

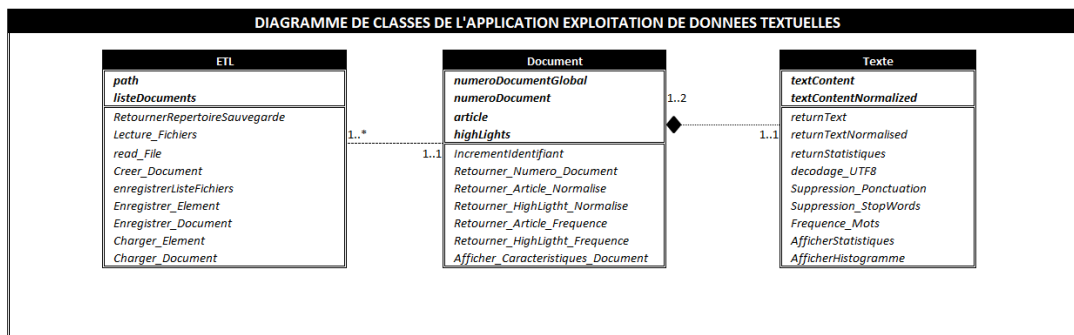
TP - Nettoyage textuel et analyse exploratoire - 2018-06-15

June 15, 2018

1 TP - Nettoyage textuel et analyse exploratoire

```
In [2]: from IPython.display import Image
        from IPython.core.display import HTML
        Image(filename = "DiagrammeClasses.png")
```

Out [2]:



2 Classes du programme

2.1 Classe ETL

```
In [120]: import pickle
          import os
          import numpy as np

          from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer, TfidfVectorizer

          #from Document import Document

          class ETL:

              def __init__(self,path):
```

```

self.path = path
os.chdir(path)

self.listeDocuments = []

self.FrequenceMotsHighLights = {}
self.FrequenceMotsArticles = {}

self.Matrice_TF_IDF_HighLights = {}
self.Matrice_TF_IDF_Articles = {}

def retourner_calcul_frequence_mots(self):
    for index,document in enumerate(self.listeDocuments):
        print('Document nř {} -----'.format(index+1))
        print('Fréquence de mots HighLight : -----')
        print(document.Retourner_HighLigtht_Frequence())

        print('Fréquence de mots Article : -----')
        print(document.Retourner_Article_Frequence())
        print('-----')

def retourner_calcul_matrice_idf(self):
    for index,document in enumerate(self.listeDocuments):
        print('Document nř {} -----'.format(index+1))
        print('Matrice TF-IDF HighLight : -----')
        print(document.Retourner_HighLigtht_Matrice_TF_IDF())

        print('Matrice TF-IDF Article : -----')
        print(document.Retourner_Article_Matrice_TF_IDF())
        print('-----')

def retourner_calcul_matrice_idf_global_Highlight(self):
    corpus = []

    for index,document in enumerate(self.listeDocuments):
        corpus.append(document.highLights.returnText())

    cvec = CountVectorizer(
        lowercase=True,
        tokenizer=None,
        token_pattern=r'\w+')

    term_freq = cvec.fit_transform(corpus)

    matrice_tfidf = TfidfTransformer().fit_transform(term_freq)

    return matrice_tfidf

```

```

def retourner_calcul_matrice_idf_global_Article(self):
    corpus = []

    for index,document in enumerate(self.listeDocuments):
        corpus.append(document.article.returnText())

    cvec = CountVectorizer(
        lowercase=True,
        tokenizer=None,
        token_pattern=r'\w+')

    term_freq = cvec.fit_transform(corpus)

    matrice_tfidf = TfidfTransformer().fit_transform(term_freq)

    return matrice_tfidf

def creation_Dictionnaire_HighLight_TF_IDF(self):
    for index,document in enumerate(self.listeDocuments):
        for items in document.Retourner_HighLightt_Matrice_TF_IDF():
            try:
                self.Matrice_TF_IDF_HighLights[items.todok().keys()] += 1
            except:
                try:
                    self.Matrice_TF_IDF_HighLights[items.todok().keys()] = 1
                except:
                    pass

    return self.Matrice_TF_IDF_HighLights

def creation_Dictionnaire_Article_TF_IDF(self):
    for index,document in enumerate(self.listeDocuments):
        for items in document.Retourner_Article_Matrice_TF_IDF():
            try:
                self.Matrice_TF_IDF_Article[items.todok().keys()] += 1
            except:
                try:
                    self.Matrice_TF_IDF_Article[items.todok().keys()] = 1
                except:
                    pass

    return self.Matrice_TF_IDF_Article

def creation_Dictionnaire_HighLight_Frequence_Mots(self):
    for index,document in enumerate(self.listeDocuments):
        for items in document.Retourner_HighLightt_Frequence():
            try:
                self.FrequenceMotsHighLights[items[0]] += items[1]

```

```

        except:
            self.FrequenceMotsHighLights[items[0]]=items[1]

    return self.FrequenceMotsHighLights

def creation_Dictionnaire_Article_Frequence_Mots(self):
    for index,document in enumerate(self.listeDocuments):
        for items in document.Retourner_Article_Frequence():
            try:
                self.FrequenceMotsArticles[items[0]]+=items[1]
            except:
                self.FrequenceMotsArticles[items[0]]=items[1]

    return self.FrequenceMotsArticles

def retournerRepertoireSauvegarde(self):
    repertoireBackUp = self.path
    return repertoireBackUp

def Lecture_Fichiers(self):
    listeFichiers = os.listdir(self.path)

    for fichier in listeFichiers:
        self.Creer_Document(fichier)

def read_File(self,nameFile):

    Content = {}

    with open(nameFile,"rb") as myFile:
        myString = str(myFile.read())
        Content['Texte'] = myString[:int(myString.find('@highlight'))]
        Content['HighLight'] = myString[int(myString.find('@highlight')):]

    return Content

def Creer_Document(self,nomFichier):
    fichier = self.read_File(nomFichier)

    document = Document(
        fichier['Texte'],
        fichier['HighLight'])

    self.listeDocuments.append(document)

def enregistrerListeFichiers(self):
    for document in self.listeDocuments:
        self.Enregistrer_Document(

```

```

document.Retourner_Numero_Document())

def Enregistrer_Element(self,numeroDocument,contenu,suffixe):
    os.chdir(self.retournerRepertoireSauvegarde())

    nameFile = "Document_" + str(numeroDocument) + suffixe

    with open(nameFile,"wb") as myFichier:
        monPickler = pickle.Pickler(myFichier)
        monPickler.dump(contenu)

def Enregistrer_Document(self,numeroDocument):
    self.Enregistrer_Element(
        numeroDocument,
        self.listeDocuments[numeroDocument-1].Retourner_Article_L
        '_Article')

    self.Enregistrer_Element(
        numeroDocument,
        self.listeDocuments[numeroDocument-1].Retourner_HighLigh
        '_HighLights')

def Charger_Element(self,numeroDocument,suffixe):
    os.chdir(self.retournerRepertoireSauvegarde())

    nameFile = "Document_" + str(numeroDocument) + suffixe

    with open(nameFile,"rb") as myFichiers:
        monDe_Pickler = pickle.Unpickler(myFichiers)
        return monDe_Pickler.load()

def Charger_Document(self,numeroDocument):
    print('-----')
    print('ARTICLE N° {} '.format(numeroDocument))
    print('-----')
    print(
        self.Charger_Element(
            numeroDocument,
            '_Article'))

    print('-----')
    print('HIGHLIGHTS N° {} '.format(numeroDocument))
    print('-----')
    print(
        self.Charger_Element(
            numeroDocument,
            '_HighLights'))

```

2.2 Classe Document

In [79]: *#from Texte import Texte*

```
class Document:

    numeroDocumentGlobal = 0

    @classmethod
    def IncrementIdentifiant(self):
        self.numeroDocumentGlobal += 1

    def __init__(self, textArticle, textHighLights):
        self.IncrementIdentifiant()
        self.numeroDocument = self.numeroDocumentGlobal

        self.article = Texte(textArticle)
        self.highLights = Texte(textHighLights)

        self.Afficher_Caracteristiques_Document()

    def Retourner_Numero_Document(self):
        return self.numeroDocument

    def Retourner_Article_Normalise(self):
        return self.article.returnTextNormalised()

    def Retourner_HighLigtht_Normalise(self):
        return self.highLights.returnTextNormalised()

    def Retourner_Article_Frequence(self):
        return self.article.ListeMotsTrie

    def Retourner_HighLigtht_Frequence(self):
        return self.highLights.ListeMotsTrie

    def Retourner_Article_Matrice_TF_IDF(self):
        return self.article.retourner_Matrice_TF_IDF()

    def Retourner_HighLigtht_Matrice_TF_IDF(self):
        return self.highLights.retourner_Matrice_TF_IDF()

    def Afficher_Caracteristiques_Document(self):
        print('-----')
        print('Caractéristiques du document {}'.format(self.Retourner_Numero_Document()))
        print('-----')
        print('HIGHLIGHTS du document {}'.format(self.Retourner_Numero_Document()))
        self.highLights.AfficherStatistiques()
```

```

print('ARTICLES du document {}'.format(self.Retourner_Numero_Document()))
print('-----')
self.article.AfficherHistogramme()

```

2.3 Classe Texte

```

In [80]: import nltk
import pandas as pd
from operator import itemgetter

class Texte:

    def __init__(self,stringData):
        self.textContent = stringData

        self.textContentNormalized = self.Suppression_StopWords(
            self.Suppression_Ponctuation(
                self.decodage_UTF8(
                    self.returnText()))))

    def returnText(self):
        return self.textContent

    def returnTextNormalised(self):

        return self.textContentNormalized

    def returnStatistiques(self):
        self.Frequence_Mots()

        return self.statsMots

    def decodage_UTF8(self,text):
        by = bytes(text,'utf-8')
        textDecode = by.decode('utf-8').lower()

        return textDecode

    def Suppression_Ponctuation(self,text):
        tokenizer = nltk.RegexpTokenizer(r'\w+')
        tokenisation = tokenizer.tokenize(text)

        return tokenisation

    def Suppression_StopWords(self,text):
        texteSansStopWords = []
        nltk.download('stopwords')

