**Candidate crawler**

**Guide utilisateur**

Table des matières

[But du programme : 3](#_Toc398299024)

[Sites de recherche d’emplois utilisés : 3](#_Toc398299025)

[Pré-requis : 3](#_Toc398299026)

[Installation de l’environnement : 4](#_Toc398299027)

[Python : 4](#_Toc398299028)

[PyQT4 : 4](#_Toc398299029)

[Beautifulsoup4 : 4](#_Toc398299030)

[Schedule : 4](#_Toc398299031)

[Utilisation du programme: 5](#_Toc398299032)

[Interface de recherche ponctuelle : 5](#_Toc398299033)

[Interface de recherche planifiée : 7](#_Toc398299034)

[Interface d’administration : 9](#_Toc398299035)

[Annexe : 10](#_Toc398299036)

[Description des fichiers du programme : 10](#_Toc398299037)

# But du programme :

Ce programme, à partir d’une liste de critères, parcours différents sites web de recherche d’emplois, et extrait les annonces répondant aux critères.

L’utilisateur peut choisir de faire une recherche ponctuelle, ou de créer un profil de recherche qui sera lancé automatiquement à intervalles réguliers.

Le programme utilise une base CSV pour gérer les adresses trouvées.

# Sites de recherche d’emplois utilisés :

Aéro contact : <http://www.aerocontact.com>

Aéro emploi formation : <http://www.aeroemploiformation.com>

APEC : <http://www.apec.fr>

CAO emplois <http://www.cao-emplois.com>

Indeed : <http://www.indeed.fr>

Monster: <http://www.monster.fr>

Pôle emploi: <http://www.pole-emploi.fr>

Régions job : <http://www.regionsjob.com>

# Pré-requis :

Ce programme utilise le langage de programmation Python, qui est libre et gratuit, ainsi que certains modules complémentaires. Voici la liste des programmes à installer :

* **Python** (version 3.4 minimum)
* **PyQT4** (version 4.11)
* Le module Python **Beautifulsoup4**
* Le module Python **Schedule**

# Installation de l’environnement :

## Python :

Python est un langage de programmation libre et gratuit.

Commencer par télécharger le programme (<https://www.python.org/downloads>) puis lancer l’installation.

Afin que l’installation des modules et compléments se fassent ensuite sans soucis, il est préférable de laisser le dossier d’installation par défaut (C:\Python34), ainsi que de laisser toutes les options par défaut. Il n’y a pas de redémarrage nécessaire.

## PyQT4 :

PyQT4 est un module complémentaire de Python permettant d’utiliser la bibliothèque graphique QT, elle aussi libre et gratuite

Commencer par télécharger le programme (<http://www.riverbankcomputing.co.uk/software/pyqt/download>) puis lancer l’installation. Il est préférable de laisser toutes les options par défaut. Il n’y a pas de redémarrage nécessaire.

## Beautifulsoup4 :

Beautifulsoup4  est un module complémentaire de Python permettant de parcourir, interpréter et extraire des informations des pages web.

Il s’installe en lançant la commande suivante dans un terminal (cmd par exemple) :

*C:\Python34\Scripts\pip install beautifulsoup4*

## Schedule :

Schedule  est un module complémentaire de Python permettant de lancer des tâches à intervalles régulier.

Il s’installe en lançant la commande suivante dans un terminal (cmd par exemple) :

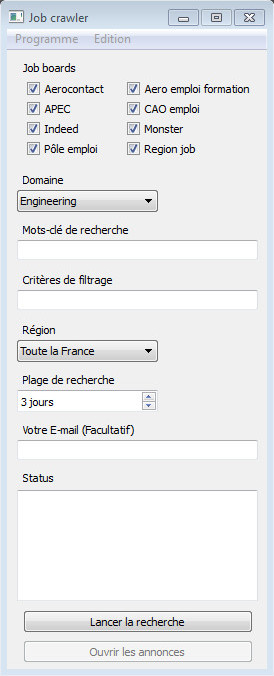
*C:\Python34\Scripts\pip install schedule*

# Utilisation du programme:

Ce programme utilise 3 interfaces principales: l’interface de recherche ponctuelle, l’interface de recherche planifiée, et l’interface d’administration. Nous allons décrire ces 3 interfaces par la suite.

