**INSTALLATION DE APACHE SOLR**

* Installer d’abord la version 8 de Java, version minimale pour exécuter apache solr à partir de la version 7.

sudo apt install openjdk-11-jdk

* Télécharger apache solr 8.6.2

cd /opt

wget <https://downloads.apache.org/lucene/solr/8.6.2/solr-8.6.2.tgz>

* Extraire de l’archive le fichier qui permettra l’installation de solr

tar xzf solr-8.6.2.tgz solr-8.6.2/bin/install\_solr\_service.sh –strip-components=2

* Installation de solr la mise à jour de solr

sudo ./install\_solr\_service.sh solr-8.6.2.tgz -f

* Démarrer, arrêter, redémarrer ou voir l’état de solr

sudo service solr stop (arrêter solr)

sudo service solr start (lancer solr)

sudo service solr restart (redémarrer solr)

sudo service solr status (vérifier l’état du service solr)

**MAMIPULATION D’UN CŒUR DANS SOLR**

* Création d’un cœur

sudo su - solr -c "/opt/solr/bin/solr create -c <Nom du cœur>"

* Suppression d’un cœur (se placer dans le dossier

sudo /opt/solr/bin/solr delete -c <Nom du cœur>

**ACCESS AU PANEL ADMINISTRATION DE SOLR**

http://localhost:8983/solr/

**PROCÉDURE D’INDEXATION D’UN FICHIER CSV**

* Créer un fichier CSV

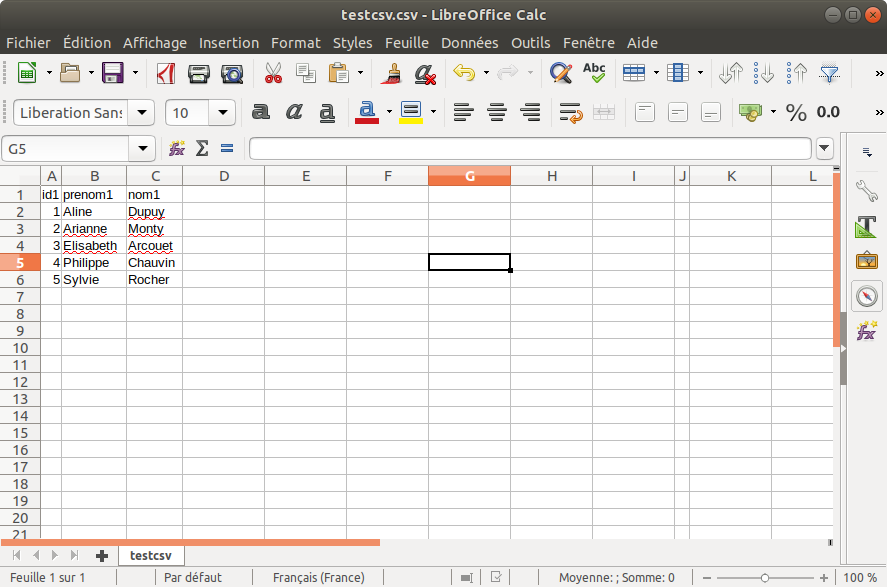


figure 1 : fichier csv crée personnellement

* Modifier le fichier managed-schema.xml qui se trouve dans le dossier **var/solr/data/<nom du core>/conf** *et insérer les spécifications de domaines de recherche correspondant à notre fichier csv*

***<field name="nom1" type="string" indexed="true" stored="true"/>***

***<field name="prenom1" type="string" indexed="true" stored="true"/>***

* Commandes d’indexation

Se placer dans le dossier ***opt/solr/bin*** *et entrer la commmande* ***sudo ./post -c <nom du core> <chemin d’acces au fichier csv à indexer>***

