

**Projet INF203 : Outil d'aide à la réalisation
d'état de l'art**

Membre du groupe :

- Lina Mouss
- Samia Baraka
- Eya Chemli
- Yasmine Hedeya
- Celine Rohee

Enseignant : Gayo Diallo

Année Universitaire : 2020 – 2021

Sommaire

Introduction	3
1. Réalisation du projet	4
1.1 Outil utilisé	4
1.2 Démarche	4
2. Création de l'interface	6
3. Indication d'utilisation	8
4. Conclusion	8

Introduction

L'objectif de ce projet est de mettre en œuvre un outil qui permette d'aider dans la réalisation d'un état de l'Art sur un domaine d'intérêt caractérisé par un ensemble de termes initiaux. L'idée principale est d'aider l'utilisateur, qui souhaite faire une revue littéraire sur un domaine, à réaliser une recherche sur la base PubMed et sélectionner les articles susceptibles d'être pertinents.

Nous avons donc réalisé une interface permettant d'effectuer une recherche d'articles pertinents à partir de mots clés sur Pubmed et de restituer les termes prépondérants des titres et résumés de ces articles.

Nous allons présenter notre solution dans ce rapport.

1. Réalisation du projet :

1.1 Outil utilisé :

Nous avons décidé d'utiliser R/Rstudio

1.2 Démarche :

1.2.1 Requête sur Pubmed

Tout d'abord, afin d'interroger à distance Pubmed nous avons eu recours à l'API E-utilities. L'API E-utilities permet d'effectuer des recherches sur les bases de données du NCBI notamment Pubmed , Pour cela nous avons installé le package “**reutils**” sur R qui permet d'interroger le corpus de PubMed à distance.

Ensuite , on demandera à l'utilisateur d'entrer ces mots clés ce qui nous permettra de faire une recherche avec la fonction “**esearch**” et récupérer les articles les plus pertinents avec la fonction “**efetch**” . Le seuil des documents retournés depuis Pubmed a été fixé à 100 .

1.2.2 Prétraitement et indexation

Une fois ces 100 documents retournés, nous allons procéder aux étapes de prétraitement et à l'indexation (packages (“**tm**”) et packages (“**SnowballC**”)). Les étapes de prétraitement utilisées sont les suivantes:

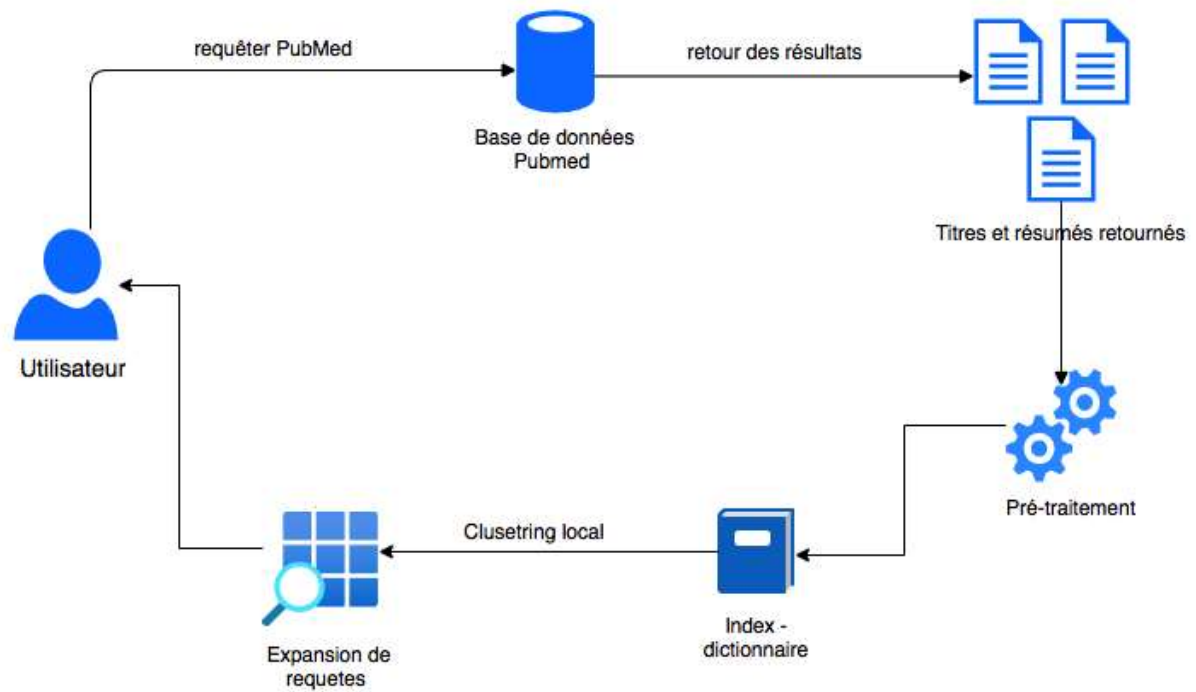
- Conversion du texte en minuscule
- Suppression des nombres
- Suppression des mots vides
- Suppression des ponctuations
- Suppression des espaces vides
- Stemming

On indexe parallèlement les titres ainsi que les abstracts.