```

```

listeStopWord = nltk.corpus.stopwords.words('english')
listeStopWord.append('â')
listeStopWord.append('n')
listeStopWord.append('xc2')
listeStopWord.append('xc3')
listeStopWord.append('000')

texteSansStopWords += [word for word in text if not word in listeStopWord]

return texteSansStopWords

def Frequence_Mots(self):
    stats = nltk.FreqDist(self.textContentNormalized)
    self.NombreMots = len(self.textContentNormalized)
    self.NombreMotsUniques = len(stats)
    self.ListeMotsTrie = sorted(stats.items(), key=itemgetter(1), reverse=True)
    self.statsMots = stats

def Analyse_Matrice_TF_IDF(self, Phrases):
    cvec = CountVectorizer(
        lowercase=True,
        tokenizer=None,
        token_pattern=r'\w+')

    term_freq = cvec.fit_transform(Phrases)
    phrases_tfidf = TfidfTransformer().fit_transform(term_freq)

    self.matrice_TF_IDF = phrases_tfidf

def retourner_Matrice_TF_IDF(self):
    try:
        self.Analyse_Matrice_TF_IDF(self.textContentNormalized)

        return self.matrice_TF_IDF
    except:
        return [[1]]

def AfficherStatistiques(self):
    self.Frequence_Mots()

    print('-----')
    print('Le texte comporte {} mot(s) '.format(self.NombreMots))
    print('-----')
    print('Le texte comporte {} mot(s) unique(s) '.format(self.NombreMotsUniques))
    print('-----')

    for occurrence in self.ListeMotsTrie:
        print('Le mot {} a {} occurrence(s)'.format(occurrence[0], occurrence[1]))

```



```

print('-----')

def AfficherHistogramme(self):
    donnees = pd.Series(self.returnStatistiques())

    donneesTop = donnees.nlargest(20)
    donneesTrieesTop = donneesTop.sort_values(ascending=True)

    donneesTrieesTop.plot.barh(color="#318CE7",title="Distribution des 20 occurrences")

```

3 Mise en oeuvre du programme

3.1 Importation des données

```

In [121]: imp = ETL("E:\\Data\\RawData\\Exploit_data_texte\\Test")
          imp.Lecture_Fichiers()

```

```

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

```

Caractéristiques du document 298

HIGHLIGHTS du document 298

Le texte comporte 50 mot(s)

Le texte comporte 46 mot(s) unique(s)

Le mot highlight a 4 occurrence(s)

Le mot nhe a 2 occurrence(s)

Le mot nnicholas a 1 occurrence(s)

Le mot levene a 1 occurrence(s)

Le mot must a 1 occurrence(s)

Le mot pay a 1 occurrence(s)

Le mot nominal a 1 occurrence(s)

Le mot sum a 1 occurrence(s)

Le mot bankrupt a 1 occurrence(s)

Le mot used a 1 occurrence(s)

Le mot money a 1 occurrence(s)

Le mot private a 1 occurrence(s)

Le mot jets a 1 occurrence(s)

Le mot super a 1 occurrence(s)

Le mot yachts a 1 occurrence(s)

Le mot xa3150 a 1 occurrence(s)

Le mot year a 1 occurrence(s)
Le mot box a 1 occurrence(s)
Le mot ascot a 1 occurrence(s)
Le mot host a 1 occurrence(s)
Le mot xa310 a 1 occurrence(s)
Le mot day a 1 occurrence(s)
Le mot pheasant a 1 occurrence(s)
Le mot shoots a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot city a 1 occurrence(s)
Le mot trader a 1 occurrence(s)
Le mot spent a 1 occurrence(s)
Le mot xa3588 a 1 occurrence(s)
Le mot son a 1 occurrence(s)
Le mot bar a 1 occurrence(s)
Le mot mitzvah a 1 occurrence(s)
Le mot hiring a 1 occurrence(s)
Le mot girl a 1 occurrence(s)
Le mot band a 1 occurrence(s)
Le mot saturday a 1 occurrence(s)
Le mot tricked a 1 occurrence(s)
Le mot britain a 1 occurrence(s)
Le mot successful a 1 occurrence(s)
Le mot businessmen a 1 occurrence(s)
Le mot including a 1 occurrence(s)
Le mot stagecoach a 1 occurrence(s)
Le mot boss a 1 occurrence(s)
Le mot sir a 1 occurrence(s)
Le mot brian a 1 occurrence(s)
Le mot souter a 1 occurrence(s)

ARTICLES du document 298

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 299

HIGHLIGHTS du document 299

Le texte comporte 61 mot(s)

Le texte comporte 54 mot(s) unique(s)

Le mot highlight a 5 occurrence(s)
Le mot four a 3 occurrence(s)
Le mot tuesday a 2 occurrence(s)
Le mot nbella a 1 occurrence(s)
Le mot rodriguez a 1 occurrence(s)
Le mot torres a 1 occurrence(s)
Le mot diagnosed a 1 occurrence(s)
Le mot stage a 1 occurrence(s)
Le mot alveolar a 1 occurrence(s)
Le mot nrhabdomyosarcoma a 1 occurrence(s)
Le mot age a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot brave a 1 occurrence(s)
Le mot 10 a 1 occurrence(s)
Le mot year a 1 occurrence(s)
Le mot old a 1 occurrence(s)
Le mot fought a 1 occurrence(s)
Le mot illness a 1 occurrence(s)
Le mot came a 1 occurrence(s)
Le mot back a 1 occurrence(s)
Le mot times a 1 occurrence(s)
Le mot nher a 1 occurrence(s)
Le mot parents a 1 occurrence(s)
Le mot chronicled a 1 occurrence(s)
Le mot countless a 1 occurrence(s)
Le mot chemotherapy a 1 occurrence(s)
Le mot radiation a 1 occurrence(s)
Le mot treatments a 1 occurrence(s)
Le mot facebook a 1 occurrence(s)
Le mot 62 a 1 occurrence(s)
Le mot followers a 1 occurrence(s)
Le mot website a 1 occurrence(s)
Le mot nshe a 1 occurrence(s)
Le mot passed a 1 occurrence(s)
Le mot away a 1 occurrence(s)
Le mot peacefully a 1 occurrence(s)
Le mot surrounded a 1 occurrence(s)
Le mot family a 1 occurrence(s)
Le mot friends a 1 occurrence(s)
Le mot nat a 1 occurrence(s)
Le mot night a 1 occurrence(s)
Le mot miami a 1 occurrence(s)
Le mot heat a 1 occurrence(s)
Le mot vs a 1 occurrence(s)
Le mot pacers a 1 occurrence(s)
Le mot game a 1 occurrence(s)
Le mot lebron a 1 occurrence(s)
Le mot james a 1 occurrence(s)

Le mot wrote a 1 occurrence(s)
Le mot livelikebella a 1 occurrence(s)
Le mot sneakers a 1 occurrence(s)
Le mot honor a 1 occurrence(s)
Le mot young a 1 occurrence(s)
Le mot fan a 1 occurrence(s)

ARTICLES du document 299

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 300

HIGHLIGHTS du document 300

Le texte comporte 50 mot(s)

Le texte comporte 41 mot(s) unique(s)

Le mot highlight a 6 occurrence(s)
Le mot months a 2 occurrence(s)
Le mot cancer a 2 occurrence(s)
Le mot nshe a 2 occurrence(s)
Le mot travel a 2 occurrence(s)
Le mot njessica a 1 occurrence(s)
Le mot weller a 1 occurrence(s)
Le mot set a 1 occurrence(s)
Le mot epic a 1 occurrence(s)
Le mot adventure a 1 occurrence(s)
Le mot lifetime a 1 occurrence(s)
Le mot 2013 a 1 occurrence(s)
Le mot nbut a 1 occurrence(s)
Le mot three a 1 occurrence(s)
Le mot travelling a 1 occurrence(s)
Le mot diagnosed a 1 occurrence(s)
Le mot breast a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot 29 a 1 occurrence(s)
Le mot year a 1 occurrence(s)
Le mot old a 1 occurrence(s)
Le mot received a 1 occurrence(s)
Le mot 18 a 1 occurrence(s)
Le mot chemotherapy a 1 occurrence(s)

Le mot five a 1 occurrence(s)
Le mot surgeries a 1 occurrence(s)
Le mot treated a 1 occurrence(s)
Le mot like a 1 occurrence(s)
Le mot companion a 1 occurrence(s)
Le mot named a 1 occurrence(s)
Le mot mr a 1 occurrence(s)
Le mot lumpy a 1 occurrence(s)
Le mot xa0 a 1 occurrence(s)
Le mot pushed a 1 occurrence(s)
Le mot continued a 1 occurrence(s)
Le mot work a 1 occurrence(s)
Le mot ncancer a 1 occurrence(s)
Le mot removed a 1 occurrence(s)
Le mot last a 1 occurrence(s)
Le mot january a 1 occurrence(s)
Le mot returned a 1 occurrence(s)

ARTICLES du document 300

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 301

HIGHLIGHTS du document 301

Le texte comporte 45 mot(s)

Le texte comporte 41 mot(s) unique(s)

Le mot highlight a 3 occurrence(s)
Le mot nthe a 2 occurrence(s)
Le mot jacket a 2 occurrence(s)
Le mot glowfaster a 1 occurrence(s)
Le mot sports a 1 occurrence(s)
Le mot syncs a 1 occurrence(s)
Le mot runner a 1 occurrence(s)
Le mot xe2 a 1 occurrence(s)
Le mot x80 a 1 occurrence(s)
Le mot x99s a 1 occurrence(s)
Le mot smartphone a 1 occurrence(s)
Le mot uses a 1 occurrence(s)
Le mot flashing a 1 occurrence(s)

Le mot lights a 1 occurrence(s)
Le mot indicate a 1 occurrence(s)
Le mot whether a 1 occurrence(s)
Le mot nthey a 1 occurrence(s)
Le mot running a 1 occurrence(s)
Le mot right a 1 occurrence(s)
Le mot pace a 1 occurrence(s)
Le mot xa399 a 1 occurrence(s)
Le mot claims a 1 occurrence(s)
Le mot waterproof a 1 occurrence(s)
Le mot breathable a 1 occurrence(s)
Le mot created a 1 occurrence(s)
Le mot ex a 1 occurrence(s)
Le mot marine a 1 occurrence(s)
Le mot features a 1 occurrence(s)
Le mot zip a 1 occurrence(s)
Le mot lighting a 1 occurrence(s)
Le mot strips a 1 occurrence(s)
Le mot nit a 1 occurrence(s)
Le mot works a 1 occurrence(s)
Le mot app a 1 occurrence(s)
Le mot tracks a 1 occurrence(s)
Le mot person a 1 occurrence(s)
Le mot performance a 1 occurrence(s)
Le mot look a 1 occurrence(s)
Le mot biofeedback a 1 occurrence(s)
Le mot set a 1 occurrence(s)
Le mot goals a 1 occurrence(s)