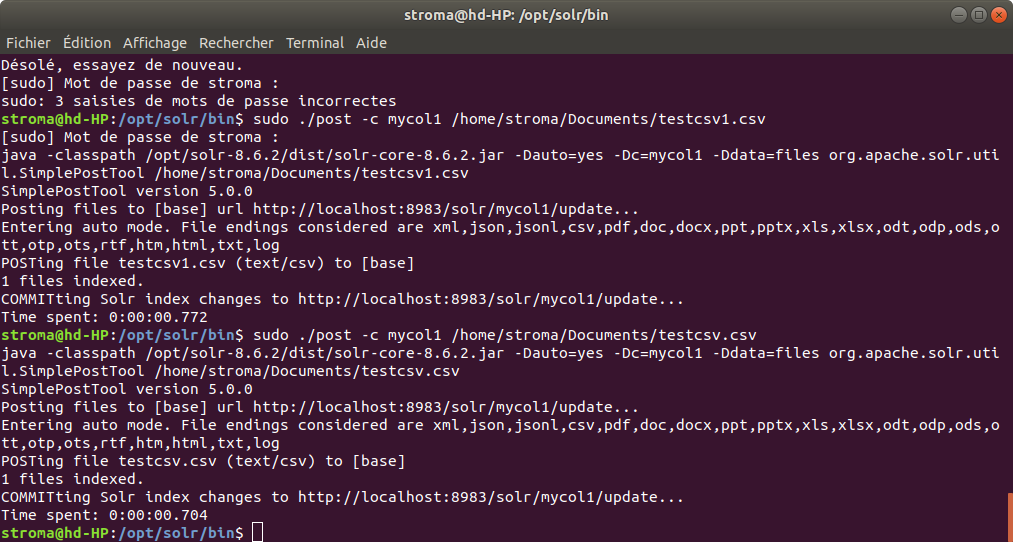
**

Figure 3 : visualisation de la commande dans le terminale avec notre fichier exemple

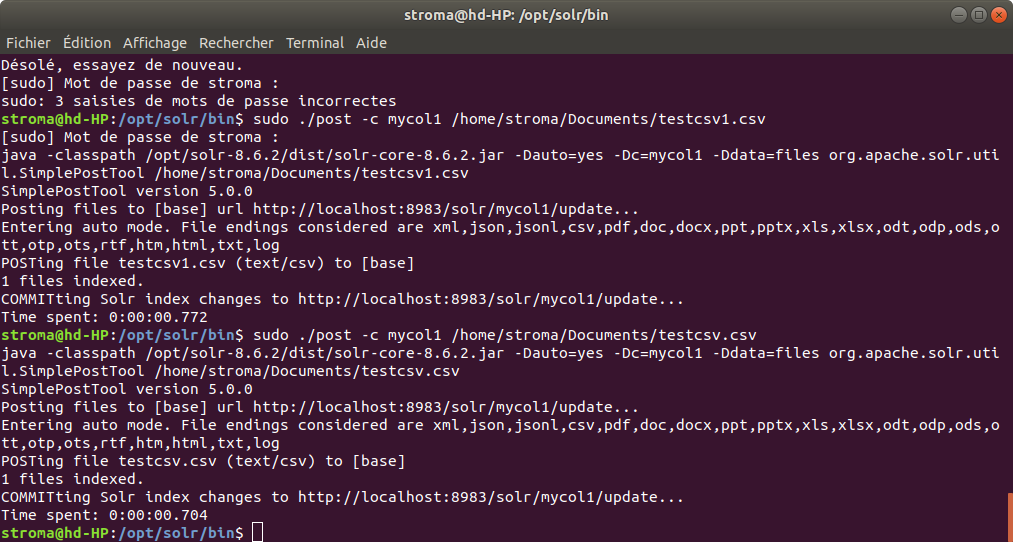


Figure 4 : Visualisation du résultat

* Effectuer une requête basique pour vérification de la bonne indexation

On veut rechercher les personnes dont le nom est ‘Aline’ ou l’id est ‘2’

