Projets PFE

CHRONOLOGIE DU DOCUMENT

Stagiaire/Collaborateur	Action	Date	Version
ES-SAMERY ABDELOUADOUD	Livraison du rapport	27/03/2020	1.0

INTITULE DU PROJET :

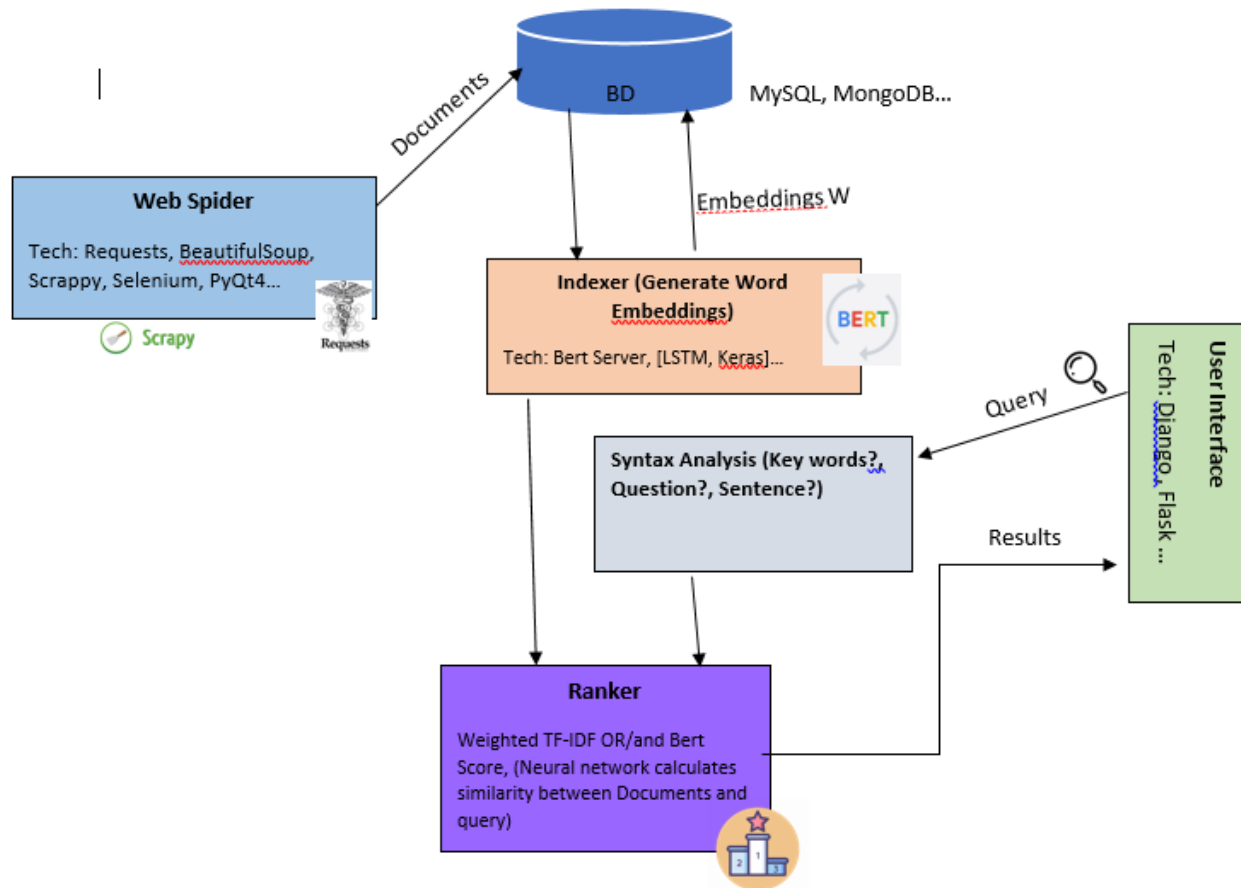| Titre du projet :				
Pôle	Encadrant	Co-Encadrant	Stagiaire	Date de début
BIG DATA	Lamalem Nizar	Haitam El Maaloumi	ES-SAMERY Abdelouadoud	02/03/2020

DESCRIPTION DU PROJET :

Search Engines are one of the most used tools in internet like Google, they are based on bots called spiders that crawl the internet, extract and index useful information, existing engines use **keyword matching** or **tf-idf** based matching to map the query to the web-documents and rank them; however they do not perform well if the query becomes too big and complex.

This Project "**Searchy**" is an improvement of classical search engines; it is an intelligent engine that uses machine learning and deep learning techniques to give more performances and reliable results. This is reachable thanks to the **Google Bert Algorithm** through understanding not only the syntax of sentence but the context of a word in the query which makes it in purpose of fully and give more related results based on it.

