

Projet Tableau de Bord Master 1 SID

*Rapport du projet*

Suivi des modifications

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Objet de la modification** | **Nom** | **Date** |
| 1.00 | Initialisation du document | Célia SARTORI, Samba Diallo WADE, Elhadji Fallou FALL, Serigne DIAW | 08/03/18 |
|  |  |  |  |
|  |  |  |  |

Sommaire

[1 Objet du document 4](#_Toc509927515)

[2 Problématique 4](#_Toc509927516)

[3 Documents applicables et de référence 5](#_Toc509927517)

[3.1 Documents applicables 5](#_Toc509927518)

[3.2 Documents de référence 5](#_Toc509927519)

[3.3 Terminologie 6](#_Toc509927520)

[4 Organisation de travail 7](#_Toc509927521)

[4.1 Répartition des rôles 7](#_Toc509927522)

[4.2 Organisation et répartition des tâches 7](#_Toc509927523)

[5 Processus de développement 8](#_Toc509927524)

[5.1 Recherche d’informations et Collecte des données 8](#_Toc509927525)

[5.2 Préparation des données 9](#_Toc509927526)

[5.3 Valorisation des données 9](#_Toc509927527)

[5.4 Visualisation des données 9](#_Toc509927528)

[6 Gestion de configuration 10](#_Toc509927529)

[7 Assurance et Contrôle qualité 11](#_Toc509927530)

[8 Bilan du projet 12](#_Toc509927531)

Table des illustrations

**Aucune entrée de table d'illustration n'a été trouvée.**

# Objet du document

L’objet de ce document est d’expliquer la démarche que nous avons suivie et les résultats que nous avons produits dans le cadre de la mise en place du Projet Tableau de Bord au cours de l’année de Master 1 SID de notre cursus. Il est destiné au client (enseignants) en réponse au Cahier des Charges qu’il a établi.

# Problématique

Dans le cadre de ce projet Tableau de Bord, nous devons développer un système d’aide à la décision pour le client.

Nous avons ainsi choisi pour ce projet d’étudier l’évolution du e-commerce à travers des articles d’actualités.

Pour cela, nous avons effectué des analyses et produit une visualisation de données textuelles extraites à partir d’un site Web.

# Documents applicables et de référence

## Documents applicables

Les documents applicables définissent les documents qui seront créés pendant la vie du projet.

[CCO] Charte de codage

[CG] Charte graphique

[RR] Règles à suivre pour le rapport

[CR] Cahier de recette

## Documents de référence

Les documents de référence sont les documents qui ne sont pas créés durant le projet, mais sur lesquels les documents applicables s’appuient.

[CDC] Cahier des charges W*. Bahsoun, R. Mokadem*

[GL] Cours « Génie Logiciel » *W.* *Bahsoun*

[GPRO1] Cours et TD « Gestion de Projet » *W. Bahsoun*

[BDD2] Cours et TPS « SQL Server » *R. Mokadem*

[DAWA] Cours « Data Warehouse » *G. Hubert*

## Terminologie

*Web Crawling* *:* Parcourir de manière programmée une collection de pages Web et extraire des données.

# Organisation de travail

## Répartition des rôles

* Introduire la répartition des rôles du tableau

|  |  |
| --- | --- |
| **Fonctions assurées** | **Membres de l’équipe** |
| **Chef de projet** | **Serigne DIAW** |
| **Responsable de Gestion de Configuration** | **Samba Diallo WADE** |
| **Responsable Assurance et Contrôle Qualité** | **Célia SARTORI** |
| **Développeur** | **Elhadji Fallou FALL** |

Selon la dimension du projet, une même personne peut prendre différentes responsabilités.

## Organisation et répartition des tâches

* Planification du projet (suivi du projet) Introduire choix méthodes SCRUM, GANT, SADT
* Planning prévisionnel = Diagramme de Gant
* (Modèle SADT)
* le projet un outil de gestion des sources.

Afin de suivre l’avancement de notre travail et de partager les documents de notre projet, nous avons utilisé des outils de gestion de projet en ligne : Trello et Github.

.

# Processus de développement

Processus de développement (=Démarche et réalisation du projet) - Démarche et outils utilisés

Règles de développement appliquées aux différentes phases de développement

Introduire les différentes phases du projet

Préciser à chaque fois les méthodes et outils utilisés

Démarche de développement

## Recherche d’informations et Collecte des données

Dans le cadre de ce projet, nous devons rechercher des informations sur le e-commerce et son évolution.

Afin de définir le périmètre du projet, nous avons choisis d’extraire les informations utiles à notre projet d’une unique source.

