



université
virtuelle
Burkina ★ Faso

RAPPORT DE FIN DE PROJET

***Scraper pour rassembler les deux
derniers articles d'actualités de sources
différentes (avec Python)***



KABORE Wendgouda Marie Innocent

KOUCHANOU Juan Prince Christel Djidenou

PLAN

INTRODUCTION

I. LA CONCEPTION

II. LES PROBLEMES

RENCONTRES

III. LECONS APPRISES

IV. L'IMPLEMENTATION

CONCLUSION

INTRODUCTION

Scraper signifie racler en français. C'est une technique qui consiste à questionner un site web sans pour autant être un utilisateur direct. Ainsi dans notre cas scraper un web consiste à écrire des lignes de code (scripts) bien ordonner afin de ressortir le maximum d'information d'un site web qui nous sont utiles. Ici dans notre cas nous avons eu a développer un ensemble de code (script) afin de questionner cinq sites web d'information du Burkina Faso afin que nos utilisateurs grâce à notre code puisse obtenir les deux dernières actualités des cinq sites et ainsi s'informer. Sans plus tarder nous allons entrer dans le vif du sujet. Dans un premier temps nous allons décrire comment utiliser notre site web tous en expliquant sa conception puis nous allons décrire les difficultés rencontrer et enfin nous allons terminer avec les notions apprissent sans omettre la conclusion.

I. LA CONCEPTION

Dans un premier temps notre application est constitué des fichier (python) permettant de faire des requêtes vers les cinq cites. Il s'agit de :

'fasozine.py', 'Lefaso.py', 'List_URL_BURKINA24.py', 'List_URL_BURKINADEMain.py', 'et 'wakasera.py' qui sont

charger de faire des requêtes ciblées vers les cinq sites afin de nous ramener les deux dernières actualités de chaque site notamment le titre, le nom de l'auteur, la description, la date de publication.

Pour cela, nous avons inspecter minutieusement chacun de ses sites pour pourvoir trouver les parties de code html que nous pouvons utiliser pour soutirer les informations voulus.

Parlons maintenant de notre base de données : Notre base de données est constituée de deux tables :la table 'actualiter ' qui contiennent les propriétés suivantes : 'identifiant_de_actualite', 'titre_actualiter', 'date_publication', 'date_entrer_base_de_donne', 'identifiant_site'

Enfin, nous avons conçu l'interface graphique avec tkinter de sorte à ce que chaque action effectuer par l'utilisateur sur cette fenetre soit gérer de sorte à interagir avec les scripts ou fonctions préétablis.

II. LES PROBLEMES

RENCONTRES

L'une des difficultés rencontrer régis au niveau des requêtes sur les sites web. En effet pour avoir accès aux différents contenus du site notamment les

informations qui nous intéressent nous sommes passées par des systèmes de lien afin d'avoir accès à nos différents sites. Notons aussi que chaque site qu'on a eu à scraper avait un squelette particulier et différents des autres du coup les mêmes scripts ne pouvaient pas être utilisés pour scraper les autres sites, ce qui rendait complexe et redondant notre travail. Aussi au niveau de la conservation des données dans la base de données nous avons dû passer par des requêtes préparées pour directement recevoir les informations des requêtes et les insérer directement dans la base de données.

L'autre difficulté importante rencontrée était dans l'affichage de l'ensemble des requêtes déjà entrées par l'utilisateur ainsi en fonction de l'url à afficher sous forme de listbox nous avons eu à ajouter deux boutons qui permettent d'agir sur la donnée en question.

L'autre difficulté se situe au niveau de l'affichage des résultats sur le LabelFrame ainsi après l'affichage d'une donnée la prochaine requête exécutée par l'utilisateur nettoyait d'abord l'ancienne information avant d'afficher les informations lui concernant alors nous avons dû corriger cela afin de conserver les anciennes informations.

III. LECONS APPRISES

Au cours du développement de ce projet, nous avons eu à apprendre de nombreuses leçons à différents niveaux. Nous avons appris comment faire des requêtes python avec la bibliothèque `requetés` et comment les récupérer puis l'analyser avec `BeautifulSoup4` pour en tirer les informations désirées. Aussi, on a appris à insérer ses résultats dans une base de données SQLite avec python et comment les manipuler à notre guise. Enfin, on a appris à faire une interface graphique conviviale avec `tkinter` pour pouvoir permettre aux utilisateurs d'interagir avec la base de données et les scripts.

IV. L'IMPLEMENTATION

Pour pouvoir implémenter notre logiciel python faut juste s'assurer que tous les fichiers suivants sont dans le même répertoire : `'fasozine.py'`, `'Lefaso.py'`, `'List_URL_BURKINA24.py'`, `'List_URL_BURKINADEMain.py'`, `'wakasera.py'`, `'ScrapperBD.py'`, `'info.py'` et la base de données `'gestion_scrapper.db'`

CONCLUSION

Ce projet de scraping de site web nous a permis d'aborder la conception d'un logiciel avec python, comment on la structure, les étapes par les lesquelles faut passer pour déboucher sur un rendu fonctionnel. On y a rencontré beaucoup de difficulté mais on n'a réussi à presque tous les surmonter. Malgré tout, nous savons que notre programme possède des limites. Ce projet nous a permis d'apprendre énormément de nouvelles notions sur python, de nouveaux modules comme requests, tkinter, sqllite, etc.