## Interface de recherche ponctuelle :

Cette interface permet de faire une recherche ponctuelle sur un ensemble choisi de sites de recherche d’emploi, et une fois la recherche terminée, soit d’ouvrir les liens dans le navigateur web par défaut, ou de recevoir un mail compilant les dits-liens.



Barre de menus

Sélection des sites de

recherche d’emploi

Domaine de recherche

Mots-clé de recherche

Critères de filtrage

Région de recherche

Plage de recherche

E-mail

Fenêtre de status

Bouton de lancement

Bouton d’ouverture des annonces

**Barre de menus :** elle contient 2 menus principaux, *Programme* et *Edition*, contenant des sous-menus.

* *Programme* contient les sous-menus :
  + *A propos* : affiche une fenêtre indiquant le but du programme
  + *Quitter*: quitte le programme
* *Edition* contient le sous-menu :
  + *Initialiser* : Nettoie la fenêtre de status

**Sélection des sites de recherche d’emploi :** L’utilisateur peut cocher les sites de recherche d’emploi qu’il souhaite utiliser

**Domaine de recherche :** certains sites de recherche d’emplois utilisent non pas des mots-clé, mais des domaines de compétences. Ce champ permet de sélectionner le domaine recherché

**Mots-clé de recherche :** définit les mots-clé utilisés pour lancer les recherches sur les sites de recherche d’emploi. Les mots-clé doivent être séparés par des virgules, sans espaces. Les accents n’étant pas pris en compte, si votre mot en contient, il est conseillé de l’entrer avec et sans.

**Critères de filtrage :** définit les mots-clé utilisés pour parcourir les annonces trouvés lors de l’étape de recherche. Si l’un des mots est reconnu, l’annonce et retenue. Les mots-clé doivent être séparés par des virgules, sans espaces. Les accents n’étant pas pris en compte, si votre mot en contient, il est conseillé de l’entrer avec et sans.

**Région de recherche :** définit la zone géographique de recherche. Toutes les régions Françaises sont proposées, ainsi qu’une option pour choisir toute la France

**Plage de recherche :** définit la plage pour laquelle l’annonce est sélectionnée. Si l’annonce a été publiée hors de la plage, elle ne sera pas retenue

**E-mail :** l’utilisateur, s’il souhaite recevoir les annonces trouvées par e-mail, doit remplir ce champ, soit avec une seule adresse, soit plusieurs séparées par des virgules. Ce champ est facultatif, aucun e-mail ne sera envoyé si ce champ est vide.

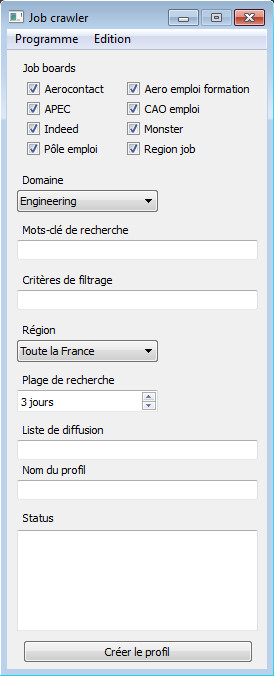
**Fenêtre de status :** l’interface de sortie du programme, annonce quand le programme se lance et se termine, ainsi que le nombre d’annonces trouvées.

**Bouton de lancement :** lance le programme, à utiliser quand tous les champs ont été renseignés.

**Bouton d’ouverture des annonces :** ouvre les annonces trouvées lors de la recherche dans le navigateur web par défaut. Ce bouton est grisé tant que le programme n’a pas été lancé et la recherche terminée.