ARTICLES du document 301

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 302

HIGHLIGHTS du document 302

Le texte comporte 89 mot(s)

Le texte comporte 73 mot(s) unique(s)

Le mot highlight a 7 occurrence(s)
Le mot reeve a 5 occurrence(s)

Le mot williams a 3 occurrence(s)
Le mot nafter a 2 occurrence(s)
Le mot old a 2 occurrence(s)
Le mot friend a 2 occurrence(s)
Le mot going a 2 occurrence(s)
Le mot none a 1 occurrence(s)
Le mot treasured a 1 occurrence(s)
Le mot relationships a 1 occurrence(s)
Le mot robin a 1 occurrence(s)
Le mot life a 1 occurrence(s)
Le mot enduring a 1 occurrence(s)
Le mot friendship a 1 occurrence(s)
Le mot fellow a 1 occurrence(s)
Le mot actor a 1 occurrence(s)
Le mot christopher a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot pair a 1 occurrence(s)
Le mot met a 1 occurrence(s)
Le mot 1973 a 1 occurrence(s)
Le mot roommates a 1 occurrence(s)
Le mot julliard a 1 occurrence(s)
Le mot school a 1 occurrence(s)
Le mot new a 1 occurrence(s)
Le mot york a 1 occurrence(s)
Le mot city a 1 occurrence(s)
Le mot ntheir a 1 occurrence(s)
Le mot careers a 1 occurrence(s)
Le mot later a 1 occurrence(s)
Le mot took a 1 occurrence(s)
Le mot time a 1 occurrence(s)
Le mot late a 1 occurrence(s)
Le mot 1970s a 1 occurrence(s)
Le mot mork a 1 occurrence(s)
Le mot mindy a 1 occurrence(s)
Le mot superman a 1 occurrence(s)
Le mot paralyzing a 1 occurrence(s)
Le mot horse a 1 occurrence(s)
Le mot riding a 1 occurrence(s)
Le mot accident a 1 occurrence(s)
Le mot 1995 a 1 occurrence(s)
Le mot revealed a 1 occurrence(s)
Le mot spirits a 1 occurrence(s)
Le mot lifted a 1 occurrence(s)
Le mot paid a 1 occurrence(s)
Le mot surprise a 1 occurrence(s)
Le mot visit a 1 occurrence(s)
Le mot nwilliams a 1 occurrence(s)
Le mot showed a 1 occurrence(s)

Le mot hospital a 1 occurrence(s)
Le mot npretending a 1 occurrence(s)
Le mot eccentric a 1 occurrence(s)
Le mot russian a 1 occurrence(s)
Le mot proctologist a 1 occurrence(s)
Le mot nperform a 1 occurrence(s)
Le mot rectal a 1 occurrence(s)
Le mot exam a 1 occurrence(s)
Le mot helped a 1 occurrence(s)
Le mot know a 1 occurrence(s)
Le mot somehow a 1 occurrence(s)
Le mot okay a 1 occurrence(s)
Le mot wrote a 1 occurrence(s)
Le mot autobiography a 1 occurrence(s)
Le mot death a 1 occurrence(s)
Le mot 2004 a 1 occurrence(s)
Le mot dedicated a 1 occurrence(s)
Le mot cecil a 1 occurrence(s)
Le mot b a 1 occurrence(s)
Le mot demille a 1 occurrence(s)
Le mot award a 1 occurrence(s)
Le mot golden a 1 occurrence(s)
Le mot globes a 1 occurrence(s)

ARTICLES du document 302

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 303

HIGHLIGHTS du document 303

Le texte comporte 59 mot(s)

Le texte comporte 52 mot(s) unique(s)

Le mot highlight a 7 occurrence(s)
Le mot found a 2 occurrence(s)
Le mot ndane a 1 occurrence(s)
Le mot kowalski a 1 occurrence(s)
Le mot mother a 1 occurrence(s)
Le mot said a 1 occurrence(s)
Le mot note a 1 occurrence(s)

Le mot revealed a 1 occurrence(s)
Le mot bitten a 1 occurrence(s)
Le mot snake a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot details a 1 occurrence(s)
Le mot suggests a 1 occurrence(s)
Le mot went a 1 occurrence(s)
Le mot scouring a 1 occurrence(s)
Le mot bushland a 1 occurrence(s)
Le mot sticks a 1 occurrence(s)
Le mot nsticks a 1 occurrence(s)
Le mot wedged a 1 occurrence(s)
Le mot underneath a 1 occurrence(s)
Le mot tyres a 1 occurrence(s)
Le mot traction a 1 occurrence(s)
Le mot ntributes a 1 occurrence(s)
Le mot flowing a 1 occurrence(s)
Le mot facebook a 1 occurrence(s)
Le mot page a 1 occurrence(s)
Le mot find a 1 occurrence(s)
Le mot dane a 1 occurrence(s)
Le mot nearlier a 1 occurrence(s)
Le mot week a 1 occurrence(s)
Le mot 27 a 1 occurrence(s)
Le mot year a 1 occurrence(s)
Le mot old a 1 occurrence(s)
Le mot ute a 1 occurrence(s)
Le mot remote a 1 occurrence(s)
Le mot highway a 1 occurrence(s)
Le mot nhe a 1 occurrence(s)
Le mot left a 1 occurrence(s)
Le mot melbourne a 1 occurrence(s)
Le mot home a 1 occurrence(s)
Le mot december a 1 occurrence(s)
Le mot 12 a 1 occurrence(s)
Le mot fishing a 1 occurrence(s)
Le mot trip a 1 occurrence(s)
Le mot darwin a 1 occurrence(s)
Le mot nthere a 1 occurrence(s)
Le mot trace a 1 occurrence(s)
Le mot police a 1 occurrence(s)
Le mot recovered a 1 occurrence(s)
Le mot body a 1 occurrence(s)
Le mot friday a 1 occurrence(s)
Le mot xa0 a 1 occurrence(s)

ARTICLES du document 303

```
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Caractéristiques du document 304

HIGHLIGHTS du document 304

Le texte comporte 50 mot(s)

Le texte comporte 42 mot(s) unique(s)

Le mot highlight a 6 occurrence(s)
Le mot servants a 2 occurrence(s)
Le mot master a 2 occurrence(s)
Le mot would a 2 occurrence(s)
Le mot nlife a 1 occurrence(s)
Le mot late a 1 occurrence(s)
Le mot victorian a 1 occurrence(s)
Le mot early a 1 occurrence(s)
Le mot edwardian a 1 occurrence(s)
Le mot britain a 1 occurrence(s)
Le mot brutal a 1 occurrence(s)
Le mot nnew a 1 occurrence(s)
Le mot staff a 1 occurrence(s)
Le mot often a 1 occurrence(s)
Le mot given a 1 occurrence(s)
Le mot generic a 1 occurrence(s)
Le mot names a 1 occurrence(s)
Le mot henry a 1 occurrence(s)
Le mot sarah a 1 occurrence(s)
Le mot ntasks a 1 occurrence(s)
Le mot included a 1 occurrence(s)
Le mot clipping a 1 occurrence(s)
Le mot toenails a 1 occurrence(s)
Le mot ironing a 1 occurrence(s)
Le mot shoelaces a 1 occurrence(s)
Le mot nmaids a 1 occurrence(s)
Le mot able a 1 occurrence(s)
Le mot turn a 1 occurrence(s)
Le mot sexual a 1 occurrence(s)
Le mot advances a 1 occurrence(s)
Le mot nover a 1 occurrence(s)
Le mot 1 a 1 occurrence(s)
Le mot 5 a 1 occurrence(s)

Le mot million a 1 occurrence(s)
Le mot british a 1 occurrence(s)
Le mot people a 1 occurrence(s)
Le mot worked a 1 occurrence(s)
Le mot nmost a 1 occurrence(s)
Le mot uk a 1 occurrence(s)
Le mot population a 1 occurrence(s)
Le mot downstairs a 1 occurrence(s)
Le mot upstairs a 1 occurrence(s)

ARTICLES du document 304

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 305

HIGHLIGHTS du document 305

Le texte comporte 43 mot(s)

Le texte comporte 40 mot(s) unique(s)

Le mot highlight a 3 occurrence(s)
Le mot 99 a 2 occurrence(s)
Le mot napple a 1 occurrence(s)
Le mot approached a 1 occurrence(s)
Le mot dozen a 1 occurrence(s)
Le mot names a 1 occurrence(s)
Le mot like a 1 occurrence(s)
Le mot florence a 1 occurrence(s)
Le mot machine a 1 occurrence(s)
Le mot effort a 1 occurrence(s)
Le mot get a 1 occurrence(s)
Le mot top a 1 occurrence(s)
Le mot talent a 1 occurrence(s)
Le mot beats a 1 occurrence(s)
Le mot exclusively a 1 occurrence(s)
Le mot nbeats a 1 occurrence(s)
Le mot music a 1 occurrence(s)
Le mot launched a 1 occurrence(s)
Le mot coming a 1 occurrence(s)
Le mot months a 1 occurrence(s)
Le mot 9 a 1 occurrence(s)

Le mot month a 1 occurrence(s)
Le mot subscription a 1 occurrence(s)
Le mot individuals a 1 occurrence(s)
Le mot family a 1 occurrence(s)
Le mot plan a 1 occurrence(s)
Le mot 14 a 1 occurrence(s)
Le mot njay a 1 occurrence(s)
Le mot z a 1 occurrence(s)
Le mot tidal a 1 occurrence(s)
Le mot thing a 1 occurrence(s)
Le mot tries a 1 occurrence(s)
Le mot convince a 1 occurrence(s)
Le mot artists a 1 occurrence(s)
Le mot sign a 1 occurrence(s)
Le mot exclusive a 1 occurrence(s)
Le mot deals a 1 occurrence(s)
Le mot beat a 1 occurrence(s)
Le mot competition a 1 occurrence(s)
Le mot xa0 a 1 occurrence(s)

ARTICLES du document 305

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 306

HIGHLIGHTS du document 306

Le texte comporte 41 mot(s)

Le texte comporte 35 mot(s) unique(s)

Le mot highlight a 4 occurrence(s)
Le mot plague a 2 occurrence(s)
Le mot nthey a 2 occurrence(s)
Le mot samples a 2 occurrence(s)
Le mot nscientists a 1 occurrence(s)
Le mot examining a 1 occurrence(s)
Le mot skeletons a 1 occurrence(s)
Le mot victims a 1 occurrence(s)
Le mot unearthed a 1 occurrence(s)
Le mot clerkenwell a 1 occurrence(s)
Le mot london a 1 occurrence(s)

Le mot compared a 1 occurrence(s)
Le mot dna a 1 occurrence(s)
Le mot taken a 1 occurrence(s)
Le mot bones a 1 occurrence(s)
Le mot recent a 1 occurrence(s)
Le mot outbreak a 1 occurrence(s)
Le mot madagascar a 1 occurrence(s)
Le mot killed a 1 occurrence(s)
Le mot 60 a 1 occurrence(s)
Le mot people a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot two a 1 occurrence(s)
Le mot almost a 1 occurrence(s)
Le mot perfect a 1 occurrence(s)
Le mot match a 1 occurrence(s)
Le mot meaning a 1 occurrence(s)
Le mot virulent a 1 occurrence(s)
Le mot today a 1 occurrence(s)
Le mot believe a 1 occurrence(s)
Le mot disease a 1 occurrence(s)
Le mot airborne a 1 occurrence(s)
Le mot spread a 1 occurrence(s)
Le mot coughs a 1 occurrence(s)
Le mot sneezes a 1 occurrence(s)

ARTICLES du document 306

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 307

HIGHLIGHTS du document 307

Le texte comporte 18 mot(s)

Le texte comporte 17 mot(s) unique(s)

Le mot highlight a 2 occurrence(s)
Le mot nbrave a 1 occurrence(s)
Le mot pensioner a 1 occurrence(s)
Le mot janet a 1 occurrence(s)
Le mot martyn a 1 occurrence(s)
Le mot 73 a 1 occurrence(s)

Le mot circled a 1 occurrence(s)
Le mot three a 1 occurrence(s)
Le mot creatures a 1 occurrence(s)
Le mot nthe a 1 occurrence(s)
Le mot londoner a 1 occurrence(s)
Le mot volunteer a 1 occurrence(s)
Le mot cango a 1 occurrence(s)
Le mot wildlife a 1 occurrence(s)
Le mot ranch a 1 occurrence(s)
Le mot south a 1 occurrence(s)
Le mot africa a 1 occurrence(s)

ARTICLES du document 307

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 308

HIGHLIGHTS du document 308

Le texte comporte 15 mot(s)

Le texte comporte 15 mot(s) unique(s)

Le mot highlight a 1 occurrence(s)
Le mot nduke a 1 occurrence(s)
Le mot cambridge a 1 occurrence(s)
Le mot reveals a 1 occurrence(s)
Le mot tore a 1 occurrence(s)
Le mot original a 1 occurrence(s)
Le mot guest a 1 occurrence(s)
Le mot list a 1 occurrence(s)
Le mot included a 1 occurrence(s)
Le mot names a 1 occurrence(s)
Le mot 777 a 1 occurrence(s)
Le mot people a 1 occurrence(s)
Le mot neither a 1 occurrence(s)
Le mot bride a 1 occurrence(s)
Le mot knew a 1 occurrence(s)

ARTICLES du document 308

[nltk_data] Downloading package stopwords to

