

Organigramme

Sonny Klotz - Jean-Didier Pailleux - Malek Zemni

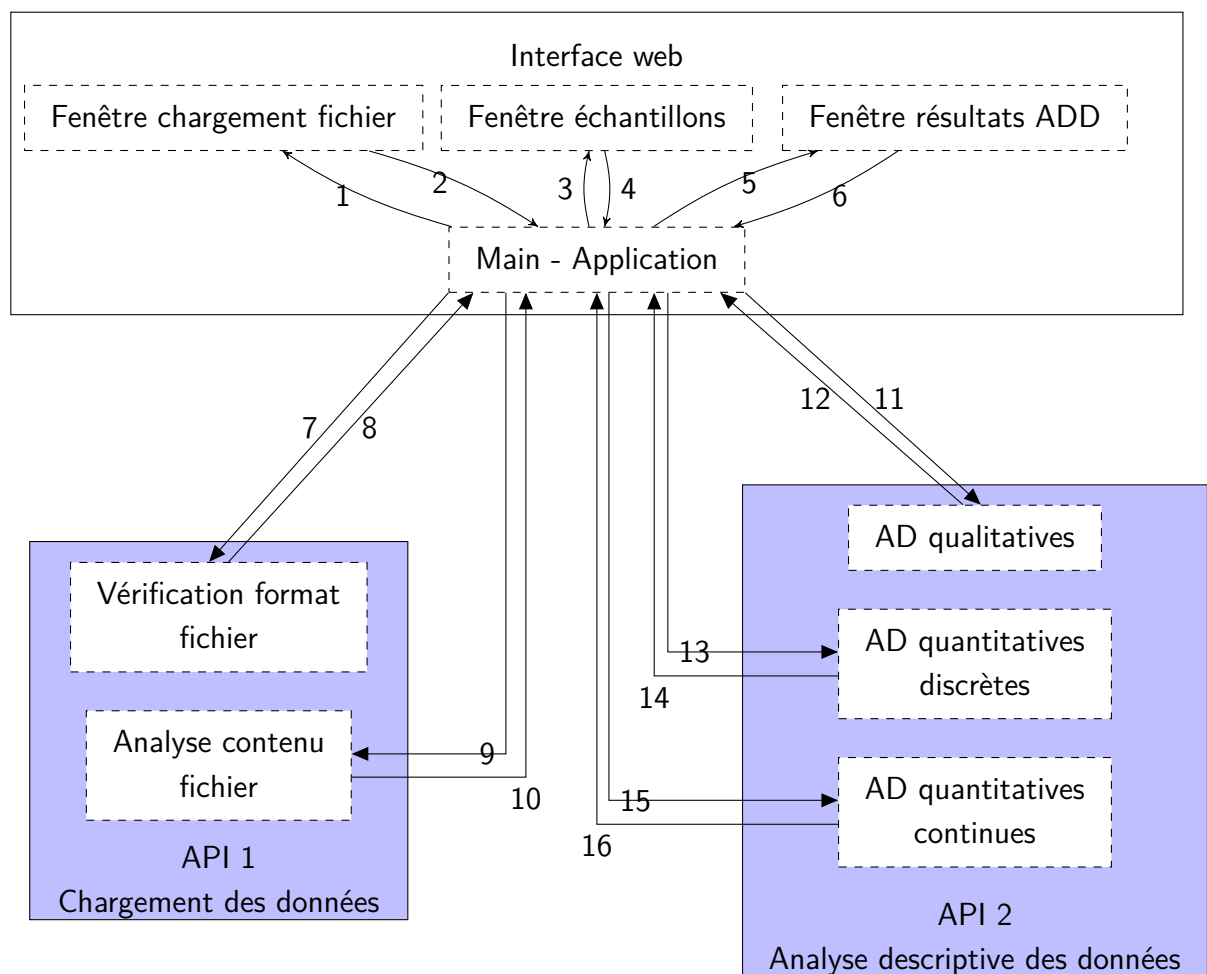
*Interface de chargement, de contrôle
et d'analyse statistique des données
pour la constitution d'un graphe de flux*

20/03/2017

1 Présentation de l'organigramme

Le projet peut être décomposé en 3 grands paquets :

- Une **interface graphique web** : module d'interface utilisateur qui va faire transiter les informations entre les différentes APIs
- Une **API 1 de chargement des données** : composée de 2 modules
- Une **API 2 d'analyse descriptive des données** : composée de 3 modules



Légende :

Famille

Module

informations transmises

FIGURE 1 – Organigramme des différents modules du logiciel

Notes :