## Interface de recherche planifiée :

Cette interface permet de créer un profil de recherche, qui sera ensuite régulièrement exécuté par un ordonnanceur.



Barre de menus

Sélection des sites de

recherche d’emploi

Domaine de recherche

Mots-clé de recherche

Critères de filtrage

Région de recherche

Plage de recherche

Liste de diffusion

Nom du profil

Fenêtre de status

Bouton de lancement

**Barre de menus :** elle contient 2 menus principaux, *Programme* et *Edition*, contenant des sous-menus.

* *Programme* contient les sous-menus :
  + *A propos* : affiche une fenêtre indiquant le but du programme
  + *Quitter*: quitte le programme
* *Edition* contient le sous-menu :
  + *Initialiser* : Nettoie la fenêtre de status

**Sélection des sites de recherche d’emploi :** L’utilisateur peut cocher les sites de recherche d’emploi qu’il souhaite utiliser

**Domaine de recherche :** certains sites de recherche d’emplois utilisent non pas des mots-clé, mais des domaines de compétences. Ce champ permet de sélectionner le domaine recherché

**Mots-clé de recherche :** définit les mots-clé utilisés pour lancer les recherches sur les sites de recherche d’emploi. Les mots-clé doivent être séparés par des virgules, sans espaces. Les accents n’étant pas pris en compte, si votre mot en contient, il est conseillé de l’entrer avec et sans.

**Critères de filtrage :** définit les mots-clé utilisés pour parcourir les annonces trouvés lors de l’étape de recherche. Si l’un des mots est reconnu, l’annonce est retenue. Les mots-clé doivent être séparés par des virgules, sans espaces. Les accents n’étant pas pris en compte, si votre mot en contient, il est conseillé de l’entrer avec et sans.

**Région de recherche :** définit la zone géographique de recherche. Toutes les régions Françaises sont proposées, ainsi qu’une option pour choisir toute la France

**Plage de recherche :** définit la plage pour laquelle l’annonce est sélectionnée. Si l’annonce a été publiée hors de la plage, elle ne sera pas retenue

**Liste de diffusion :** l’utilisateur doit remplir ce champ, soit avec une seule adresse, soit plusieurs séparées par des virgules. Comme ce programme n’a pas de sortie visuelle, ce champ est obligatoire.

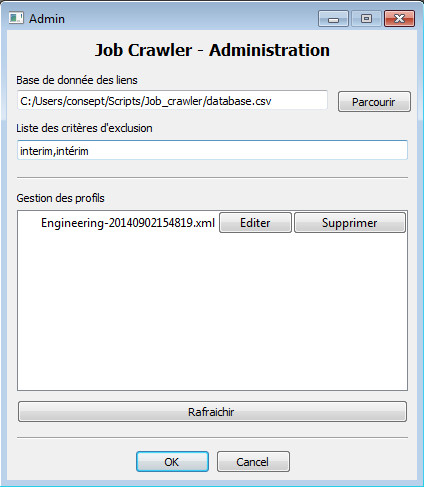
**Nom du profil :** définit le nom qui sera donné au profil de recherche. Un timestamp y sera ajouté afin de rendre le nom de profil unique, l’utilisateur peut donc choisir son nom sans crainte de doublons, apr contre sans espaces et sans accents.

**Fenêtre de status :** l’interface de sortie du programme, annonce quand le programme se lance et se termine, ainsi que le nom du profil créé.

**Bouton de lancement :** crée le profil de recherche, à utiliser quand tous les champs ont été renseignés.

## Interface d’administration :

Cette interface permet les réglages principaux de l’application, ainsi que la gestion des profils de recherche. En raison de son orientation, elle ne devrait pas être utilisée par les utilisateurs classiques du programme.