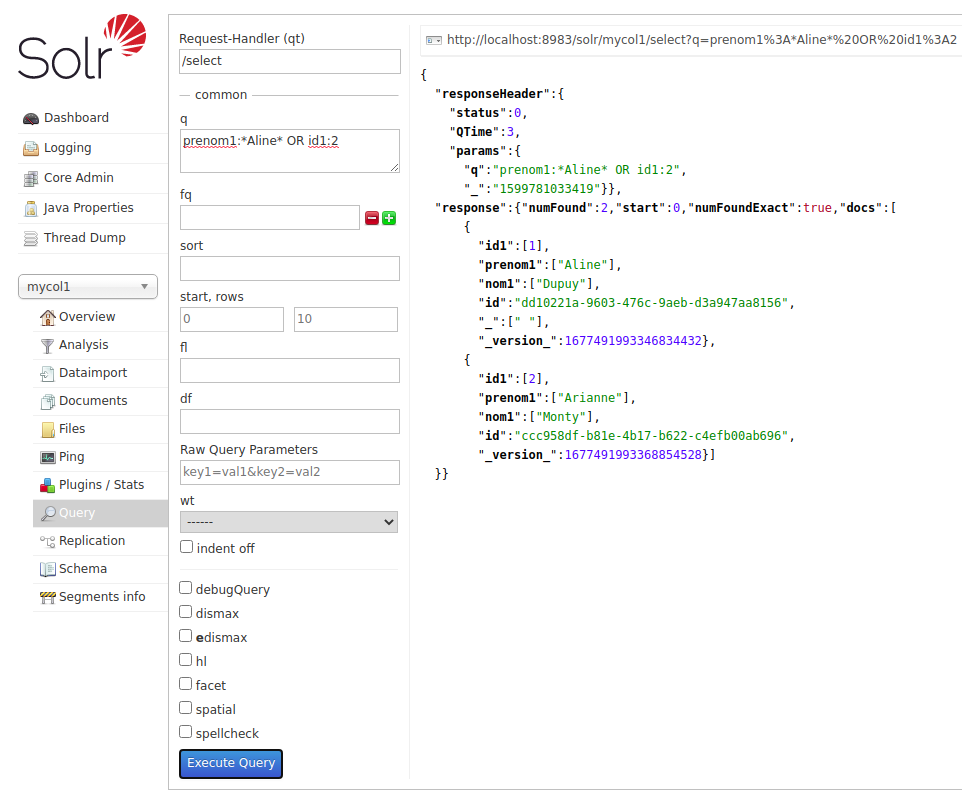


Figure 5 : Requête et affichage du résultat

* Syntaxe de quelques requête sous forme de liens
* Rechercher un seul terme : <http://localhost:8983/solr/><nom du sore>/select?q=<mot à rechercher>
* Recherche de phrase : <http://localhost:8983/solr/><nom du sore>/select?q=’’<phrase à rechercher>’’

**MISE EN PLACE DE L’OCR**

1- Commençons par éditer le fichier solrconfig.xml du core ocr situé dans le dossier **« admin:///var/solr/data/ocr/conf »** et y placer la configuration suivante :

**<requestHandler name="/update/extract" class="solr.extraction.ExtractingRequest**

**<lst name="defaults">**

**<str name="fmap.content">text</str>**

**<str name="lowernames">true</str>**

**<str name="uprefix">attr\_</str>**

**<str name="captureAttr">true</str>**

**</lst>**

**</requestHandler>**

2- Ensuite, créer un dossier « extract » n’importe où sur notre système (j’ai crée le mien dans le dossier du core ocr (admin:///var/solr/data) puis placer y le fichier apache-solr-cell-8.6.2.jar (qui se trouve dans le dossier **/opt/solr/dist**) y placer également tous les fichiers qui se trouvent dans le dossier **« /opt/solr/contrib/extraction/lib »**

3- En plus de cela, nous avons besoin de l’entrée suivantes ajoutée au fichier solrconfig.xml du core ocr

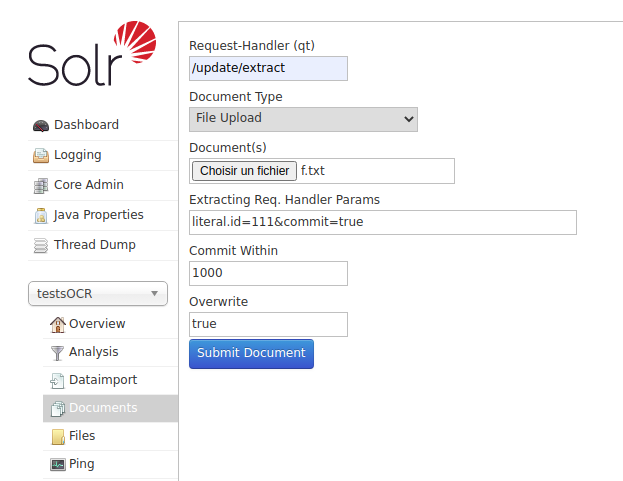
**<lib dir="admin:///var/solr/data/ocr/extract" regex=".\*\.jar" />**

4- placer la structure d’index suivante dans la section fields du fichier **« schema.xml »** ou **«  managed-schema »** qui se trouve dans le dossier **admin:///var/solr/data/ocr/conf**