- (1) Lancement de la fenêtre de chargement du fichier `.csv` par le Main-Application
- (2) Récupération du fichier `.csv`

- (3) Transmission des informations contenus dans le `.csv`
- (4) Demande d'accès à la Fenêtre résultats ADD
- (5) Envoi de la colonne à analyser
- (6) Demande d'exportation des résultats de l'ADD
- (7) Fichier `.csv` : lancement de la vérification de son format
- (8) Code d'erreur : fichier OK ou ERREUR
- (9) Fichier `.csv` : lancement de l'analyse de son contenu
- (10) **structure 1** : contenant les données du fichier, le nombre de lignes et le nombre de colonnes (connus à partir de la taille de la structure)
structure 2 : contenant 3 informations sur chaque colonne : le type, le rôle et les positions des données erronées ou manquantes
- (11) demande du compte-rendu AD qualitatif
- (12) succès : infos et représentations graphiques correspondantes, erreur : message d'erreur
- (13) demande du compte-rendu AD quantitatif discret
- (14) succès : infos et représentations graphiques correspondantes, erreur : message d'erreur
- (15) demande du compte-rendu AD quantitatif continu
- (16) succès : infos et représentations graphiques correspondantes, erreur : message d'erreur

2 Fonctionnalités des modules

2.1 API 1 - Chargement des données

L'API de chargement des données va être divisée en 2 modules :

2.1.1 Vérification format fichier :

Ce module va vérifier le fichier fourni en entrée en plusieurs points :

- l'ouverture du fichier a réussi
- le fichier est un `.csv` contenant du texte brut non formaté (pas de mise en forme avec des balises ou autres)
- le fichier est accessible en lecture

Critère de satisfaction : le fichier est bien ouvert, accessible en lecture et ne contient que du texte brut.

2.1.2 Analyse contenu fichier :

Ce module comprend deux fonctionnalités principales :

1. Lecture du contenu du fichier `.csv` :

On initialise une première structure (**structure 1**) pour y sauvegarder le contenu du fichier.

On lit ligne par ligne des caractères du fichier. A chaque fois qu'on détecte un caractère de séparation (une virgule, un point-virgule ou une tabulation), on stocke les caractères lus (la donnée) dans la structure.

Cette fonctionnalité fournit la **structure 1** contenant les données du fichier, le nombre de lignes et le nombre de colonnes (connus à partir de la taille de la structure).

Critère de satisfaction : l'ensemble des entrées du fichier sont renseignées dans une structure (donc exploitables)

2. Descriptions préliminaires des données de chaque colonne du fichier `.csv` :

On initialise une deuxième structure (**structure 2**) pour y stocker des informations sur chaque colonne.

On lit une par une les données de chaque colonne. On vérifie le type de la donnée en le comparant au type attendu. Si le type ne correspond pas, on signale dans la structure que la colonne contient une donnée erronée ou manquante (repérée par sa position dans la colonne). A la fin du parcours d'une colonne, on pourra lui attribuer un rôle/nom.

Cette fonctionnalité fournit donc la **structure 2** contenant 3 informations sur chaque colonne : le type, le rôle et les positions des données erronées ou manquantes.

Critère de satisfaction : pour chaque colonne, on connaît son type, son rôle et les éventuels données erronées.

Ce module va donc fournir les deux structures **structure 1** et **structure 2** décrites ci-dessus.

2.2 API 2 - Analyse descriptive des données

2.2.1 Module 1 : Calcul de la distribution des données - Analyse de données qualitatives

- effectifs, effectifs cumulés
- fréquence et fréquence cumulée
- représentations graphiques : diagramme en secteur, histogramme, ...

2.2.2 Module 2 : Analyse de données quantitatives discrètes

- 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
- représentations graphiques : distribution, cumulatif, boîte à moustaches, ...

2.2.3 Module 3 : Découpage en classes d'intervalle - Analyse de données quantitatives continues

- découpage en classe selon une précision définie
- découpage en classe selon une précision par défaut (échelle définie après un parcours des valeurs)
- représentations graphiques : distribution, boîte à moustaches, ...

