

*

0cm10ex plus 1ex minus .2ex1.5ex plus .2ex *

0cm4.5ex plus 1ex minus .2ex1.5ex plus .2ex *

*

*

Organigramme

Sonny Klotz - Jean-Didier Pailleux - Malek Zemni

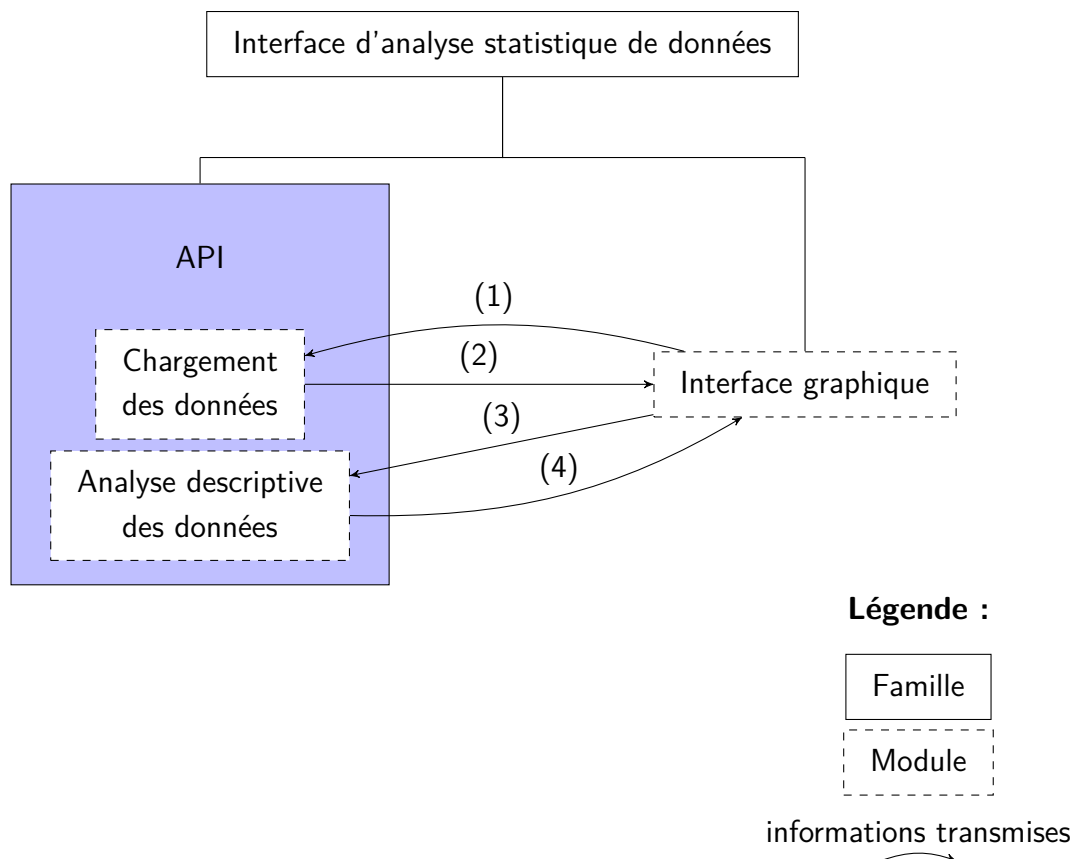
*Interface de chargement, de contrôle
et d'analyse statistique des données
pour la constitution d'un graphe de flux*

7 mars 2017

1 Présentation de l'organigramme

Le travail peut être décomposé en 3 modules :

- Une interface graphique : va faire transiter les informations entre les différentes APIs de l'application et fournir des rendus visuels des analyses
- Une API de chargement des données : va effectuer des vérifications sur les données fournies (.csv puis retourne le résultat à l'interface graphique
- Une API d'analyse descriptive des données : va permettre de calculer les informations statistiques nécessaires



Notes :

- (1) Fichier .csv formaté
- (2) Structure décrivant les données du fichier .csv après vérification : nombre de lignes et de colonnes, entrées erronées, noms des colonnes...
- (3) Colonne de données d'un type attendu
- (4) Résultats statistiques (moyenne, quantiles, écart-type, ...)

FIGURE 1 – Organigramme des différents modules du logiciel

2 Fonctionnalités des modules

2.1 API

Application Programming Interface constituent les paquets utilisables par les développeurs (intégrée), qu'on va livrer au client en plus de l'application elle-même.

Parmi les modules du logiciel, on a 2 APIs :

- Une API de chargement des données
- Une API d'analyse statistique des données

2.1.1 Chargement des données

Le module de chargement des données s'occupe de la préparation et de la vérification du fichier fourni en entrée. Ces principales fonctionnalités sont :

- Ouverture du fichier en vérifiant qu'il a la bonne extension `.csv` et qu'il est accessible en lecture.
- Donner le nombre de lignes et de colonnes
- Détection des type de colonnes : ça sera plutôt une détection de la cohérence des types, on parcourt les entrées de chaque colonne et on vérifie si leur type correspond bien au type attendu.

Dans le même parcours, on pourra aussi gérer les erreurs :

- détecter les valeurs erronées (valeurs qui ne correspondent pas au type attendu)
- détecter les cases vides
- Nommer les colonnes.

En sortie, on aura donc une structure décrivant les données du fichier `.csv` après vérification : nombre de lignes et de colonnes, entrées erronées, noms des colonnes...

2.1.2 Analyse descriptive des données

Ce module est utilisé pour fournir des informations de statistiques descriptives sur les colonnes de données qui lui seront fournies. Ses fonctionnalités se décomposent en trois pour les différents types de données à traiter :

1. qualitatif : calcul des effectifs et fréquences d'apparition
2. quantitatif discret :
 - effectifs, effectifs cumulés, fréquence et fréquence cumulée
 - moyenne
 - médiane et autres quantiles
 - variance et écart-type

- anomalies : boîte à moutaches de Tukey
- symétrie : coeff de Pearson ou coeff de Yule
- aplatissement : coeff de Fisher

3. quantitatif continu : on regroupe les valeurs en classe d'intervalles, on peut ensuite appliquer les mêmes techniques que pour les variables discrètes

2.2 Interface graphique

JD