Base de donnée des liens

Liste des critères d’exclusion

Gestion des profils

Bouton de rafraichissement

Boutons de validation/annulation

**Base de donnée des liens :** Le programme utilise une base de donnée au format CSV pour garder une trace des annonces déjà affichée, afin de ne pas proposer continuellement les mêmes. Ce champ permet de donner son emplacement. Un bouton *Parcourir* permet de faire la sélection de façon graphique

**Liste des critères d’exclusion :** définit les mots-clé utilisés pour parcourir les annonces trouvés lors de l’étape de recherche. Si l’un des mots est reconnu, l’annonce est rejetée. Les mots-clé doivent être séparés par des virgules, sans espaces. Les accents n’étant pas pris en compte, si votre mot en contient, il est conseillé de l’entrer avec et sans.

**Gestion des profils :** permet de voir les profils créés, ainsi que de les éditer ou de les supprimer. La fenêtre d’édition est similaire à celle de recherche planifiée, seul le nom du profil ne peut être modifié. Quand à la suppression, un pop-up permet de confirmer son choix.

**Bouton de rafraichissement :** actualise la zone de gestion des profils

**Boutons de validation/annulation :** un click sur *Ok* permet de sauver les champs choisis et de quitter la fenêtre, le bouton *Cancel* quitte la fenêtre sans sauver.

# Annexe :

## Description des fichiers du programme :

Cette partie liste les fichiers et dossiers du programme, et explique sommairement leurs fonctions.

**\_\_pycache\_\_:** dossier où Python crée des fichiers précompilés des modules, pour une exécution plus rapide, généré automatiquement.

**predefined\_searches:** dossier de stockage des profils de recherche.

**database.csv:** base de donnée des liens au format CSV, contient les liens trouvés et la date d’ajout à la base. Toutes les annonces datant de plus d’un mois sont automatiquement supprimées, afin de ne pas trop alourdir la base.

**database.backup:** sauvegarde automatique de la base de donnée des liens

**Job\_crawler\_Config.xml:** configuration du programme, au format XML, géré par l’interface d’administration.

**Job\_crawler\_profile\_template.xml:** template pour la création des profils de recherche au format XML.

**Job\_crawler\_statics.xml:** définit les régions de France, les domaines, et l’emplacement des profils de recherché, afin de donner une base commune pour ces champs à tous les modules.

**Chronos.py:** ordonnanceur, permet de lancer des programmes à intervalles réguliers

**crawler\_aerocontact.py:** module de recherché d’annonce pour Aerocontact

**crawler\_aeroemploiformation.py:** module de recherché d’annonce pour Aeroemploiformation

**crawler\_apec.py:** module de recherché d’annonce pour APEC

**crawler\_caoemploi.py:** module de recherché d’annonce pour CAOemploi

**crawler\_indeed.py:** module de recherché d’annonce pour Indeed

**crawler\_monster.py:** module de recherché d’annonce pour Monster

**crawler\_poleemploi.py:** module de recherché d’annonce pour Pôle emploi

**crawler\_regionjob.py:** module de recherché d’annonce pour Région job

**Job\_crawler\_About\_window\_ponctual.py:** module de génération de la fenêtre à propos pour la recherché ponctuelle

**Job\_crawler\_About\_window\_scheduled.py:** module de génération de la fenêtre à propos pour la recherché planifiée

**Job\_crawler\_Admin\_window.py:** fenêtre d’administration

**Job\_crawler\_batch.py:** module permettant de lancer les profils de recherche planifié

**Job\_crawler\_core.py:** module central, gère l’extraction des annonces, leur traitement, stockage et envoie

**Job\_crawler\_db\_management.py:** module de gestion de la base de donnée CSV

**Job\_crawler\_Edit\_profile\_window.py:** module d’édition des profils de recherche

**Job\_crawler\_email.py:** module de gestion des e-mail

**Job\_crawler\_profiles\_management.py:** module de gestion profils

**Job\_crawler\_QT\_ponctual.py:** fenêtre de recherché ponctuelle

**Job\_crawler\_QT\_scheduled.py:** fenêtre de création de profil pour la recherché planifiés

**Job\_crawler\_toolbox.py:** module contenant des utilitaires divers pour tous les autres modules