Nous avons ainsi décidé d’extraire les informations à partir du site du Journal Du Net (JDN) <https://www.journaldunet.com/> .

Nous avons ensuite cherché la rubrique concernant le e-commerce accessible via le lien : <https://www.journaldunet.com/ebusiness/commerce/list> .

Nous avons définis notre équation de recherche de la forme suivante : journaldunet + ebusiness + commerce.

Cette rubrique relate des articles d’actualités concernant le e-commerce, depuis l’année 2006. Pour chaque année, les articles sont triés par mois.

**Web Crawling**

Nous avons utilisé Scrapy, qui est un Framework d'application open source permettant d'explorer des sites Web et d'extraire des données structurées qui peuvent être utilisées pour un large éventail d'applications utiles, telles que l'exploration de données, le traitement de l'information.

Ce qui nous a permis d’extraire nos articles concernant l’e-commerce à l’ère du digital. On a récupéré au total 2759 articles.

Pour chaque article on a récupéré les éléments nécessaires à notre analyse par la suite. C’est-à-dire par exemple : le titre, l’auteur, l’introduction, le contenu de l’article, ses tags et sa date.

Après avoir constitué notre corpus cible, nous avons procédé au nettoyage des textes.

## Préparation des données

Avant d’insérer les données dans la base, il est nécessaire d’effectuer un traitement sur les données textuelles. En ce sens nous avons créé un algorithme qui effectue cette tâche.

Les principaux objectifs sont :

-Enlever les caractères spéciaux dans les contenus textuels

-Ne - Conserver que le texte

-Remplacer les articles n’ayant pas de contenu par « NAN »

Ce nettoyage a été effectué sur l’introduction des contenus textuels et sur les noms des éditeurs de chaque article avant d’être insérer dans la base donnée.

Cependant pour obtenir les mots clés un filtrage supplémentaire plus complexe a été effectué, qui intervient à la suite de notre premier algorithme de filtrage. Les principaux objectifs de cet algorithme sont les suivantes :

-Enlever les mots ayant une taille inferieur a 3 lettres.

-Prédire les entités nommées (c.-à-d. si le mot est un nom, un adjectif, nom commun etc…) pour chaque mot dans l’introduction d’un article.

-Retenir les mots ayant les entités nommées suivantes (Nom commun, Nom ou Nom Propre)

- Enfin retenir comme mot clé l’association des deux mots ayant le plus de fréquence dans l’introduction.

Ainsi, cet algorithme nous permet après un filtrage très élaborer, d’avoir nos mots clés pour chaque article prés à être insérer dans la base de données.

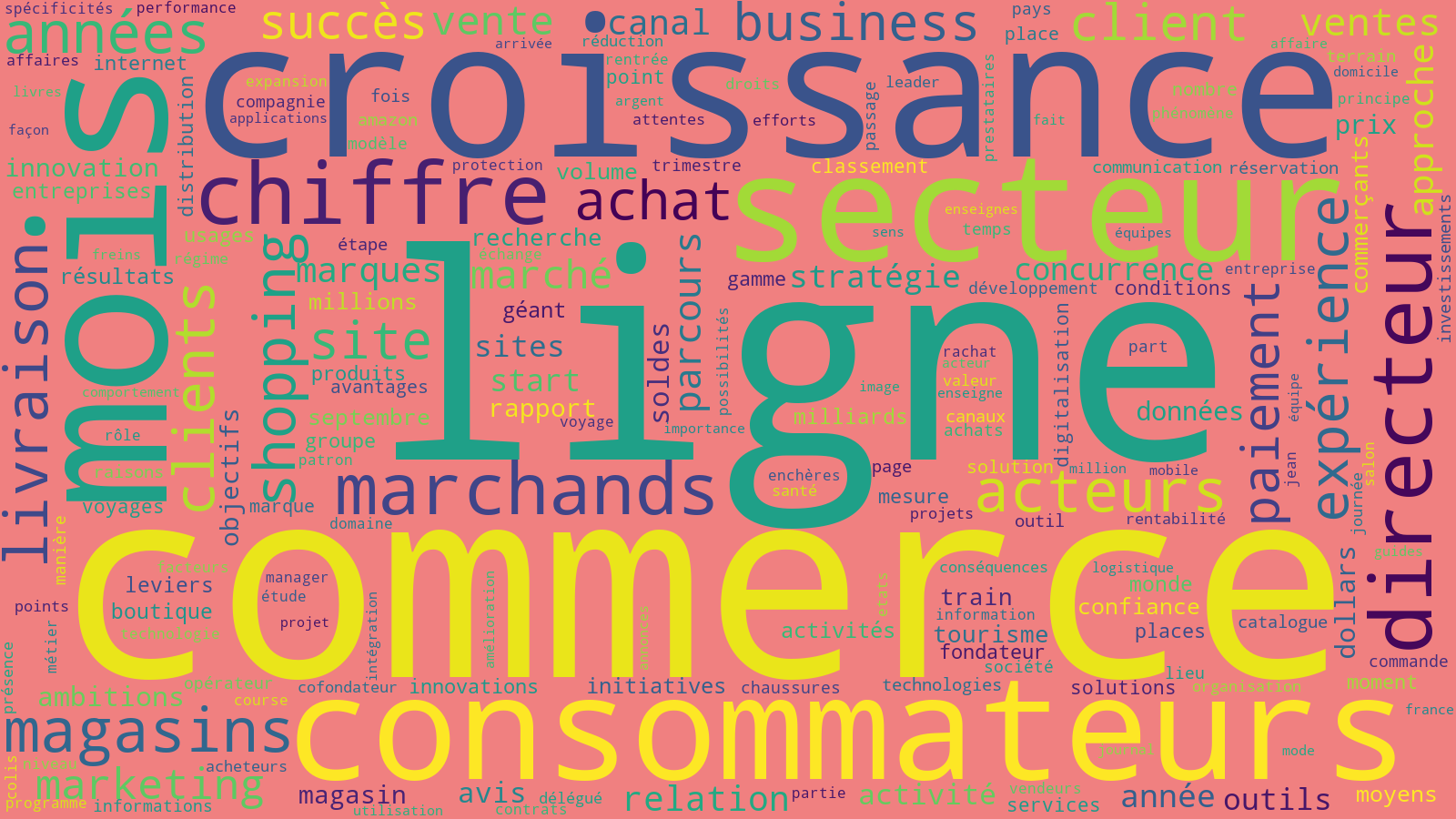
Puis nous avons structuré les informations afin de construire la base de données.