L'ARCHITECTURE FONCTIONNELLE/TECHNIQUE :
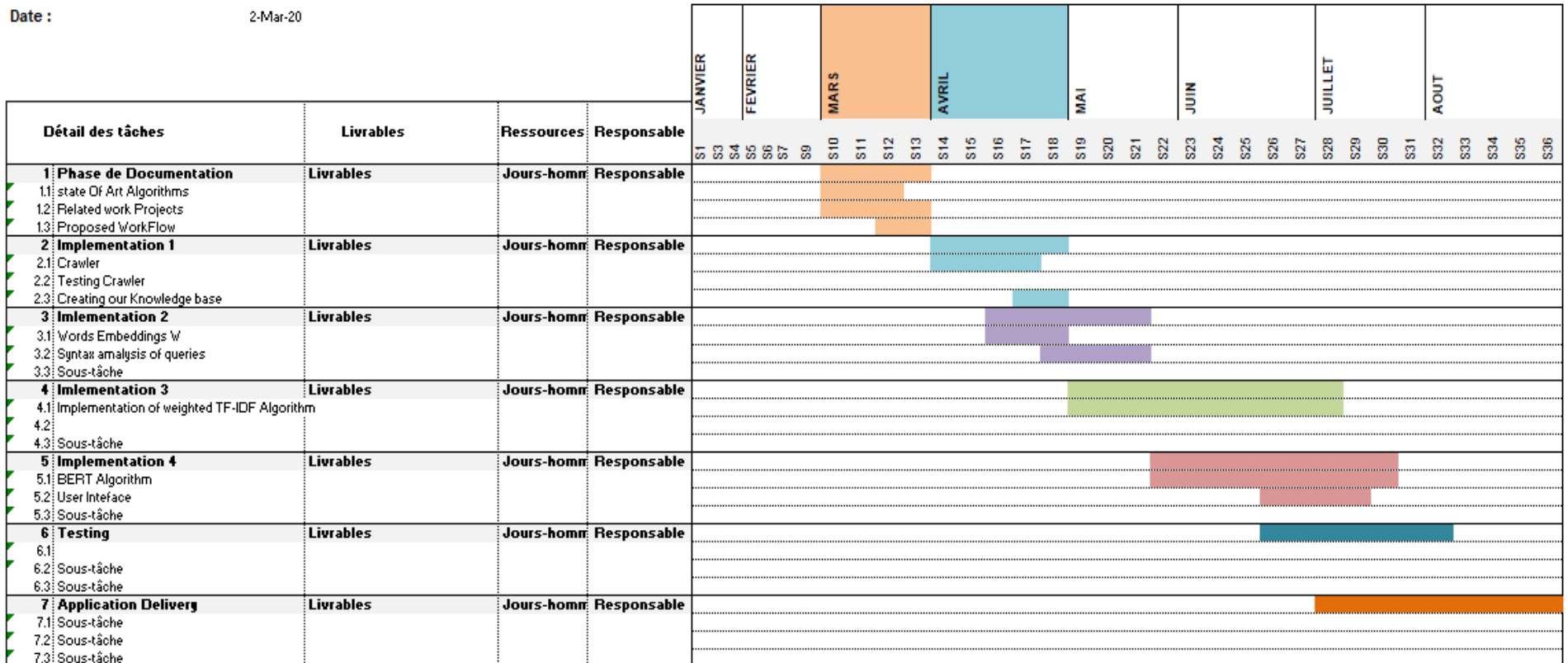

WorkFlow of the Project

DIAGRAMME DE GANTT :

Intitulé du projet : ES-SAMERY Abdelouadoud

Date : 2-Mar-20

	Détail des tâches	Livrables	Ressources	Responsable
1	**Phase de Documentation**	**Livrables**	**Jours-homm**	**Responsable**
1.1	state Of Art Algorithms			
1.2	Related work Projects			
1.3	Proposed WorkFlow			
2	**Implementation 1**	**Livrables**	**Jours-homm**	**Responsable**
2.1	Crawler			
2.2	Testing Crawler			
2.3	Creating our Knowledge base			
3	**Imlemention 2**	**Livrables**	**Jours-homm**	**Responsable**
3.1	Words Embeddings W			
3.2	Syntax analysis of queries			
3.3	Sous-tâche			
4	**Imlemention 3**	**Livrables**	**Jours-homm**	**Responsable**
4.1	Implementation of weighted TF-IDF Algorithm			
4.2				
4.3	Sous-tâche			
5	**Implementation 4**	**Livrables**	**Jours-homm**	**Responsable**
5.1	BERT Algorithm			
5.2	User Inteface			
5.3	Sous-tâche			
6	**Testing**	**Livrables**	**Jours-homm**	**Responsable**
6.1				
6.2	Sous-tâche			
6.3	Sous-tâche			
7	**Application Delivery**	**Livrables**	**Jours-homm**	**Responsable**
7.1	Sous-tâche			
7.2	Sous-tâche			
7.3	Sous-tâche			

Gantt Diagram

PLANIFICATION DES ACTIVITES :

Activité	Livrable(s)	Date prévue de livraison
Report and proposed workflow	Report	27/03/2020
Implementation 1 : Crawler	Code	
Implementation 2 : Word Embeddings, syntax Analysis	Code (+ report)	
Implementation 3 : weighted TF-IDF	Code	
Implementation 4 : Bert Algorithm, UserInterface	Code (+ report)	
Testing	Report	
Application Delivery	Code + Report	

AVANCEMENT DES TRAVAUX :

Semaine	Activités Réalisées/En cours	Avancement	Estimation en J/H
23 au 27/03/2020	Reading Resarch paper on StateOfArt Algorithms for Ranking	80%	4j
	Reading Research Project on weighted TFIDF and Bert Algorithm	90%	4j
	Reading these papers on DL Models and NLP representations	80%	4j
	Looking for more Github Projects related	95%	4j
	Django tutorial (en cours)	35%	2j
	Proposed Workflow	100%	2j

LIVRABLES REALISES :

Livrable I	Report Of The Week (23 au 27/03/2020)
Livrable II	
Livrable N	