



Travail préparatoire groupe 2 - Index

Ce document est une restitution du travail préparatoire effectué par le groupe n°2, en vue du projet inter-promo qui se tiendra du 8 au 19 janvier 2018. Il décrit les méthodes des gestions de projets, les outils et objectifs fixés. Nous spécifions ici les langage utilisés et le format d'échange des données.

Organisation et présentation du projet :

Nous avons organisé 2 réunions en amont du projet avec le second groupe de base de données.

Première réunion :

- Réponses aux questions liées au context du projet
- Présentation des objectifs (simple et fonctionnel)
- Présentation des outils (Git, SourceTree, méthodes agiles, API, MYSQL, JMERISE)
- Mise au point du planning

Seconde réunion :

- Réalisation du MCD en collaboration avec le second groupe de BD, les chefs de groupe et le groupe supervision
- Précisions sur les cas d'utilisation

Les chefs de groupes (filtrage, statistique et machine learning) se sont rendu disponible pour préciser les use case. Florence Canal et Loann Michon (supervision) ont participés à la mise en place du planning et du MCD et nous ont accompagné.

Nous développerons l'interface de programmation applicative REST avec le langage de programmation PHP.

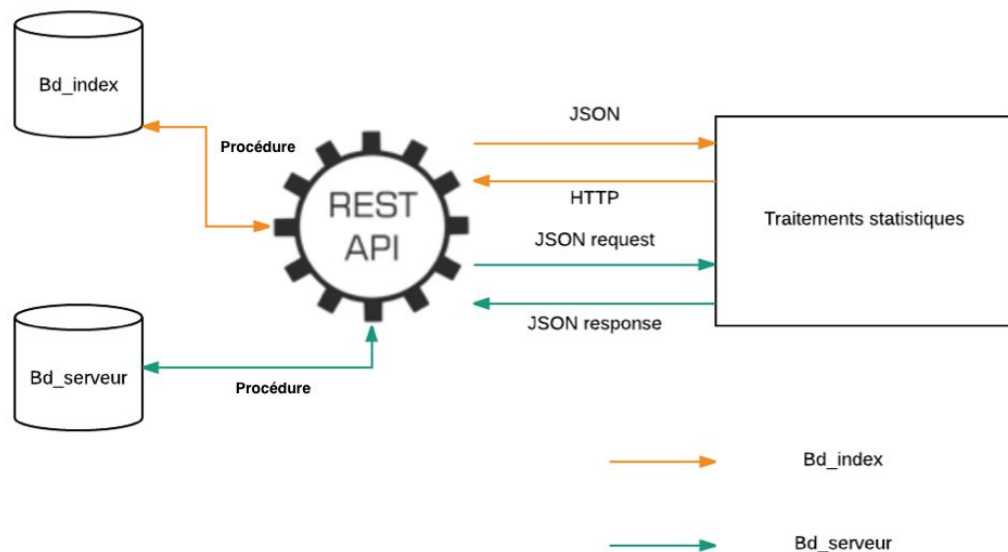
Forces :

- Les membres du groupe 2 et 3 ont eu l'occasion de travailler en mode projet avec le langage de programmation PHP
- Le langage de programmation PHP est intégrable par tous les serveurs

Architecture :

Les groupes communiqueront avec l'API en JSON. L'API REST analysera le JSON et exécutera les procédures. Nous allons créer une url de type POST par groupe. (filtrage, sémantique, machine learning, statistique).

L'API exécutera pour chaque url les procédures d'insertions dédiées.



Le graphique ci-dessus a été réalisé par le groupe statistique. Il représente un cas d'interaction entre la base de données et le groupe statistique. Le processus sera similaire pour les autres groupes.