```
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]  Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]  Package stopwords is already up-to-date!
```

Caractéristiques du document 309

HIGHLIGHTS du document 309

Le texte comporte 68 mot(s)

Le texte comporte 56 mot(s) unique(s)

Le mot highlight a 7 occurrence(s)
Le mot xe2 a 2 occurrence(s)
Le mot x80 a 2 occurrence(s)
Le mot x99s a 2 occurrence(s)
Le mot bid a 2 occurrence(s)
Le mot employer a 2 occurrence(s)
Le mot said a 2 occurrence(s)
Le mot nhacker a 1 occurrence(s)
Le mot list a 1 occurrence(s)
Le mot lets a 1 occurrence(s)
Le mot hacker a 1 occurrence(s)
Le mot jobs a 1 occurrence(s)
Le mot posted a 1 occurrence(s)
Le mot users a 1 occurrence(s)
Le mot called a 1 occurrence(s)
Le mot employers a 1 occurrence(s)
Le mot nan a 1 occurrence(s)
Le mot describes a 1 occurrence(s)
Le mot hacking a 1 occurrence(s)
Le mot task a 1 occurrence(s)
Le mot need a 1 occurrence(s)
Le mot budget a 1 occurrence(s)
Le mot nregistered a 1 occurrence(s)
Le mot hackers a 1 occurrence(s)
Le mot tasks a 1 occurrence(s)
Le mot selects a 1 occurrence(s)
Le mot work a 1 occurrence(s)
Le mot nall a 1 occurrence(s)
Le mot members a 1 occurrence(s)
Le mot listed a 1 occurrence(s)
Le mot anonymously a 1 occurrence(s)
Le mot provide a 1 occurrence(s)
Le mot email a 1 occurrence(s)
Le mot physical a 1 occurrence(s)

Le mot address a 1 occurrence(s)
Le mot registering a 1 occurrence(s)
Le mot nsite a 1 occurrence(s)
Le mot endorse a 1 occurrence(s)
Le mot condone a 1 occurrence(s)
Le mot illegal a 1 occurrence(s)
Le mot activities a 1 occurrence(s)
Le mot nit a 1 occurrence(s)
Le mot believed a 1 occurrence(s)
Le mot registered a 1 occurrence(s)
Le mot new a 1 occurrence(s)
Le mot zealand a 1 occurrence(s)
Le mot founders a 1 occurrence(s)
Le mot revealed a 1 occurrence(s)
Le mot identities a 1 occurrence(s)
Le mot nsecurity a 1 occurrence(s)
Le mot expert a 1 occurrence(s)
Le mot xa0the a 1 occurrence(s)
Le mot site a 1 occurrence(s)
Le mot people a 1 occurrence(s)
Le mot ethics a 1 occurrence(s)
Le mot skills a 1 occurrence(s)

ARTICLES du document 309

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 310

HIGHLIGHTS du document 310

Le texte comporte 27 mot(s)

Le texte comporte 24 mot(s) unique(s)

Le mot highlight a 3 occurrence(s)
Le mot sinclair a 2 occurrence(s)
Le mot nleonora a 1 occurrence(s)
Le mot found a 1 occurrence(s)
Le mot guilty a 1 occurrence(s)
Le mot manslaughter a 1 occurrence(s)
Le mot cleared a 1 occurrence(s)
Le mot murder a 1 occurrence(s)

Le mot ncourt a 1 occurrence(s)
Le mot heard a 1 occurrence(s)
Le mot lloyd a 1 occurrence(s)
Le mot frightened a 1 occurrence(s)
Le mot wife a 1 occurrence(s)
Le mot attacked a 1 occurrence(s)
Le mot previous a 1 occurrence(s)
Le mot occasions a 1 occurrence(s)
Le mot nhe a 1 occurrence(s)
Le mot stabbed a 1 occurrence(s)
Le mot thigh a 1 occurrence(s)
Le mot kitchen a 1 occurrence(s)
Le mot knife a 1 occurrence(s)
Le mot trying a 1 occurrence(s)
Le mot watch a 1 occurrence(s)
Le mot football a 1 occurrence(s)

ARTICLES du document 310

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 311

HIGHLIGHTS du document 311

Le texte comporte 35 mot(s)

Le texte comporte 33 mot(s) unique(s)

Le mot highlight a 3 occurrence(s)
Le mot nexum a 1 occurrence(s)
Le mot speight a 1 occurrence(s)
Le mot dna a 1 occurrence(s)
Le mot found a 1 occurrence(s)
Le mot body a 1 occurrence(s)
Le mot douglas a 1 occurrence(s)
Le mot stumler a 1 occurrence(s)
Le mot 30 a 1 occurrence(s)
Le mot cold a 1 occurrence(s)
Le mot case a 1 occurrence(s)
Le mot investigation a 1 occurrence(s)
Le mot advances a 1 occurrence(s)
Le mot technology a 1 occurrence(s)

Le mot nthe a 1 occurrence(s)
Le mot pair a 1 occurrence(s)
Le mot argued a 1 occurrence(s)
Le mot stolen a 1 occurrence(s)
Le mot property a 1 occurrence(s)
Le mot murder a 1 occurrence(s)
Le mot nspeight a 1 occurrence(s)
Le mot 50 a 1 occurrence(s)
Le mot continued a 1 occurrence(s)
Le mot boxing a 1 occurrence(s)
Le mot 2001 a 1 occurrence(s)
Le mot famously a 1 occurrence(s)
Le mot lost a 1 occurrence(s)
Le mot current a 1 occurrence(s)
Le mot heavyweight a 1 occurrence(s)
Le mot champion a 1 occurrence(s)
Le mot wladimir a 1 occurrence(s)
Le mot klitschko a 1 occurrence(s)
Le mot 1996 a 1 occurrence(s)

ARTICLES du document 311

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 312

HIGHLIGHTS du document 312

Le texte comporte 30 mot(s)

Le texte comporte 26 mot(s) unique(s)

Le mot highlight a 3 occurrence(s)
Le mot theory a 2 occurrence(s)
Le mot xa0 a 2 occurrence(s)
Le mot nbig a 1 occurrence(s)
Le mot bang a 1 occurrence(s)
Le mot star a 1 occurrence(s)
Le mot mayim a 1 occurrence(s)
Le mot bialik a 1 occurrence(s)
Le mot follows a 1 occurrence(s)
Le mot attachment a 1 occurrence(s)
Le mot parenting a 1 occurrence(s)

Le mot ncontroversial a 1 occurrence(s)
Le mot philosophy a 1 occurrence(s)
Le mot sees a 1 occurrence(s)
Le mot parents a 1 occurrence(s)
Le mot sleep a 1 occurrence(s)
Le mot alongside a 1 occurrence(s)
Le mot babies a 1 occurrence(s)
Le mot night a 1 occurrence(s)
Le mot baby a 1 occurrence(s)
Le mot wearing a 1 occurrence(s)
Le mot mothers a 1 occurrence(s)
Le mot carry a 1 occurrence(s)
Le mot offspring a 1 occurrence(s)
Le mot sling a 1 occurrence(s)
Le mot day a 1 occurrence(s)

ARTICLES du document 312

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 313

HIGHLIGHTS du document 313

Le texte comporte 28 mot(s)

Le texte comporte 24 mot(s) unique(s)

Le mot highlight a 3 occurrence(s)
Le mot united a 2 occurrence(s)
Le mot right a 2 occurrence(s)
Le mot nander a 1 occurrence(s)
Le mot herrera a 1 occurrence(s)
Le mot started a 1 occurrence(s)
Le mot last a 1 occurrence(s)
Le mot eight a 1 occurrence(s)
Le mot games a 1 occurrence(s)
Le mot manchester a 1 occurrence(s)
Le mot nmidfielder a 1 occurrence(s)
Le mot says a 1 occurrence(s)
Le mot team a 1 occurrence(s)
Le mot league a 1 occurrence(s)
Le mot xa0 a 1 occurrence(s)

Le mot nherrera a 1 occurrence(s)
Le mot scored a 1 occurrence(s)
Le mot twice a 1 occurrence(s)
Le mot aston a 1 occurrence(s)
Le mot villa a 1 occurrence(s)
Le mot man a 1 occurrence(s)
Le mot 3 a 1 occurrence(s)
Le mot 1 a 1 occurrence(s)
Le mot victory a 1 occurrence(s)

ARTICLES du document 313

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 314

HIGHLIGHTS du document 314

Le texte comporte 33 mot(s)

Le texte comporte 29 mot(s) unique(s)

Le mot highlight a 4 occurrence(s)
Le mot police a 2 occurrence(s)
Le mot nformer a 1 occurrence(s)
Le mot head a 1 occurrence(s)
Le mot west a 1 occurrence(s)
Le mot yorkshire a 1 occurrence(s)
Le mot would a 1 occurrence(s)
Le mot faced a 1 occurrence(s)
Le mot dismissal a 1 occurrence(s)
Le mot nsir a 1 occurrence(s)
Le mot norman a 1 occurrence(s)
Le mot questioned a 1 occurrence(s)
Le mot fairness a 1 occurrence(s)
Le mot ipcc a 1 occurrence(s)
Le mot process a 1 occurrence(s)