1.2.3 Expansion de la requête

Finalement, nous procédons à l'expansion de la requête de départ . La méthode utilisée pour l'expansion de la requête est le “**clustering local** “ en construisant une matrice de corrélation terme-document à partir des articles retournés.

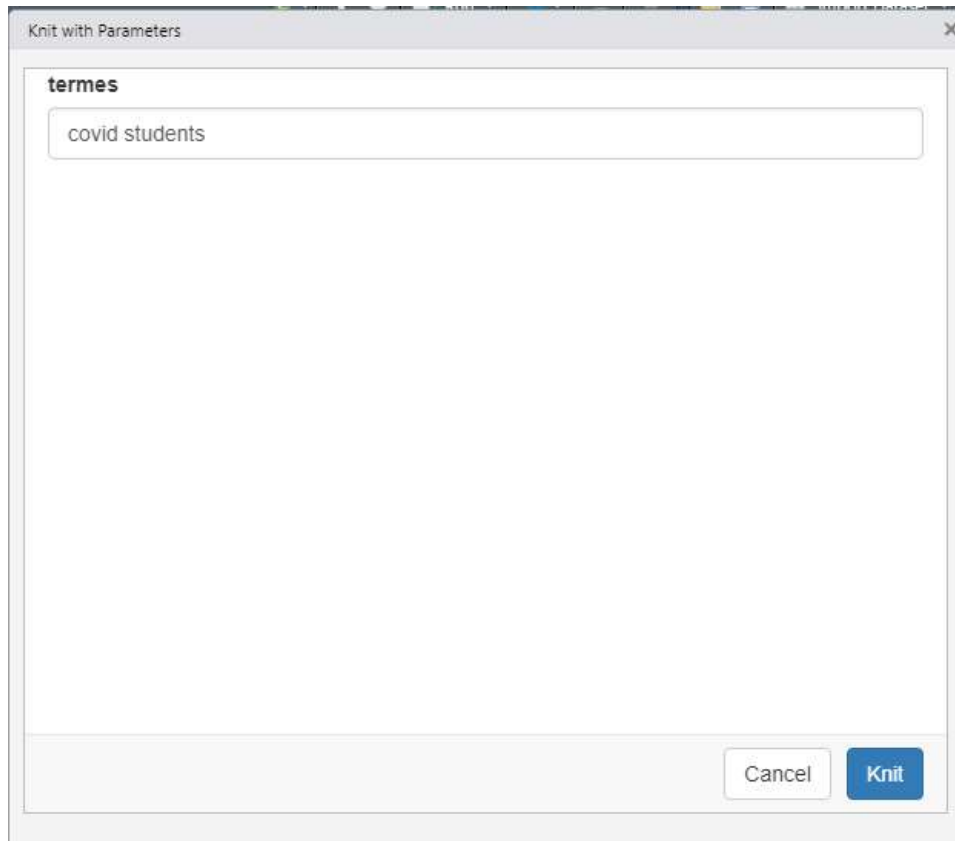
Le schéma ci-dessous résume notre démarche dans la réalisation de l'outil:



2. Affichage sur Rmarkdown :

L’affichage est effectué avec Rmarkdown qui comprend les pages suivantes:

1. Une première page “termes” permet d’écrire les mots clé (Exemple: Covid students)



The image shows a screenshot of the 'Knit with Parameters' dialog box in RStudio. The dialog has a title bar 'Knit with Parameters' and a close button. Inside, there is a section labeled 'termes' with a text input field containing 'covid students'. At the bottom right, there are two buttons: 'Cancel' and 'Knit'.

2. Page “ articles pertinents”: L’utilisateur obtient la liste des 10 meilleurs résultats correspondants : leurs titres, abstracts et un lien vers pubmed
ce lien a été obtenu grâce au PMID qu’on récupère avec les titres et abstracts

3. Indications d'utilisation :

Cet outil a été créé pour aider les utilisateurs à écrire un état de l'art en leur donnant un nombre d'articles pertinents .

Voici un petit guide d'utilisation pour pouvoir faciliter la mise en oeuvre de l'outil :

- Etape 1 : Ouvrir Rstudio
- Etape 2 : Télécharger les packages
- Etape 3 : Appuyer sur Knit Paramètres (Une fenêtre va s'ouvrir)
- Etape 4 : Ecrire les mots clés
- Etape 5 : Affichage d'une fenêtre qui présente les articles pertinents et comment affiner sa recherche si les résultats ne sont pas satisfaisants
- Etape 6 : Refaire si nécessaire une recherche avec les nouveaux mots clés proposés

4. Conclusion:

Nous avons atteint notre objectif qui était de réaliser un moteur de recherche à partir de Pubmed afin de retourner des articles pertinents dans le cadre de la réalisation d'un état de l'art.

Cependant, nous pourrions améliorer les points suivants:

- ajout de critères de recherche (exemple: recherche par date)
- création d'une notification à l'utilisateur lorsque le terme recherché n'existe pas afin qu'il puisse refaire une recherche
- ajout d'une fonction qui rende impossible le fait de récupérer plus d'une fois le même terme après l'expansion de requête
- réaliser une interface interactive
- utilisation de stopwords selon la langue de l'article et/ou des stopwords spécifiques au domaine de la santé
- ajout d'une ontologie/terminologie comme source externe pour faire l'expansion de requêtes

