

Décembre 2022

---

# Devoir Maison

TP : NLP KM

---

Alexandre Bellebon - 21808613

Léo Vincent - 21805239

M2 Informatique - IDC

Année 2022-2023

## Table des matières

<b>1</b>	<b>Objectif du travail</b>	<b>2</b>
<b>2</b>	<b>Expansion de requête</b>	<b>2</b>
<b>3</b>	<b>Recherche par entités nommées</b>	<b>3</b>
<b>4</b>	<b>Démonstration</b>	<b>4</b>
4.1	Expansion d'une requête classique . . . . .	4
4.2	Requête avec recherche par entités nommées . . . . .	5

# 1 Objectif du travail

L'objectif de notre travail est de réaliser un moteur de recherche d'offres d'emploi. Pour cela, nous avons utilisés différents outils :

- Python. Pour réaliser notre moteur de recherche, nous avons décidé de le faire en python. Cela va nous faciliter la tâche à l'aide des nombreuses librairies disponibles.
- La librairie spaCy. Comme nous réalisons notre moteur de recherche en python, nous avons besoin d'une librairie pour traiter les termes utilisés dans le champ de recherche. SpaCy va donc nous permettre de pouvoir réaliser des expansions de requêtes. C'est-à-dire qu'en plus de faire une recherche avec le mot entré par l'utilisateur, nous allons en faire avec d'autres avec des mots similaires. SpaCy va nous permettre aussi de pouvoir implémenter la recherche par entités. C'est-à-dire que si l'utilisateur fait une recherche en fonction d'un lieu, d'une entreprise ou autres, spaCy va nous aider à trouver des offres avec ces entités (surtout si le nom de l'entreprise est un nom commun ou autres...)
- L'API Pôle emploi. Nous utilisons la librairie python de Pôle emploi pour faire nos recherches d'emploi. Ces recherches se font à l'aide de mot-clef.

Afin de rendre notre moteur de recherche plus pertinent et d'augmenter le nombre de résultat, nous avons développé 2 systèmes de requêtes : expansion de requête et la recherche par entités. Nous les développerons dans les sections suivantes. Enfin, nous ferons une démonstration de notre moteur de recherche et des points qu'ils restent à améliorer.

## 2 Expansion de requête

Le fonctionnement du moteur de recherche en lui-même est assez simple. Quand l'utilisateur rentre un mot (pour peu qu'il est du sens), on retourne les 20 premières offres contenant ce mot-clé. Si l'utilisateur rentrera une phrase, on fera une recherche pour chaque terme dans cette phrase. Cependant l'intérêt d'expandre la requête permettra d'augmenter drastiquement le nombre d'offres retournés et de les triés en fonction du nombre d'apparition de celles-ci.

Tout d'abord il faut stocker tous les termes intéressants utilisés dans la recherche (nom commun, verbe, nom propre) et chercher, à l'aide de spaCy, les termes les plus similaires. Nous décidons donc de récupérer les 100 mots les plus proches de celui utilisé par l'utilisateur et de garder ceux qui ont un score de similarité supérieur à 0.65. Par exemple pour le mot **boulang**er, nous obtenons ceci :

```
['boulang
```

er', 'boulanger', 'boulange', 'boulangers', 'boulangerie',  
'charcutier', 'boulangère', 'épicier', 'fromager', 'cafetier',  
'pâtissier', 'cuisinier', 'boucher', 'confiturier', 'Boulanger',  
'boulangeries', 'boulangères', 'maraîcher', 'patissier', 'confiseur', 'chocolatier']

Nous faisons ensuite des requêtes avec tous ses termes en enregistrant toutes les offres dans le même dictionnaire. Ainsi, nous allons pouvoir compter l'occurrence de chaque offre et ensuite les afficher dans l'ordre décroissants, soit des offres qui sont le plus apparues à celles qui sont le moins apparues. Pour éviter de se perdre avec les offres qui ont le même intitulé, nous les numérotons afin de les différencier.

### **3 Recherche par entités nommées**

L'intérêt ici est de pouvoir rechercher un emploi en fonction d'une entreprise, d'un groupe, d'un lieu, d'une personnalité ou divers autres termes (événements, produits, ...). Ainsi, lorsque l'utilisateur rentre un mot dans le champ de recherche, il doit préciser le type d'entité (ORG, LOC, PER, MISC). Ensuite nous faisons des recherches avec des mots similaire au terme qu'il a utilisé (comme pour l'expansion de requête), en majuscule (s'il en a utilisé) et en minuscule. Enfin, pour booster les offres contenant notre entité, nous parcourons les descriptions et vérifions si notre entité est présente.

## 4 Démonstration

Pour illustrer tout cela, nous allons faire 2 démonstrations.

