Il y a 9 programmes qui tournent sur la machine avec des heures précises pour leur lancement à partir d’un planificateur de tâches. On peut les diviser en 2 selon leurs fonctions, les programmes de récupération de liens (recup\_liste\_a\_vendre, recup\_liste\_a\_construire, recup\_liste\_a\_louer, recup\_liste\_vacance, 01\_mise\_a\_jour\_agence, 02\_recup\_liste\_agence) et les programmes d’extractions des données (a\_vendre, terrain et 03\_agence). Donc le système est qu’il faut tout d’abord récupérer les liens de toutes les annonces et après on pourra extraire les informations.

L’ordre chronologique avec leurs heures de lancement est ci-dessous :

* recup\_liste\_a\_vendre => 17:00
* recup\_liste\_a\_construire => 19:00
* recup\_liste\_a\_louer => 21:00
* recup\_liste\_vacance => 23:00
* a\_vendre => 01:00
* terrain => 09:00
* 01\_mise\_a\_jour\_agence => 11:00
* 02\_recup\_liste\_agence => 12 :00
* 03\_agence => 18:00

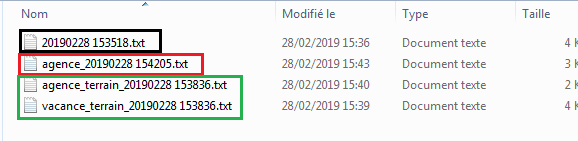
Les exportations des fichiers txt se trouvent dans « **c:\outils\resultats** ». Et ils sont regroupés dans un répertoire nommé « date du jour » pour le jour des extractions.

Dans les exportations des fichiers txt, on peut regrouper en 3 sortes selon que les informations se correspondent et sont communes et cela est dû également au ressemblance de l’apparition de chaque page web.

Lorsque vous ouvrez le répertoire « resultats », il y a un répertoire comme ceci :



Ensuite ouvert, on voit des fichiers qui peuvent être regroupées en 3 sortes :



1. Le noir regroupe toutes les annonces « à vendre, à louer, maisons ou appartements … » => traité par le programme « a\_vendre »
2. Le rouge regroupe toutes les annonces déclarées par agence. La liste de ce dernier a été récupéré à partir de 1) => traité par le programme « 03\_agence »
3. Le vert regroupe toutes les annonces nommées « terrain » qui veulent dire tout ce qui concerne « terrain, garage, construire, vacance, emplacement ». Ces derniers concernant les agences s’y trouvent également. => traité par le programme « terrain »

Ces fichiers peuvent être plusieurs selon que la machine a été éteinte et donc le reste des lignes non récupérées seront stockées vers d’autres fichiers nommés avec leur date, heure de récupération après le lancement automatique ultérieur du programme concerné.

Il est à noter qu’en backend, toutes les annonces sont regroupées dans une seule unique table.