## Valorisation des données

* Construction + Administration BD 🡪 SGBD
  + MCD
* Méthodes statistiques – Choix données

## Visualisation des données

### Analyse des mots les plus fréquemment utilisés parmi nos articles



Interprétation :

Sur ce graphique on a les mots les plus utilisés dans nos articles de 2006 à 2018, on constante que parmi les mots les plus fréquents on a « Commerce, ligne, Croissance » ce qui nous permet de déduire que le E-commerce en plein expansion.

### Analyse de sentiment :

Dans cette partie nous avons essayé de comprendre le ressentis des personnes vis à vis de l’e-commerce à travers nos articles depuis 2006 à 2018.

Interprétation :

Ainsi nous avons constaté que dans nos articles la plus part ne porte pas de jugement sur ce nouveau phénomène, c’est des articles que l’on considère comme neutre. On a 1744 articles considérés comme étant neutre. Par contre on a 1015 articles qui portent un jugement sur cette nouvelle tendance. Parmi les 1015 il y n a 872 publications qui ont une opinion positif et 143 articles qui ont une opinion négatif.

Interprétation :

Ce graphique permet de voir le pourcentage d’évolution de la polarité des articles chaque année depuis 2006. On constate les années où il y a le plus d’article qui porte un jugement sur le e-commerce sont de 2013 à 2017. En particulier l’année 2014 est l’année ou il y a plus d’article subjectif. Globalement les articles positifs ont une large dominance ce qui nous permet de déduire que l’E-commerce est plein expansion. Par contre les articles négatifs ne sont pas négligés car on peut voir que ces dernières années leur proportion augmente légèrement tandis que celui des articles positifs diminue légèrement.

