SAE 15 : Traiter des données

Intervenant : Sébastien DRUON / Yassine HADDAB

10 heures encadré dont 8 de TP

En tout 20 heures

Apprentissage critique= truc à mettre dans le portfolio

Ressources mobilisées :

R107/R108/R109/R110/R111/R115

Type de livrable/production :

Codes informatiques

Démonstration technique

Rapport technique (compte rendu qui décrit ce qu’on a fait et comment ça marche)

Soutenance orale présentant le travail réalisé

Objectifs visés :

Collecter traiter présenter et publier des données.

Réaliser un premier projet de développement informatique.

S’initier aux différentes étapes d’un projet informatique.

Faire un schéma/croquis de ce que l’on veut faire dans notre cas un schéma du programme (fonction, commande…)

Mettre en place un mécanisme de collecte de données

Selon la nature et quantité des données à collecter il conviendra de mettre en œuvre des mécanismes d’automatisation de la récupération des données (scripts, programmes, temporisation…)

Traitement des données :

Une fois récup il faut vérifier les consistances et leur authenticité.

Il conviendra également de les traiter (filtrage, élimination des données aberrantes…)

Extraction d’informations :

L’étape suivante consiste à extraire des infos pertinentes de ces données (caractéristiques : moyenne, écart type …)

Présentation des données :

Mise en forme de ces données de sorte à en extraire des informations pertinentes et exploitables par un humain ou une machine.

Publication :

Une fois les données mises en forme, dans certain cas il conviendra de les publier (site web, dépôt logiciel, lieu de publication…)

Principe très utilisé sur internet : le web scrapping

C’est une méthode automatique qui permet de récup de grande quantité de données à partir de sites web. La plupart des données sont non structurées (html, json…), elles sont ensuite converties en données structurées dans un tableur/base de données afin de pouvoir être utilisées dans diverse application.

data.montpellier3m.fr

Cas étude :

Outils logiciel : python

Librairies à étudier/utiliser :

Librairie « requests » qui permet d’envoyer des requetés http

Librairie « lxml » qui permet de manipuler les données d’un fichier XML

Librairie « time » qui permet de gérer le temps