Le mot nleft a 1 occurrence(s)
Le mot force a 1 occurrence(s)
Le mot october a 1 occurrence(s)
Le mot last a 1 occurrence(s)
Le mot year a 1 occurrence(s)
Le mot cannot a 1 occurrence(s)

Le mot face a 1 occurrence(s)
Le mot disciplinary a 1 occurrence(s)
Le mot action a 1 occurrence(s)
Le mot nvictim a 1 occurrence(s)
Le mot mother a 1 occurrence(s)
Le mot calls a 1 occurrence(s)
Le mot stripped a 1 occurrence(s)
Le mot knighthood a 1 occurrence(s)

ARTICLES du document 314

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 315

HIGHLIGHTS du document 315

Le texte comporte 61 mot(s)

Le texte comporte 47 mot(s) unique(s)

Le mot highlight a 7 occurrence(s)
Le mot nshe a 3 occurrence(s)
Le mot lives a 2 occurrence(s)
Le mot works a 2 occurrence(s)
Le mot son a 2 occurrence(s)
Le mot child a 2 occurrence(s)
Le mot money a 2 occurrence(s)
Le mot xa0 a 2 occurrence(s)
Le mot nevita a 1 occurrence(s)
Le mot kossman a 1 occurrence(s)
Le mot husband a 1 occurrence(s)
Le mot bristol a 1 occurrence(s)
Le mot poland a 1 occurrence(s)
Le mot mateusz a 1 occurrence(s)
Le mot aged a 1 occurrence(s)
Le mot 10 a 1 occurrence(s)
Le mot nfather a 1 occurrence(s)
Le mot sends a 1 occurrence(s)
Le mot xa3220 a 1 occurrence(s)
Le mot month a 1 occurrence(s)
Le mot benefit a 1 occurrence(s)
Le mot tax a 1 occurrence(s)

Le mot credit a 1 occurrence(s)
Le mot nmother a 1 occurrence(s)
Le mot said a 1 occurrence(s)
Le mot positively a 1 occurrence(s)
Le mot surprised a 1 occurrence(s)
Le mot get a 1 occurrence(s)
Le mot handouts a 1 occurrence(s)
Le mot uk a 1 occurrence(s)
Le mot full a 1 occurrence(s)
Le mot time a 1 occurrence(s)
Le mot running a 1 occurrence(s)
Le mot successful a 1 occurrence(s)
Le mot xa0restaurant a 1 occurrence(s)
Le mot nso a 1 occurrence(s)
Le mot spends a 1 occurrence(s)
Le mot extra a 1 occurrence(s)
Le mot food a 1 occurrence(s)
Le mot toys a 1 occurrence(s)
Le mot cinema a 1 occurrence(s)
Le mot trips a 1 occurrence(s)
Le mot says a 1 occurrence(s)
Le mot xa0scrounger a 1 occurrence(s)
Le mot british a 1 occurrence(s)
Le mot system a 1 occurrence(s)
Le mot entitles a 1 occurrence(s)

ARTICLES du document 315

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 316

HIGHLIGHTS du document 316

Le texte comporte 68 mot(s)

Le texte comporte 56 mot(s) unique(s)

Le mot highlight a 5 occurrence(s)
Le mot shot a 3 occurrence(s)
Le mot rodger a 2 occurrence(s)
Le mot opened a 2 occurrence(s)
Le mot fire a 2 occurrence(s)

Le mot two a 2 occurrence(s)
Le mot friends a 2 occurrence(s)
Le mot say a 2 occurrence(s)
Le mot nbianca a 1 occurrence(s)
Le mot de a 1 occurrence(s)
Le mot kock a 1 occurrence(s)
Le mot 20 a 1 occurrence(s)
Le mot five a 1 occurrence(s)
Le mot times a 1 occurrence(s)
Le mot elliot a 1 occurrence(s)
Le mot walked a 1 occurrence(s)
Le mot sorority a 1 occurrence(s)
Le mot house a 1 occurrence(s)
Le mot nshe a 1 occurrence(s)
Le mot arm a 1 occurrence(s)
Le mot chest a 1 occurrence(s)
Le mot died a 1 occurrence(s)
Le mot nin a 1 occurrence(s)
Le mot first a 1 occurrence(s)
Le mot interview a 1 occurrence(s)
Le mot recounted a 1 occurrence(s)
Le mot seeing a 1 occurrence(s)
Le mot smile a 1 occurrence(s)
Le mot saying a 1 occurrence(s)
Le mot looked a 1 occurrence(s)
Le mot happy a 1 occurrence(s)
Le mot nafter a 1 occurrence(s)
Le mot called a 1 occurrence(s)
Le mot mother a 1 occurrence(s)
Le mot thought a 1 occurrence(s)
Le mot going a 1 occurrence(s)
Le mot die a 1 occurrence(s)
Le mot stranger a 1 occurrence(s)
Le mot found a 1 occurrence(s)
Le mot cop a 1 occurrence(s)
Le mot comforted a 1 occurrence(s)
Le mot nanother a 1 occurrence(s)
Le mot survivor a 1 occurrence(s)
Le mot nick a 1 occurrence(s)
Le mot pasichuke a 1 occurrence(s)
Le mot also a 1 occurrence(s)
Le mot spoken a 1 occurrence(s)
Le mot overcome a 1 occurrence(s)
Le mot injuries a 1 occurrence(s)
Le mot get a 1 occurrence(s)
Le mot back a 1 occurrence(s)
Le mot water a 1 occurrence(s)
Le mot polo a 1 occurrence(s)

Le mot team a 1 occurrence(s)
Le mot next a 1 occurrence(s)
Le mot season a 1 occurrence(s)

ARTICLES du document 316

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 317

HIGHLIGHTS du document 317

Le texte comporte 34 mot(s)

Le texte comporte 31 mot(s) unique(s)

Le mot highlight a 4 occurrence(s)
Le mot nkate a 1 occurrence(s)
Le mot visited a 1 occurrence(s)
Le mot blessed a 1 occurrence(s)
Le mot sacrament a 1 occurrence(s)
Le mot school a 1 occurrence(s)
Le mot islington a 1 occurrence(s)
Le mot north a 1 occurrence(s)
Le mot london a 1 occurrence(s)
Le mot nviewing a 1 occurrence(s)
Le mot progress a 1 occurrence(s)
Le mot pact a 1 occurrence(s)
Le mot plus a 1 occurrence(s)
Le mot project a 1 occurrence(s)
Le mot addresses a 1 occurrence(s)
Le mot addiction a 1 occurrence(s)
Le mot families a 1 occurrence(s)
Le mot nshe a 1 occurrence(s)
Le mot joined a 1 occurrence(s)
Le mot comedian a 1 occurrence(s)
Le mot john a 1 occurrence(s)
Le mot bishop a 1 occurrence(s)
Le mot also a 1 occurrence(s)
Le mot supports a 1 occurrence(s)
Le mot cause a 1 occurrence(s)
Le mot nduchess a 1 occurrence(s)
Le mot wore a 1 occurrence(s)

Le mot salmon a 1 occurrence(s)
Le mot pink a 1 occurrence(s)
Le mot dress a 1 occurrence(s)
Le mot goat a 1 occurrence(s)

ARTICLES du document 317

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 318

HIGHLIGHTS du document 318

Le texte comporte 36 mot(s)

Le texte comporte 32 mot(s) unique(s)

Le mot highlight a 4 occurrence(s)
Le mot midwife a 2 occurrence(s)
Le mot nelleleanor a 1 occurrence(s)
Le mot bennett a 1 occurrence(s)
Le mot died a 1 occurrence(s)
Le mot 27 a 1 occurrence(s)
Le mot hours a 1 occurrence(s)
Le mot birth a 1 occurrence(s)
Le mot february a 1 occurrence(s)
Le mot 2004 a 1 occurrence(s)
Le mot na a 1 occurrence(s)
Le mot failed a 1 occurrence(s)
Le mot take a 1 occurrence(s)
Le mot baby a 1 occurrence(s)
Le mot heart a 1 occurrence(s)
Le mot rate a 1 occurrence(s)
Le mot 43 a 1 occurrence(s)
Le mot minutes a 1 occurrence(s)
Le mot labour a 1 occurrence(s)
Le mot nan a 1 occurrence(s)
Le mot investigation a 1 occurrence(s)
Le mot carried a 1 occurrence(s)
Le mot staff a 1 occurrence(s)
Le mot furness a 1 occurrence(s)
Le mot general a 1 occurrence(s)
Le mot hospital a 1 occurrence(s)

Le mot nthe a 1 occurrence(s)
Le mot probe a 1 occurrence(s)
Le mot critical a 1 occurrence(s)
Le mot elleleanor a 1 occurrence(s)
Le mot parents a 1 occurrence(s)
Le mot told a 1 occurrence(s)

ARTICLES du document 318

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 319

HIGHLIGHTS du document 319

Le texte comporte 0 mot(s)

Le texte comporte 0 mot(s) unique(s)

ARTICLES du document 319

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

Caractéristiques du document 320

HIGHLIGHTS du document 320

Le texte comporte 0 mot(s)

Le texte comporte 0 mot(s) unique(s)

ARTICLES du document 320

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!