**Analyse des mots les plus fréquemment utilisés parmi dans les articles ayant une opinion négatif sur le E-commerce**



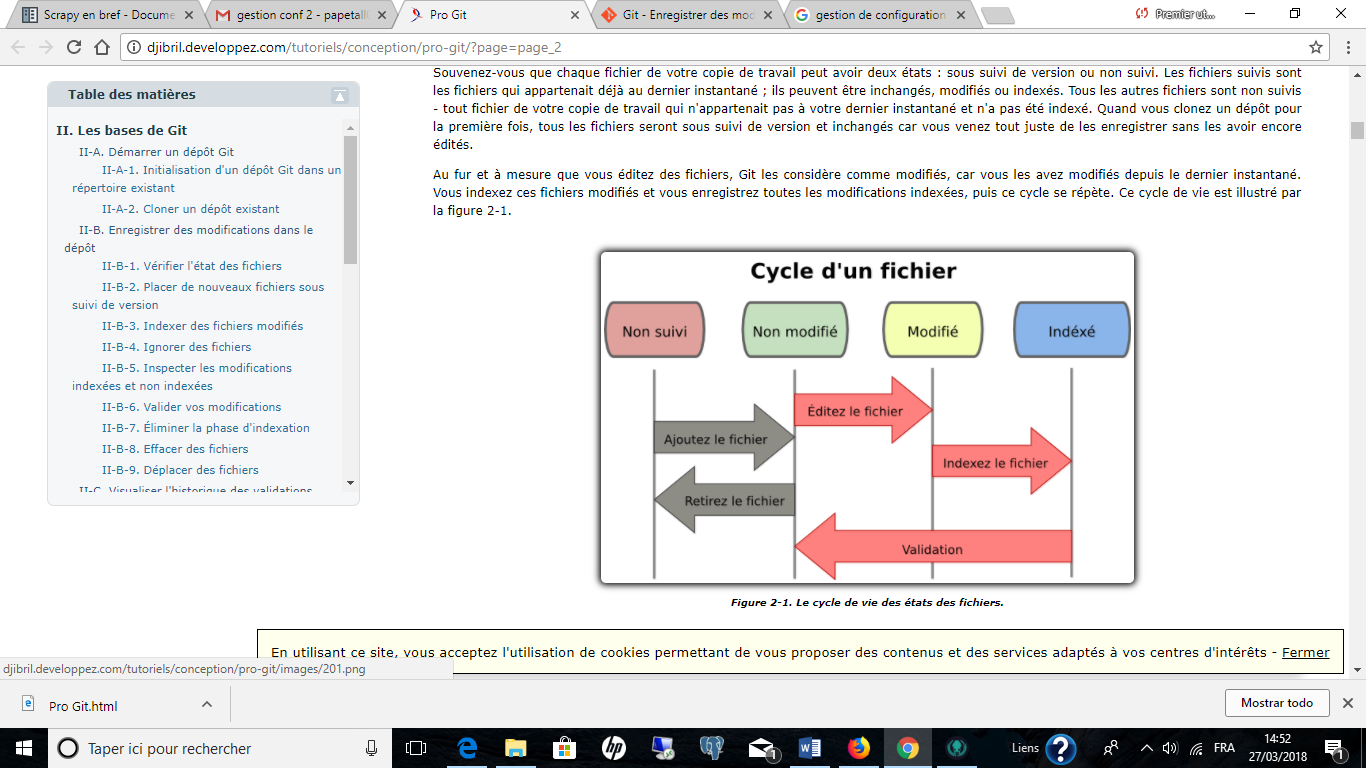
Interprétation :

Sur ce graphique on a les mots les plus utilisés dans les articles ayant une opinion négatif sur le E-commerce de 2006 à 2018, on constante que les mots les plus cités dans ces articles sont «Commerce, Ligne, France, Euros, Internet etc..»

