	Bilan et conclusions	9
6	Bibliographie et annexes	10

Chapitre 1

Organisation du projet

1.1 Méthodes d'organisation

Afin de mener à bien le développement du projet, nous avons décidé de travailler un maximum de temps ensemble et de manière très régulière. Nous nous sommes réunis trois à quatre fois par semaine, en vue de faire le point sur l'avancement du projet et de définir les objectifs restant à atteindre.

Ainsi, selon l'état de progression de la conception du moteur de requêtes, nous réalismes les tâches en retard durant le week-end pour ne pas cumuler de retard et respecter l'intégralité du cahier des charges.

Toutes les semaines, nous nous sommes réunis avec notre encadrante, M^{me} Anne-Muriel Chiffolleau. Lors de ces réunions, des mises au point relatives au projet, nous furent prodiguées, cela nous a permis de bénéficier de précieux conseils.

1.2 Decoupage du projet

Nous avons découpé la réalisation du projet en trois grandes phases.

1.2.1 Phase de modélisation

Durant cette étape, nous nous sommes réunis pour définir les fonctionnalités demandées par le projet. Notamment séparer les fonctionnalités importantes de celle moins importantes. Nous avons également choisi les outils de travail collaboratifs et les principales technologies utilisées, ainsi qu'une première modélisation du projet.

1.2.2 Phase de développement

Durant cette phase, nous avons commencé à implémenter les différentes fonctionnalités que nous avons modélisées lors de l'étape précédente, tout en améliorant la modélisation au fur et à mesure de l'avancement de notre projet. Nous avons notamment réalisé des tests pour les différents modules afin de s'assurer de leur bon fonctionnement.

1.2.3 Finalisation du projet

Cette étape a consisté en la réalisation des tests finaux afin de s'assurer que le moteur de requêtes fonctionne en toute circonstance et éventuellement corriger les bogues qui peuvent apparaître.

1.3 Outils de collaboration

Afin de s'organiser, nous avons décidé d'utiliser Git au travers du serveur GitLab hébergé par le service informatique de la faculté. En effet le logiciel libre Git a facilité grandement la collaboration entre nous. Le serveur GitLab quant à lui est fourni gratuitement par le service informatique de la faculté.

En ce qui concerne la rédaction de ce rapport, nous avons utilisé \LaTeX , système de composition de documents créé par Leslie Lamport, pour faciliter la rédaction à plusieurs.

Schéma 1.1 – Logo du GitLab

Schéma 1.2 – Logo de Latex

Chapitre 2

Introduction au sujet

2.1 Présentation du problème

2.1.1 Interet à resoudre le problèmes

2.1.2 Présenter l'approches choisie pour résoudre le problème

2.1.3 Cachier des charges détaillé

2.2 Technologies utilisées

Quesqu'on a utilisé comme techno (python, Traduction et autres librairies)

Chapitre 3

Conception Modélisation Implementation

3.1 Conception et Modélisation

Lors de cette première phase, le plus important a été de comprendre et s'imprégner du code déjà mis en place et mis à notre disposition (Claimskg), comprendre les outils utilisés ainsi que la structure déjà établie.

Comme le projet est d'une envergure immense bien comprendre la structure était primordiale afin de respecter un maximum les outils utilisés, nous avons premièrement analysé les classes principales comme celle de " Claim.py ", puis comprendre comment l'exécution de l'extraction se faisait avec " init.py " . nous avons ensuite analysé les différents scripts d'extraction de sites de fact checking. De ce que nous avons compris du projet il fallait écrire une class qui représente notre site web, puis écrire des méthodes d'extraction (extraction du titre, extraction de la claim ... etc.) sur une page que nous appliquerons sur tout le site via un script python, chaque méthode représente une colonne du fichier csv final où tous les éléments extraits sont stockés, tous les résultats de chaque méthode sont regroupés dans une méthode qui les regroupent dans la classe principale " Claim.py " et c'est sur celle-ci que le script " init.py " est appliqué.

L'étape suivante était de trouver un site web respectant les critères de fact checking ainsi que les critères (une véracité, une claim écrite, un auteur, des tags ... etc) du projet initial (Claimskg), pour apporter un maximum de diversité ainsi que notre contribution personnelle spécifique nous avons décidé avec l'accord de notre encadrant d'extraire les données d'un site d'une langue différente pas encore présente dans le projet, comme tous les étudiants présents dans ce projet ont une connaissance de la langue arabe il a été logique de commencer par cette langue, le seul site officiel de fact checking en arabe est <https://fatabyyano.net/> nous avons donc analysé la structure du site web puis commencé son l'implémentation, l'étape d'après fut de trouver comment faire une " named entity recognition " sur les mêmes principes que celle déjà présenté (TagMe qui renvoi vers des liens wikipedia).

Le second site web sur lequel nous avons travaillé est <https://www.vishvasnews.com/>, 11 langues différentes sont présentent sur ce site : L'anglais, Hindi, Pendjabi, Ourdou, Bengali, Tamoul, Malayalam, Goudjerati, Télougou, Marathi et L'odia , nous avons analysé la structure du site puis extrait ses données.

3.2 Implémentation

Chapitre 4

Analyse des résultats

4.1 résultats

4.2 Problèmes rencontrés

Chapitre 5

Bilan et conclusions

Chapitre 6

Bibliographie et annexes