### 4.1 Expansion d'une requête classique

Dans cette première démonstration, nous allons chercher un emploi de boulanger. Notre moteur de recherche va donc chercher des offres avec comme mots-clef la liste des 21 mots énoncé précédemment. Nous obtenons alors ceci :

```
Bonjour a vous,  
Bienvenue dans ce moteur de recherche de contrat de travail, merci de patienter un instant...  
Souhaitez vous faire :  
1 - Une recherche classique.  
2 - Une recherche par entités.  
1  
Veuillez entrer votre recherche : boulanger  
Recherche en cours...  
246 offres trouvés :  
[ 1 ] - Boulanger-pâtissier / Boulangère-pâtissière  
[ 2 ] - BOULANGER(E) - PATISSIER(E)  
[ 3 ] - Boulanger-pâtissier-traiteur / Boulangère-pâtissière-traite (H/F)  
[ 4 ] - Boulanger / Boulangère  
[ 5 ] - Boulanger / Boulangère (H/F)  
[ 6 ] - Boulanger / Boulangère (H/F)  
[ 7 ] - Boulanger (H/F)  
[ 8 ] - boulanger(H/F) (H/F)  
[ 9 ] - Boulanger H/F (H/F)  
[ 10 ] - Boulanger Briocheur (H/F)  
Tapez le numéro d'une offre pour plus de détail. Si vous voulez continuer, tapez sur entrée. Si vous voulez arreter, tapez stop :
```

Nous pouvons observer que les premières offres sont en adéquations avec notre recherche.

```
Intitule : Boulanger-pâtissier / Boulangère-pâtissière  
Description : Sous le contrôle du responsable de rayon vous exécutez toutes les taches nécessaires à la fabrication des produits (gâteaux , ou pains) proposés dans votre rayon, dans le respect des règles d'hygiène, de sécurité et des procédures du guide des bonnes pratiques.Vous intégrez une équipe de pâtissiers et de boulangers professionnels afin de développer et proposer l'offre dans un rayon traditionnel à fort potentiel .  
Vous respectez des procédures de nettoyage et de désinfection de votre environ de travail (labos, frigos, réserves, rayons, vitrines , matériels et vérifiez la conformité des marchandises livrées en quantité et en qualité, ainsi que les dates de conservation, et la température des produits selon les procédures en vigueur.  
Travail du lundi au samedi de 5h à 12h ou 12h30 ou 11h selon les jours de la semaine.  
Lieu : 76 - ST PIERRE LES ELBEUF  
Type de contrat : CDI  
Salaire : Mensuel de 1762,90 Euros sur 12 mois  
Horaire de travail : 35H Travail le samedi  
Vous pouvez retrouver l'offre à cette adresse : https://candidat.pole-emploi.fr/offres/recherche/detail/145TXYW  
Tapez le numéro d'une offre pour plus de détail. Si vous voulez continuer, tapez sur entrée. Si vous voulez arreter, tapez stop :
```

FIGURE 1 – Détails de l'offre n°1

Cependant si on va plus loin dans les résultats, on observe que les offres paraissent parfois hors-sujet. Cela est dû à certains termes qui sont trop éloignés du mot initial (cafetier, fromager, ...).

```
[ 141 ] - (1)Confiseur H/F (H/F)
[ 142 ] - (2)Confiseur H/F (H/F)
[ 143 ] - (3)Confiseur H/F (H/F)
[ 144 ] - (5)Confiseur H/F (H/F)
[ 145 ] - Pâtissier-chocolatier / Pâtissière-chocolatière (H/F)
[ 146 ] - (5)Chocolatier / Chocolatière
[ 147 ] - Pâtissier-chocolatier / Pâtissière-chocolatière
[ 148 ] - AIDE FROMAGER (H/F)
[ 149 ] - Commercial Café hôtel Restaurant H/F
[ 150 ] - (0)Boulangier / Boulangère
```

## 4.2 Requête avec recherche par entités nommées

Dans ce second exemple, nous cherchons un emploi dans l'entreprise Orange. La difficulté ici est qu'orange peut faire référence à la fois au fruit, à la ville, et à l'entreprise. Par conséquent, lors de la phase de prérecherche, on indique bien que l'on cherche une entreprise (ORG). Voici ce que nous obtenons :

```
Bonjour a vous,

Bienvenue dans ce moteur de recherche de contrat de travail, merci de patienter un instant...

Souhaitez vous faire :
1 - Une recherche classique.
2 - Une recherche par entités.
2

Quel est le nom de votre entité ? Orange

Quel est le type de cet entité :
ORG - compagnie, agence, institution.
LOC - groupe géographique, pays, ville.
PER - personnalité, nom de famille.
MISC - autres (nationalités, produits, ...)
ORG

Recherche en cours...

155 offres trouvés :

[ 1 ] - Technicien expert / Technicienne experte en installation intégration en télécommunications
et réseaux d'entreprise
[ 2 ] - Conseiller vendeur boutique Orange Nantes (H/F)
[ 3 ] - Technicien Fibre D3 Opérateur Orange (H/F)
[ 4 ] - Agent de sécurité - ORANGE (H/F)
[ 5 ] - Responsable Production informatique AS400 (H/F)
[ 6 ] - COMMERCIAL TERRAIN DEBUTANT F/H - DPT 84 Orange (H/F)
[ 7 ] - serveur H/F HCR ORANGE (H/F)
[ 8 ] - Manager en restauration rapide HCR ORANGE (H/F)
[ 9 ] - employé polyvalent restauration H/F HCR ORANGE
[ 10 ] - Chargé de clientèle de banque (F/H)
```

Malheureusement on voit que la probabilité de tomber sur des faux positifs est très élevée (notamment ici avec l'ambiguïté de la ville). Malgré tous les 3 premières offres correspondent bien à notre requête. Ceci dit, il faut rappeler qu'Orange a sa propre plate-forme de recrutement, et que par conséquent, peu d'offre sont présentes sur Pôle emploi.