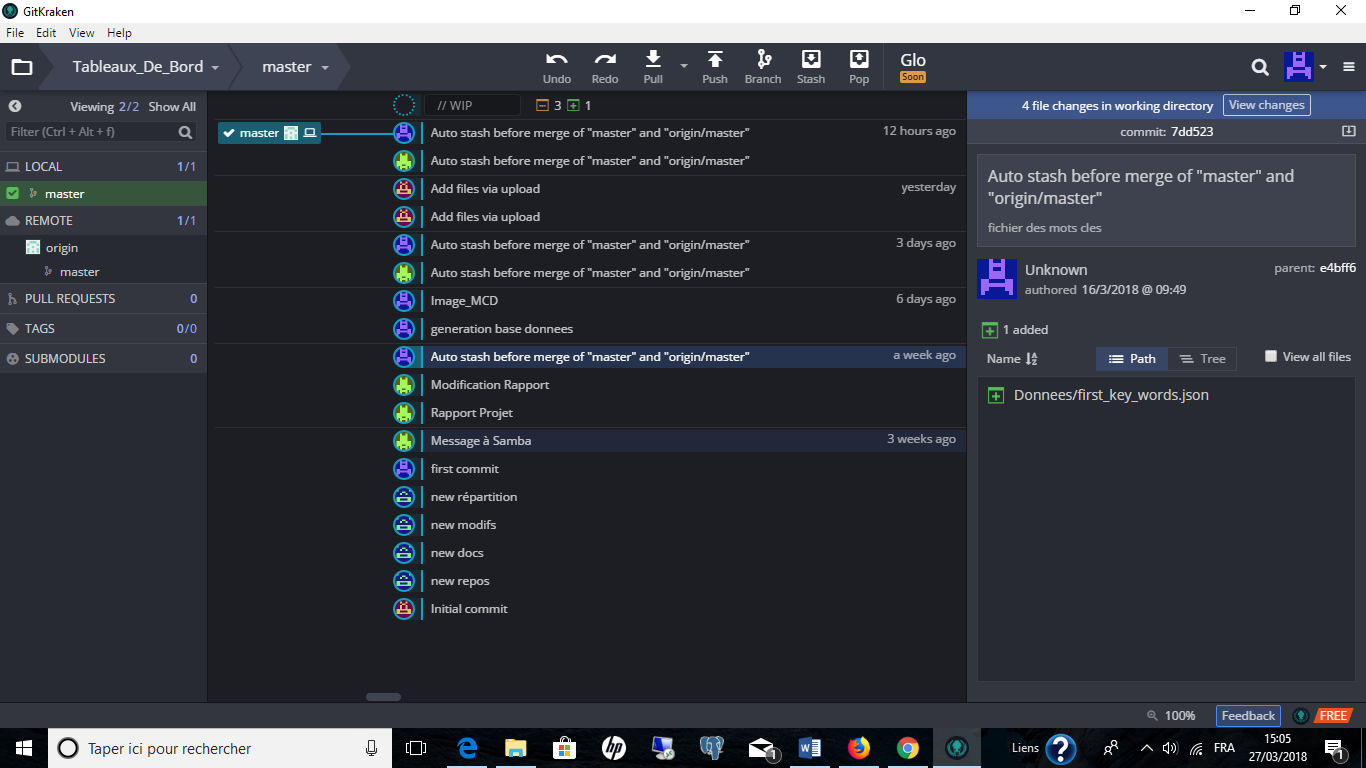
# Gestion de configuration

La gestion de configuration peut être utilisée à plusieurs fins nous l’utilisons afin de stocker et tracer les différentes versions ou révisions de toute information destinée à être utilisée par notre système (matériel, logiciel, document, donnée unitaire, etc.), en d’autre terme La gestion de la configuration gère les évolutions du produit pendant tout son cycle de vie, en termes d’adéquation entre ce qui est spécifié et ce qui est réalisé. Ainsi nous avons utilisé le logiciel Git HUB par le biais de Git Kraken pour pouvoir synchroniser notre travail et assurer la traçabilité de chaque élément produit.

Le principe de la gestion de configuration de Git hub pour un fichier est illustré par le schéma suivant :



Le schéma ci-dessous permet de voir la structure de la gestionnaire de configuration des fichiers par git kraken :



Pour chaque fichier quand on le selectionne on a la description de la modification, la version, la date et l’auteur qui apparait.

# Assurance et Contrôle qualité

* revues entre clients fournisseurs, réunions

**Démarche Qualité**

La démarche qualité consiste à concevoir et développer un service de qualité correspondant aux attentes des utilisateurs.

Selon l’ISO 9000, la qualité est définie telle que : «L’aptitude d’un ensemble de caractéristiques intrinsèques à satisfaire des exigences».

C’est-à-dire que la qualité d’un logiciel correspond à son aptitude à satisfaire les besoins explicites, mais également implicites du client.

Plus précisément, le processus de qualité consiste à identifier les exigences de qualité et les normes à respecter pour le projet et ses livrables.

Pour cela, nous avons produit de la documentation sur la façon dont le projet démontre sa conformité aux exigences et aux normes de qualité appropriées. C’est-à-dire les documents suivants : la charte de codage, la charte graphique, les règles à suivre pour le rapport ainsi que le cahier de recette.

L’intérêt principal de ce processus est qu’il fournit les directives et les orientations de management et de validation de la qualité tout au long du projet.

**Assurance et Contrôle Qualité**

L’assurance qualité est, selon la déﬁnition ISO 9000, la partie du management de la qualité visant à donner conﬁance en ce que les exigences pour la qualité soient satisfaites.

Il s’agit donc de piloter le processus de développement en mettant en place un mécanisme de prévention des défauts qui consiste à déﬁnir, au début du projet, les activités de vériﬁcation et de validation du cycle de développement .

Ce mécanisme d’assurance qualité s’appuie sur l’engagement conjoint du client et du fournisseur.

Les revues incluses dans la démarche d’assurance qualité, portant à la fois sur le processus et sur le produit, permettent de s’assurer de la conformité du produit.

Elles permettent au client de suivre, d’apprécier et d’anticiper l’avancement du projet, et à l’équipe de projet d’organiser son travail.

# Bilan du projet