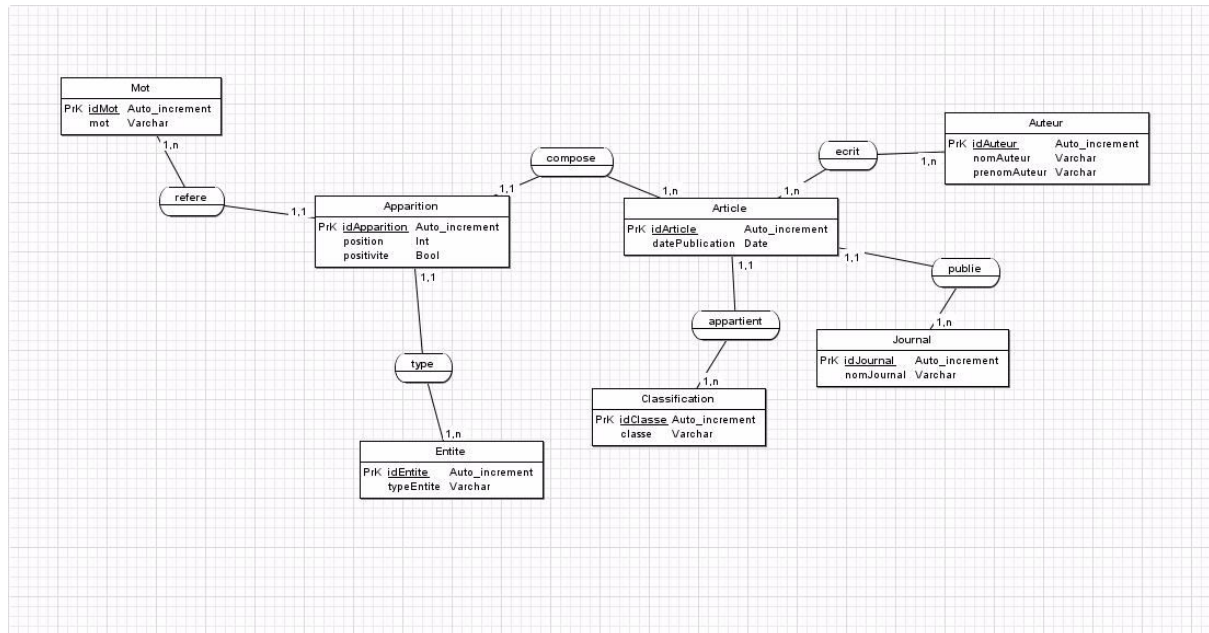
Le groupe statistique est le seul groupe où les interactions se font à double sens avec notre groupe. Nous fournirons les données des analyses statiques à partir de procédure et d'un service de type GET fourni par l'API.

Format des données pour le transfert des données du groupe index vers le groupe statistiques :

```
{
  {source : l'equipe, categorie : sport, entite_nomme : Y, mot : macron, ...}, {source : le
  figaro, categorie : politique, entite_nomme : Y, mot : macron, ...}
}
```

En interne, nous développerons des déclencheurs pour assurer la cohérence des données

Modèle conceptuel de données en 3 NF :



Il s'agit ici de la version 0.2 du modèle conceptuel de données.

Cette V0.2 permet de stocker pour le groupe filtrage :

- Liste des mots d'intérêts
- Position des mots dans le texte
- Date de publication des articles
- Journaux
- Auteurs

Cette V0.2 permet de stocker pour le groupe sémantique :

- Positivité du mot : positivite -> True si positif, False sinon
- Entité nommée d'un mot

Cette V0.2 permet de stocker pour le groupe machine learning :

- La classe de l'article (nous modifierons la cardinalité si la prédiction est multi-label)

Cette VO.2 permet de calculer la fréquence d'un mot dans le corpus, le journal ou l'article.

Nous pourrions aussi agréger les analyses du groupes sémantique à l'échelle du corpus, d'un journal ou d'un article.

Nous pourrions répondre à la question : taux de positivité du mot "Toulousain" dans le journal l'Equipe.

Gestion de projet :

Dans le cadre du projet nous utiliserons des méthodes agiles. Ce sont des outils pragmatiques dédiés au pilotage et à la réalisation de projets. Ces méthodes pérennisent le bon déroulement d'un projet.

Pour nous organiser nous utiliserons un tableau de bord. Celui-ci sera divisé en 5 parties :
New Issue - To Do - In Progress - Review / QA - Done

L'objectif étant que tous les membres de l'équipe puissent déterminer, visualiser et comprendre les problématiques future, en cours et passées. Cette méthode de travail agile permet de gagner du temps et d'optimiser le travail de groupe.

L'objectif étant pour nous de simplifier le partage des données et d'imposer une architecture unique. Nous avons commencé à mettre en place un processus qualité avec Github. Toutes les réalisations sont versionnées et datées. Nous sommes en mesure d'analyser les modifications apportés par chacun des membres du groupe à l'instant t.

| Planning prévisionnel | Semaine 1 | | | | | Semaine 2 | | | | |
|--|-----------|---|---|---|---|-----------|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Implémentation de la base de données | | | | | | | | | | |
| Mise à jour de la base de données | | | | | | | | | | |
| Création des procédures | | | | | | | | | | |
| Création des déclencheurs | | | | | | | | | | |
| Création de l'API REST en PHP | | | | | | | | | | |
| Maintenance et évolution des procédures | | | | | | | | | | |
| Maintenance et évolution des déclencheurs | | | | | | | | | | |
| Maintenance et évolution de l'API | | | | | | | | | | |
| Intégration des statistiques statique à la base de données | | | | | | | | | | |
| Intégration des logs à la base de données | | | | | | | | | | |
| Rédaction du rapport | | | | | | | | | | |
| Soutenance | | | | | | | | | | |

Nous allons évoluer par équipe de deux pour mener à bien toutes les tâches. Les membres de l'équipe pourront sélectionner les tâches au dans la tableau de bord au fur et à mesure que le projet avancera.