2.3 Interface graphique web

Le module de l'interface graphique s'occupera de la manière dont l'application sera représentée à l'écran pour l'utilisateur, ce qui correspond au positionnement des éléments textuels, boutons et des fonctionnalités disponible dans une fenêtre. Voici une liste des principales fenêtres composant ce module avec leurs fonctionnalités :

1. Fenêtre de chargement pour récupérer un fichier `.csv` avec validation du choix pour passer à la prochaine fenêtre (En renseignant son chemin dans le système de fichier, ou de la manière d'un Drag & Drop). (C.S : Le fichier devra être chargé et devra correspondre à la demande).
2. Fenêtre de contrôle préliminaire avec visualisation d'un échantillon du `.csv`.
 - Affichera le nombre de lignes/colonnes contenu dans le `.csv`. (C.S : Affiche le bon nombre de lignes et de colonnes).
 - Affichera le titre du fichier. (C.S : Affiche le bon titre du fichier).
 - Affichera un échantillon du contenu du `.csv` (environ les 1000 premières lignes) avec un système de scroll. (C.S : L'affichage de l'échantillon doit bien se faire sur les 1000 premières valeurs et doivent bien correspondre bien correspondre au contenu du fichier).
 - Affichage des lignes erronées (numéro de la ligne + contenu + type d'erreur) (C.S : Doit bien affiché les lignes contenant les erreurs, le contenu et leur type de façon précise).
 - Mise en place d'un système de navigation sous forme d'onglet (Onglet erreurs, onglet échantillon,...). Cela permettra d'éviter que la fenêtre contienne trop d'informations. (C.S : Le système de navigation doit être lisible et pratique sans trop charger la fenêtre).
 - Affichage d'un mécanisme pour passer à l'étape de l'analyse de données (C.S : Fonctionnalité bien placé, qui répond à la demande).

3. Fenêtre correspondant à l'étape de l'analyse de données.

- L'Utilisateur devra sélectionner la colonne avec un clic, puis pourra lancer l'analyse sur celle-ci. (C.S : Sélection de la colonne doit bien se faire + lancement possible après la sélection).
- La fenêtre affichera les résultats de l'étude qualitative (Médiane, Quantile et anomalie) d'une part et de l'étude quantitative (Histogramme et Diagramme de secteur) d'autre part. (C.S : résultats conformes + diagramme bien dessiné).
- Une fonctionnalité pour lancer l'exportation des résultats sera disponible (Écriture dans un nouveau fichiers).

3 Coûts

Module	Nombre de lignes	Justification
Main-Application	15	Mise en forme du main et appel de l'application
Fenêtre choix et chargement du fichier	10 + 20	Fonctions pour : Drag& Drop + Système de fichiers
Fenêtre échantillon/contrôle	5 + 20 + 10 + 10 + 20	Communication avec le module application + Affichage de l'interface+ lecteurs des valeurs + affichage des valeurs
Fenêtre ADD	10 + 3* 30	Envoie d'informations au module application + Construction des graphe pour l'ADD pour les 3 types d'analyse
Calcul distribution de données	20 + 20*3	Application des formules pour les calculs de fréquences et d'effectifs + calcul des valeurs pour la construction de 3 graphes
Quantitatif Discret	60 + 20*2	Application des formules attaché à l'analyse quantitative discret + calcul des valeurs pour la construction de 2 graphes
Quantitatif Continu	20 + 10 + 10 + 5 + 20*2	Parcours + choix précision classe d'intervalle + écriture + communication avec les modules+ calcul des valeurs pour la construction de 2 graphes

Vérification fichier	30	Ouverture fichier + vérification si ouverture en lecture + présence de texte formaté ou non
Analyse Syntaxique	20 + 5	Recopie et vérification + initialisation de la structure
Description	25 + 10	Parcours du fichiers avec condition + Fonction pour donner nom et type de colonne
Coût Total	565	Estimation totale du coût

4 Répartition des tâches

Module	Malek	Sonny	Jean-Didier	Total
Main-Application			x	1
Fenêtre choix et chargement du fichier			x	1
Fenêtre échantillon/contrôle	x			1
Fenêtre ADD		x		1
Calcul distribution de données			x	1
Quantitative Discret		x		1
Quantitative Continu		x		1
Vérification contenu fichier	x			1
Analyse colonnes fichier	x		x	2

5 Sources

- [http ://www.math.univ-toulouse.fr/ baccini/zpedago/asde.pdf](http://www.math.univ-toulouse.fr/baccini/zpedago/asde.pdf)
- [http ://www.math.univ-toulouse.fr/ besse/Wikistat/pdf/st-l-des-uni.pdf](http://www.math.univ-toulouse.fr/besse/Wikistat/pdf/st-l-des-uni.pdf)
- [http ://iml.univ-mrs.fr/ reboul/cours2.pdf](http://iml.univ-mrs.fr/reboul/cours2.pdf)
- [http ://commons.apache.org/proper/commons-math/userguide/stat.html](http://commons.apache.org/proper/commons-math/userguide/stat.html)
- [https ://hal.archives-ouvertes.fr/halshs-00287751/document](https://hal.archives-ouvertes.fr/halshs-00287751/document)
- Wikipédia : Données aberrantes.