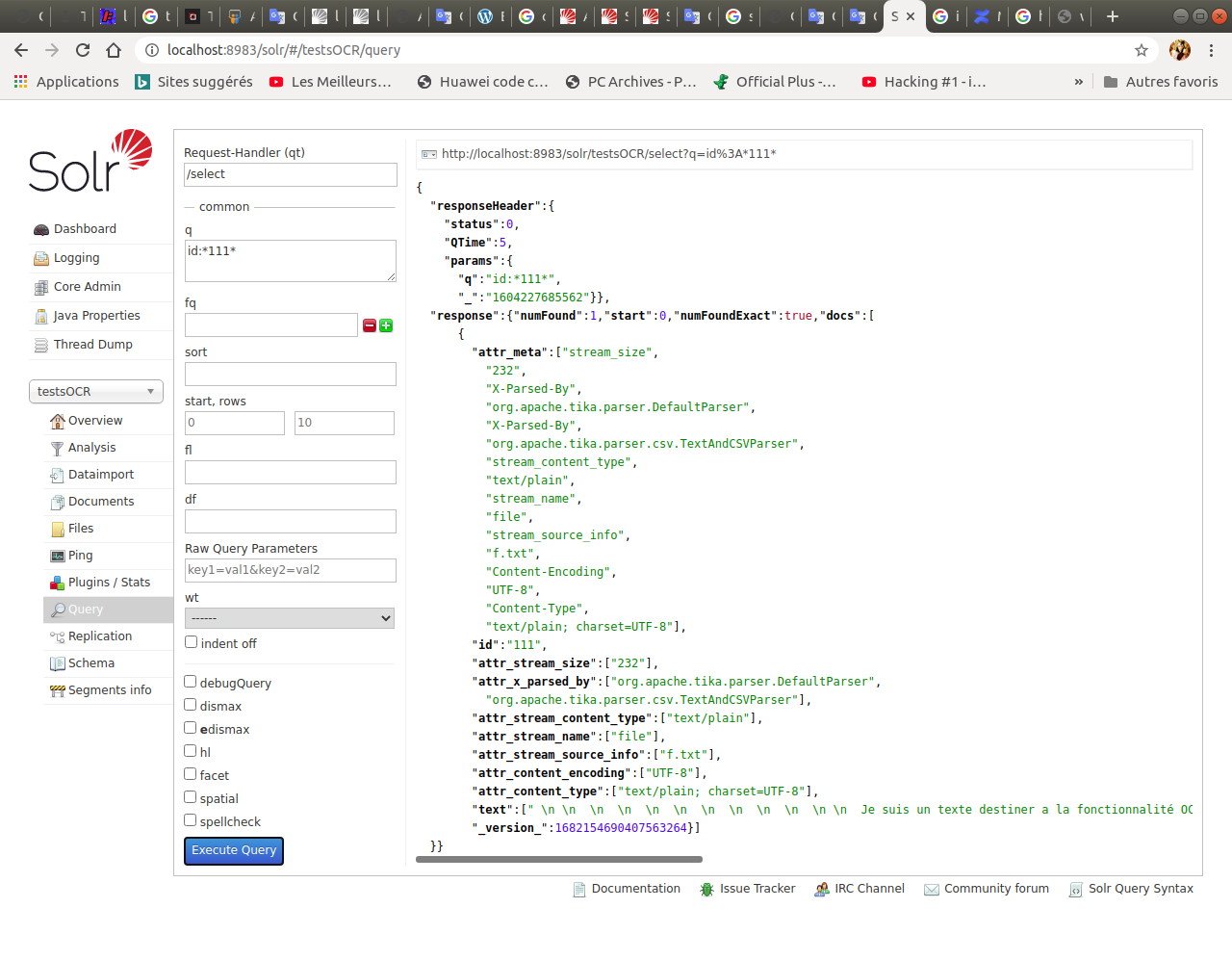
**<field name="text" type="text\_general" indexed="true" stored="true"/>**

**TEST DE L’OCR**

Mon fichier se nomme « **f.txt**» et contient le texte « Je suis un texte destiner a la fonctionnalité OCR pour apache solr. Manger bouger pour bien vivre dans la société. Installer d’abord la version 8 de Java, version minimale pour exécuter apalche solr a partir de la version 7 »



Paramétrer l’interface web de solr comme ceci (j’ai donnée l’id 111 au fichier une fois indexé)



Exécution de la requête de recherche : on s’aperçoit que l’id est bien « 111 », le nom du fichier est bien « f.txt » et la présence du texte récupéré