```
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Caractéristiques du document 321

HIGHLIGHTS du document 321

Le texte comporte 0 mot(s)

Le texte comporte 0 mot(s) unique(s)

ARTICLES du document 321

```
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Caractéristiques du document 322

HIGHLIGHTS du document 322

Le texte comporte 0 mot(s)

Le texte comporte 0 mot(s) unique(s)

ARTICLES du document 322

```
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]      C:\Users\monne\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Caractéristiques du document 323

HIGHLIGHTS du document 323

Le texte comporte 0 mot(s)

Le texte comporte 0 mot(s) unique(s)

ARTICLES du document 323

```
[nltk_data] Downloading package stopwords to  
[nltk_data]   C:\Users\monne\AppData\Roaming\nltk_data...  
[nltk_data]   Package stopwords is already up-to-date!  
[nltk_data] Downloading package stopwords to  
[nltk_data]   C:\Users\monne\AppData\Roaming\nltk_data...  
[nltk_data]   Package stopwords is already up-to-date!
```

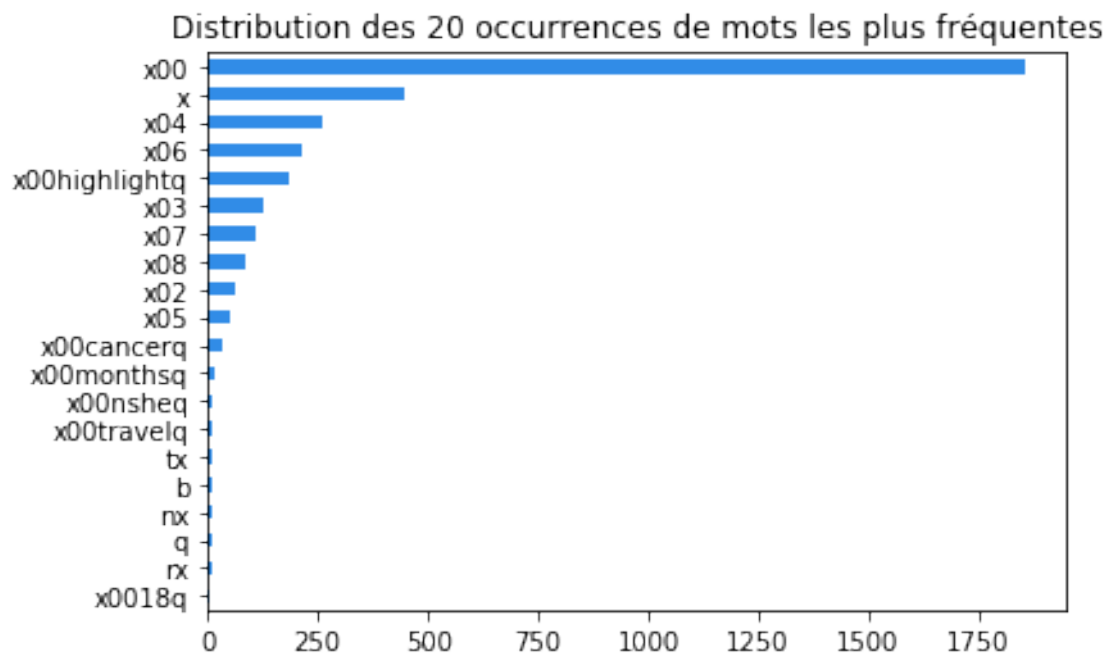
Caractéristiques du document 324

HIGHLIGHTS du document 324

Le texte comporte 0 mot(s)

Le texte comporte 0 mot(s) unique(s)

ARTICLES du document 324



3.2 Calcul des fréquences de mots pour chaque document

In [104]: `imp.retourner_calcul_frequence_mots()`

Document nř 1 -----

Fréquence de mots HighLight : -----

[('highlight', 4), ('nhe', 2), ('nnicholas', 1), ('levene', 1), ('must', 1), ('pay', 1), ('nom', 1)]

Fréquence de mots Article : -----

[('xe2', 12), ('x80', 12), ('levene', 10), ('pay', 7), ('xa31', 5), ('year', 5), ('x99s', 5), ('x99', 5)]

Document nř 2 -----

Fréquence de mots HighLight : -----

[('highlight', 5), ('four', 3), ('tuesday', 2), ('nbella', 1), ('rodriguez', 1), ('torres', 1), ('torres', 1)]

Fréquence de mots Article : -----

[('bella', 16), ('cancer', 7), ('battle', 6), ('parents', 6), ('tumors', 6), ('little', 6), ('little', 6)]

Document nř 3 -----

Fréquence de mots HighLight : -----

[('highlight', 6), ('months', 2), ('cancer', 2), ('nshe', 2), ('travel', 2), ('njessica', 1), ('njessica', 1)]

Fréquence de mots Article : -----

[('weller', 23), ('nms', 15), ('cancer', 11), ('2014', 10), ('said', 9), ('2013', 8), ('right', 8)]

Document nř 9 -----

Fréquence de mots HighLight : -----

[('highlight', 4), ('plague', 2), ('nthey', 2), ('samples', 2), ('nscientists', 1), ('examining', 1)]

Fréquence de mots Article : -----

[('nthe', 10), ('skeletons', 9), ('black', 8), ('plague', 8), ('london', 7), ('death', 7), ('x', 7)]

Document nř 10 -----

Fréquence de mots HighLight : -----

[('highlight', 2), ('nbrate', 1), ('pensioner', 1), ('janet', 1), ('martyn', 1), ('73', 1), ('one', 1)]

Fréquence de mots Article : -----

[('janet', 8), ('face', 7), ('water', 5), ('crocodiles', 5), ('species', 4), ('one', 3), ('danger', 3)]

Document nř 11 -----

Fréquence de mots HighLight : -----

[('highlight', 1), ('nduke', 1), ('cambridge', 1), ('reveals', 1), ('tore', 1), ('original', 1)]

Fréquence de mots Article : -----

[('queen', 16), ('said', 10), ('duke', 8), ('william', 7), ('prince', 7), ('wedding', 5), ('ntl', 5)]

Document nř 12 -----

Fréquence de mots HighLight : -----

[('highlight', 7), ('xe2', 2), ('x80', 2), ('x99s', 2), ('bid', 2), ('employer', 2), ('said', 2)]

Fréquence de mots Article : -----

[('xe2', 66), ('x80', 66), ('site', 28), ('xa0', 24), ('x99s', 19), ('list', 18), ('x99', 18), ('x', 18)]

Document nř 17 -----

Fréquence de mots HighLight : -----

[('highlight', 4), ('police', 2), ('nformer', 1), ('head', 1), ('west', 1), ('yorkshire', 1),

Fréquence de mots Article : -----

[('xe2', 19), ('x80', 19), ('hillsborough', 11), ('police', 10), ('norman', 9), ('said', 9), (

Document nř 18 -----

Fréquence de mots HighLight : -----

[('highlight', 7), ('nshe', 3), ('lives', 2), ('works', 2), ('son', 2), ('child', 2), ('money'

Fréquence de mots Article : -----

[('xa0', 17), ('benefits', 15), ('son', 13), ('britain', 11), ('money', 10), ('month', 9), ('s

Document nř 19 -----

Fréquence de mots HighLight : -----

[('highlight', 5), ('shot', 3), ('rodger', 2), ('opened', 2), ('fire', 2), ('two', 2), ('friend

Fréquence de mots Article : -----

[('kock', 11), ('said', 11), ('rodger', 8), ('people', 8), ('de', 7), ('also', 7), ('two', 5),

Document nř 20 -----

Fréquence de mots HighLight : -----

[('highlight', 4), ('nkate', 1), ('visited', 1), ('blessed', 1), ('sacrament', 1), ('school', 1

Fréquence de mots Article : -----

[('duchess', 26), ('children', 20), ('kate', 20), ('school', 18), ('nthe', 11), ('addiction', 1

```

Document nř 25 -----
Fréquence de mots HighLight : -----
[]
Fréquence de mots Article : -----
[('x00', 123), ('x', 20), ('x04', 12), ('x05', 9), ('x06', 7), ('x07', 6), ('x03', 5), ('x00hi
-----
Document nř 26 -----
Fréquence de mots HighLight : -----
[]
Fréquence de mots Article : -----
[('x00', 1858), ('x00x', 447), ('x01', 262), ('x02', 211), ('x04', 184), ('x06', 126), ('x05',

```

3.3 Fréquence globale des mots pour les documents de type HighLights

```
In [55]: imp.creation_Dictionnaire_HighLight_Frequence_Mots()
```

```

Out[55]: {'1': 2,
          '10': 2,
          '12': 1,
          '14': 1,
          '18': 1,
          '1970s': 1,
          '1973': 1,
          '1995': 1,
          '1996': 1,
          '20': 1,
          '2001': 1,
          '2004': 2,
          '2013': 1,
          '27': 2,
          '29': 1,
          '3': 1,
          '30': 1,
          '43': 1,
          '5': 1,
          '50': 1,
          '60': 1,
          '62': 1,
          '73': 1,
          '777': 1,

