Année Universitaire : 2023/2024	Université des Sciences et de la Technologie Houari Boumediene	TP N°6
Master 2 : SII	Faculté d'Informatique	Recherche de l'Information : Appariement
Module: Recherche d'Information	Département d'Intelligence Artificielle et Sciences des Données	Partie 3

Support:

1. Extraction de termes (Tokens) à l'aide de l'expression régulière suivante :

 $nltk.RegexpTokenizer('(?:[A-Za-z]\.)+|[A-Za-z]+[\-@]\d+(?:\.\d+)?|\d+[A-Za-z]+|\d+(?:[\.\,]\d+)?\%?|\w+(?:[\-/]\w+)*')$

2. Appariement:

2.1. Modèle Booléen

Entrée (requête):

La requête est représentée sous forme d'une **expression logique**. Dans cette expression, les termes sont reliés par des opérateurs booléens **OR**, **AND** et **NOT**. Ces opérateurs servent à effectuer des opérations d'union, d'intersection et de différence.

Q = < NOT TERME1 AND TERME2 OR TERME3 >

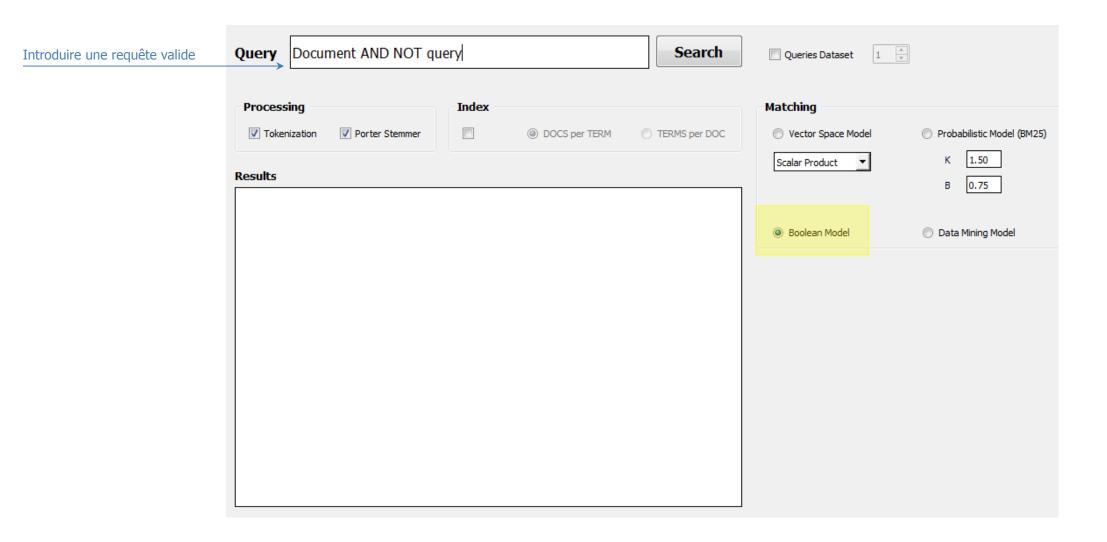
Sortie:

Les documents qui répondent à la requête logique sont considérés comme pertinents. Le degré de pertinence RSV d'un document d par rapport à une requête Q est une valeur binaire:

$$RSV(Q,d) = \{1,0\}$$

Exercice:

I. Proposer une solution permettant de vérifier la validité d'une requête logique.



Quelques exemples de requêtes valides : (les parenthèses ne sont pas autorisées dans une requête logique)

Terme

Terme **AND** Terme

Terme **OR** Terme

Terme **AND** Terme **OR** Terme

NOT Terme

NOT Terme **AND** Terme

NOT Terme **OR NOT** Terme

Terme **AND NOT** Terme

NOT Terme **AND** Terme **OR NOT** Terme

Quelques exemples de requêtes non valides :

AND

OR

Terme Terme

AND Terme

Terme **OR**

AND OR Terme

Terme **AND OR** Terme

Terme **AND** Terme **AND**

NOT

NOT NOT Terme

NOT AND Terme

Terme **AND NOT**

Terme **AND** TERME **NOT**