```


'9': 1,
'99': 2,
'able': 1,
'accident': 1,
'action': 1,
'activities': 1,
'actor': 1,
'addiction': 1,
'address': 1,
'addresses': 1,
'advances': 2,
'adventure': 1,
'africa': 1,
'age': 1,
'aged': 1,
'airborne': 1,
'almost': 1,
'alongside': 1,
'also': 2,
'alveolar': 1,
'anonymously': 1,
'app': 1,
'approached': 1,
'argued': 1,
'arm': 1,
'artists': 1,
'ascot': 1,
'aston': 1,
'attachment': 1,
'attacked': 1,
'autobiography': 1,
'award': 1,
'away': 1,
'b': 1,
'babies': 1,
'baby': 2,
'back': 2,
'band': 1,
'bang': 1,
'bankrupt': 1,
'bar': 1,
'beat': 1,
'beats': 1,
'believe': 1,
'believed': 1,
'benefit': 1,
'bennett': 1,
'bialik': 1,

'bid': 2,
'biofeedback': 1,
'birth': 1,
'bishop': 1,
'bitten': 1,
'blessed': 1,
'body': 2,
'bones': 1,
'boss': 1,
'box': 1,
'boxing': 1,
'brave': 1,
'breast': 1,
'breathable': 1,
'brian': 1,
'bride': 1,
'bristol': 1,
'britain': 2,
'british': 2,
'brutal': 1,
'budget': 1,
'bushland': 1,
'businessmen': 1,
'called': 2,
'calls': 1,
'cambridge': 1,
'came': 1,
'cancer': 2,
'cango': 1,
'cannot': 1,
'careers': 1,
'carried': 1,
'carry': 1,
'case': 1,
'cause': 1,
'cecil': 1,
'champion': 1,
'chemotherapy': 2,
'chest': 1,
'child': 2,
'christopher': 1,
'chronicled': 1,
'cinema': 1,
'circled': 1,
'city': 2,
'claims': 1,
'cleared': 1,
'clerkenwell': 1,

'clipping': 1,
'cold': 1,
'comedian': 1,
'comforted': 1,
'coming': 1,
'companion': 1,
'compared': 1,
'competition': 1,
'condone': 1,
'continued': 2,
'convince': 1,
'cop': 1,
'coughs': 1,
'countless': 1,
'created': 1,
'creatures': 1,
'credit': 1,
'critical': 1,
'current': 1,
'dane': 1,
'darwin': 1,
'day': 2,
'de': 1,
'deals': 1,
'death': 1,
'december': 1,
'dedicated': 1,
'demille': 1,
'describes': 1,
'details': 1,
'diagnosed': 2,
'die': 1,
'died': 2,
'disciplinary': 1,
'disease': 1,
'dismissal': 1,
'dna': 2,
'douglas': 1,
'downstairs': 1,
'dozen': 1,
'dress': 1,
'early': 1,
'eccentric': 1,
'edwardian': 1,
'effort': 1,
'eight': 1,
'elleenor': 1,
'elliot': 1,

'email': 1,
'employer': 2,
'employers': 1,
'endorse': 1,
'enduring': 1,
'entitles': 1,
'epic': 1,
'ethics': 1,
'ex': 1,
'exam': 1,
'examining': 1,
'exclusive': 1,
'exclusively': 1,
'expert': 1,
'extra': 1,
'face': 1,
'facebook': 2,
'faced': 1,
'failed': 1,
'fairness': 1,
'families': 1,
'family': 2,
'famously': 1,
'fan': 1,
'features': 1,
'february': 1,
'fellow': 1,
'find': 1,
'fire': 2,
'first': 1,
'fishing': 1,
'five': 2,
'flashing': 1,
'florence': 1,
'flowing': 1,
'followers': 1,
'follows': 1,
'food': 1,
'football': 1,
'force': 1,
'fought': 1,
'found': 5,
'founders': 1,
'four': 3,
'friday': 1,
'friend': 2,
'friends': 3,
'friendship': 1,

'frightened': 1,
'full': 1,
'furness': 1,
'game': 1,
'games': 1,
'general': 1,
'generic': 1,
'get': 3,
'girl': 1,
'given': 1,
'globes': 1,
'glowfaster': 1,
'goals': 1,
'goat': 1,
'going': 3,
'golden': 1,
'guest': 1,
'guilty': 1,
'hacker': 1,
'hackers': 1,
'hacking': 1,
'handouts': 1,
'happy': 1,
'head': 1,
'heard': 1,
'heart': 1,
'heat': 1,
'heavyweight': 1,
'helped': 1,
'henry': 1,
'herrera': 1,
'highlight': 91,
'highway': 1,
'hiring': 1,
'home': 1,
'honor': 1,
'horse': 1,
'hospital': 2,
'host': 1,
'hours': 1,
'house': 1,
'husband': 1,
'identities': 1,
'illegal': 1,
'illness': 1,
'included': 2,
'including': 1,
'indicate': 1,

'individuals': 1,
'injuries': 1,
'interview': 1,
'investigation': 2,
'ipcc': 1,
'ironing': 1,
'islington': 1,
'jacket': 2,
'james': 1,
'janet': 1,
'january': 1,
'jets': 1,
'jobs': 1,
'john': 1,
'joined': 1,
'julliard': 1,
'killed': 1,
'kitchen': 1,
'klitschko': 1,
'knew': 1,
'knife': 1,
'knighthood': 1,
'know': 1,
'kock': 1,
'kossman': 1,
'kowalski': 1,
'labour': 1,
'last': 3,
'late': 2,
'later': 1,
'launched': 1,
'league': 1,
'lebron': 1,
'left': 1,
'lets': 1,
'levene': 1,
'life': 1,
'lifetime': 1,
'lifted': 1,
'lighting': 1,
'lights': 1,
'like': 2,
'list': 2,
'listed': 1,
'livelikebella': 1,
'lives': 2,
'lloyd': 1,
'london': 2,

'londoner': 1,
'look': 1,
'looked': 1,
'lost': 1,
'lumpy': 1,
'machine': 1,
'madagascar': 1,
'man': 1,
'manchester': 1,
'manslaughter': 1,
'marine': 1,
'martyn': 1,
'master': 2,
'match': 1,
'mateusz': 1,
'mayim': 1,
'meaning': 1,
'melbourne': 1,
'members': 1,
'met': 1,
'miami': 1,
'midwife': 2,
'million': 1,
'mindy': 1,
'minutes': 1,
'mitzvah': 1,
'money': 3,
'month': 2,
'months': 3,
'mork': 1,
'mother': 3,
'mothers': 1,
'mr': 1,
'murder': 2,
'music': 1,
'must': 1,
'na': 1,
'nafter': 3,
'nall': 1,
'named': 1,
'names': 3,
'nan': 2,
'nander': 1,
'nanother': 1,
'napple': 1,
'nat': 1,
'nbeats': 1,
'nbella': 1,

'nbianca': 1,
'nbig': 1,
'nbrave': 1,
'nbut': 1,
'ncancer': 1,
'ncontroversial': 1,
'ncourt': 1,
'ndane': 1,
'nduchess': 1,
'nduke': 1,
'nearlier': 1,
'need': 1,
'neither': 1,
'nelleleanor': 1,
'nevita': 1,
'new': 2,
'next': 1,
'nexum': 1,
'nfather': 1,
'nformer': 1,
'nhacker': 1,
'nhe': 4,
'nher': 1,
'nherrera': 1,
'nick': 1,
'night': 2,
'nin': 1,
'nit': 2,
'njay': 1,
'njessica': 1,
'nkate': 1,
'nleft': 1,
'nleonora': 1,
'nlife': 1,
'nmaids': 1,
'nmidfielder': 1,
'nmost': 1,
'nmother': 1,
'nnew': 1,
'nnicholas': 1,
'nominal': 1,
'none': 1,
'norman': 1,
'north': 1,
'note': 1,
'nover': 1,
'nperform': 1,
'npretending': 1,

'nregistered': 1,
'nrhabdomyosarcoma': 1,
'nscientists': 1,
'nsecurity': 1,
'nshe': 8,
'nsir': 1,
'nsite': 1,
'nso': 1,
'nspeight': 1,
'nsticks': 1,
'ntasks': 1,
'nthe': 11,
'ntheir': 1,
'nthere': 1,
'nthey': 3,
'ntributes': 1,
'nvictim': 1,
'nviewing': 1,
'nwilliams': 1,
'occasions': 1,
'october': 1,
'offspring': 1,
'often': 1,
'okay': 1,
'old': 5,
'opened': 2,
'original': 1,
'outbreak': 1,
'overcome': 1,
'pace': 1,
'pacers': 1,
'pact': 1,
'page': 1,
'paid': 1,
'pair': 2,
'paralyzing': 1,
'parenting': 1,
'parents': 3,
'pasichuke': 1,
'passed': 1,
'pay': 1,
'peacefully': 1,
'pensioner': 1,
'people': 4,
'perfect': 1,
'performance': 1,
'person': 1,
'pheasant': 1,

'philosophy': 1,
'physical': 1,
'pink': 1,
'plague': 2,
'plan': 1,
'plus': 1,
'poland': 1,
'police': 3,
'polo': 1,
'population': 1,
'positively': 1,
'posted': 1,
'previous': 1,
'private': 1,
'probe': 1,
'process': 1,
'proctologist': 1,
'progress': 1,
'project': 1,
'property': 1,
'provide': 1,
'pushed': 1,
'questioned': 1,
'radiation': 1,
'ranch': 1,
'rate': 1,
'received': 1,
'recent': 1,
'recounted': 1,
'recovered': 1,
'rectal': 1,
'reeve': 5,
'registered': 1,
'registering': 1,
'relationships': 1,
'remote': 1,
'removed': 1,
'returned': 1,
'revealed': 3,
'reveals': 1,
'riding': 1,
'right': 3,
'robin': 1,
'rodger': 2,
'rodriguez': 1,
'roommates': 1,
'runner': 1,
'running': 2,

'russian': 1,
'sacrament': 1,
'said': 4,
'salmon': 1,
'samples': 2,
'sarah': 1,
'saturday': 1,
'say': 2,
'saying': 1,
'says': 2,
'school': 2,
'scored': 1,
'scouring': 1,
'season': 1,
'seeing': 1,
'sees': 1,
'selects': 1,
'sends': 1,
'servants': 2,
'set': 2,
'sexual': 1,
'shoelaces': 1,
'shoots': 1,
'shot': 3,
'showed': 1,
'sign': 1,
'sinclair': 2,
'sir': 1,
'site': 1,
'skeletons': 1,
'skills': 1,
'sleep': 1,
'sling': 1,
'smartphone': 1,
'smile': 1,
'snake': 1,
'sneakers': 1,
'sneezes': 1,
'somehow': 1,
'son': 3,
'sorority': 1,
'souter': 1,
'south': 1,
'speight': 1,
'spends': 1,
'spent': 1,
'spirits': 1,
'spoken': 1,

'sports': 1,
'spread': 1,
'stabbed': 1,
'staff': 2,
'stage': 1,
'stagecoach': 1,
'star': 1,
'started': 1,
'sticks': 1,
'stolen': 1,
'stranger': 1,
'stripped': 1,
'strips': 1,
'stumbler': 1,
'subscription': 1,
'successful': 2,
'suggests': 1,
'sum': 1,
'super': 1,
'superman': 1,
'supports': 1,
'surgeries': 1,
'surprise': 1,
'surprised': 1,
'surrounded': 1,
'survivor': 1,
'syncs': 1,
'system': 1,
'take': 1,
'taken': 1,
'talent': 1,
'task': 1,
'tasks': 1,
'tax': 1,
'team': 2,
'technology': 1,
'theory': 2,
'thigh': 1,
'thing': 1,
'thought': 1,
'three': 2,
'tidal': 1,
'time': 2,
'times': 2,
'today': 1,
'toenails': 1,
'told': 1,
'took': 1,

'top': 1,
'tore': 1,
'torres': 1,
'toys': 1,
'trace': 1,
'tracks': 1,
'traction': 1,
'trader': 1,
'travel': 2,
'travelling': 1,
'treasured': 1,
'treated': 1,
'treatments': 1,
'tricked': 1,
'tries': 1,
'trip': 1,
'trips': 1,
'trying': 1,
'tuesday': 2,
'turn': 1,
'twice': 1,
'two': 3,
'tyres': 1,
'uk': 2,
'underneath': 1,
'unearthed': 1,
'united': 2,
'upstairs': 1,
'used': 1,
'users': 1,
'uses': 1,
'ute': 1,
'victims': 1,
'victorian': 1,
'victory': 1,
'villa': 1,
'virulent': 1,
'visit': 1,
'visited': 1,
'volunteer': 1,
'vs': 1,
'walked': 1,
'watch': 1,
'water': 1,
'waterproof': 1,
'wearing': 1,
'website': 1,
'wedged': 1,

```

'week': 1,
'weller': 1,
'went': 1,
'west': 1,
'whether': 1,
'wife': 1,
'wildlife': 1,
'williams': 3,
'vladimir': 1,
'wore': 1,
'work': 2,
'worked': 1,
'works': 3,
'would': 3,
'wrote': 2,
'x80': 3,
'x99s': 3,
'xa0': 8,
'xa0restaurant': 1,
'xa0scrounger': 1,
'xa0the': 1,
'xa310': 1,
'xa3150': 1,
'xa3220': 1,
'xa3588': 1,
'xa399': 1,
'xe2': 3,
'yachts': 1,
'year': 5,
'york': 1,
'yorkshire': 1,
'young': 1,
'z': 1,
'zealand': 1,
'zip': 1}

```

3.4 Fréquence globale des mots pour les documents de type Articles

In [56]: `imp.creation_Dictionnaire_Article_Frequence_Mots()`

```

Out[56]: {'xe2': 180,
'x80': 187,
'levne': 10,
'pay': 14,
'xa31': 6,
'year': 42,
'x99s': 48,
'conned': 4,

```

'court': 7,
'nominal': 4,
'fraud': 4,
'xa0': 79,
'x99': 54,
'nlevene': 4,
'city': 5,
'investors': 3,
'jailed': 5,
'last': 24,
'november': 4,
'admitted': 4,
'scheme': 5,
'nhe': 23,
'money': 20,
'pictured': 27,
'days': 14,
'nthe': 90,
'successful': 4,
'nbut': 16,
'lost': 12,
'clients': 3,
'assets': 3,
'would': 30,
'could': 18,
'group': 9,
'file': 3,
'picture': 4,
'trader': 2,
'wealthy': 2,
'yesterday': 4,
'ordered': 2,
'back': 20,
'13': 9,
'years': 27,
'lucrative': 3,
'ponzi': 2,
'xa3316million': 2,
'used': 7,
'finance': 2,
'lifestyle': 2,
'private': 5,
'jets': 2,
'super': 2,
'yachts': 2,
'round': 5,
'world': 9,
'trips': 4,

'sum': 2,
'bankrupt': 2,
'seven': 8,
'found': 33,
'britain': 21,
'businessmen': 2,
'interest': 6,
'profits': 4,
'believed': 4,
'xa3101': 2,
'told': 32,
'order': 3,
'nicknamed': 2,
'beano': 2,
'love': 6,
'comic': 4,
'realisable': 2,
'make': 17,
'paid': 5,
'said': 115,
'high': 10,
'sir': 7,
'brian': 2,
'souter': 2,
'sister': 4,
'ann': 2,
'gloag': 2,
'founders': 3,
'stagecoach': 2,
'owner': 3,
'former': 6,
'football': 9,
'club': 3,
'xa32million': 2,
'eight': 8,
'bedroom': 4,
'property': 4,
'barnet': 2,
'north': 9,
'london': 24,
'xa3150': 2,
'box': 7,
'ascot': 2,
'gambling': 2,
'luxury': 2,
'one': 40,
'take': 11,
'funds': 2,

'financial': 2,
'israel': 3,
'fraudster': 2,
'spent': 6,
'xa3588': 2,
'second': 6,
'son': 23,
'bar': 2,
'mitzvah': 2,
'celebration': 2,
'featured': 2,
'performance': 7,
'saturdays': 2,
'several': 4,
'evidence': 6,
'amount': 3,
'b': 28,
'nalex': 1,
'ward': 1,
'na': 19,
'millions': 1,
'pounds': 2,
'nnicholas': 1,
'48': 1,
'orchestrating': 1,
'raked': 1,
'lavish': 1,
'nmust': 1,
'nicholas': 1,
'arriving': 2,
'nnow': 1,
'given': 8,
'serious': 6,
'office': 1,
'xa332': 1,
'352': 1,
'027': 1,
'685': 1,
'406': 1,
'pocket': 2,
'nit': 10,
'unclear': 1,
'much': 19,
'nthis': 4,
'victims': 5,
'recouped': 1,
'justine': 1,
'davidge': 1,

'representing': 1,
'sfo': 1,
'southwark': 1,
'crown': 1,
'subject': 1,
'nbankruptcy': 1,
'since': 7,
'october': 6,
'2009': 2,
'ongoing': 1,
'ninvestigations': 1,
'njailed': 1,
'left': 28,
'trading': 1,
'floor': 1,
'1990': 1,
'right': 28,
'wife': 7,
'tracy': 1,
'nshe': 36,
'added': 15,
'anything': 3,
'seized': 1,
'nbe': 1,
'dealt': 2,
'bankruptcy': 1,
'officials': 1,
'x98in': 1,
'respect': 3,
'namount': 2,
'suggest': 2,
'nseven': 1,
'x98it': 4,
'may': 19,
'future': 1,
'mr': 16,
'come': 15,
'ninto': 1,
'ripping': 1,
'series': 5,
'nof': 4,
'fliers': 1,
'including': 7,
'bus': 1,
'rail': 1,
'richard': 1,
'caring': 1,
'ivy': 1,

'le': 1,
'caprice': 1,
'restaurants': 1,
'west': 9,
'end': 3,
'russell': 1,
'nbartlett': 1,
'director': 2,
'r3': 1,
'investment': 1,
'hull': 1,
'ncity': 1,
'childhood': 1,
'book': 4,
'worker': 1,
'estimated': 3,
'nwealth': 1,
'xa315million': 1,
'xa320million': 1,
'2005': 1,
'nhigh': 1,
'flyer': 1,
'owning': 1,
'nlavish': 1,
'ran': 2,
'multi': 1,
'million': 5,
'pound': 1,
'illegal': 12,
'hosting': 1,
'xa310': 1,
'day': 22,
'pheasant': 1,
'shoots': 2,
'addicted': 3,
'spending': 3,
'fortunes': 1,
'spread': 9,
'betting': 1,
'insatiable': 1,
'taste': 2,
'deputy': 2,
'chairman': 1,
'nleyton': 1,
'orient': 1,
'count': 2,
'false': 1,
'accounting': 1,

'obtaining': 1,
'transfer': 2,
'deception': 1,
'12': 8,
'peter': 1,
'paul': 1,
'nand': 12,
'move': 2,
'accounts': 13,
'havens': 1,
'jersey': 1,
'nswitzerland': 1,
'nseeing': 1,
'stars': 2,
'girl': 6,
'band': 3,
'nspent': 1,
'big': 8,
'meant': 4,
'potential': 2,
'considered': 1,
'6million': 1,
'nvictim': 1,
'co': 5,
'brother': 2,
'xa310million': 1,
'nwith': 2,
'network': 1,
'contacts': 1,
'nstrong': 2,
'reputation': 3,
'people': 37,
'faith': 2,
'seemingly': 1,
'concrete': 1,
'ninvestment': 1,
'deals': 5,
'commission': 2,
'fee': 2,
'married': 4,
'father': 7,
'three': 14,
'took': 6,
'nmillions': 1,
'promising': 1,
'invest': 1,
'rights': 3,
'issue': 3,

'releases': 1,
'companies': 2,
'hsbc': 1,
'lloyds': 1,
'tsb': 1,
'nmining': 1,
'firms': 1,
'xstrata': 1,
'rio': 1,
'tinto': 1,
'dug': 1,
'ever': 3,
'deepening': 1,
'hole': 1,
'fob': 1,
'excuses': 1,
'nliving': 1,
'life': 32,
'chauffeur': 1,
'driven': 4,
'bentley': 1,
'went': 16,
'holidays': 1,
'lasting': 1,
'weeks': 6,
'ninvestigators': 1,
'yacht': 1,
'hire': 2,
'top': 2,
'hotel': 1,
'stays': 1,
'australia': 3,
'south': 13,
'africa': 3,
'fleet': 1,
'ncars': 1,
'nwhich': 1,
'girlband': 1,
'main': 1,
'nhouse': 1,
'nhis': 4,
'huge': 1,
'investigators': 1,
'finding': 4,
'blowing': 1,
'xa3720': 1,
'cricket': 2,
'match': 5,

'bet': 1,
'2007': 3,
'nhaving': 1,
'seizure': 1,
'judge': 2,
'martin': 2,
'beddow': 1,
'x98as': 1,
'nothing': 5,
'navailable': 1,
'direct': 1,
'payable': 1,
'nto': 2,
'appear': 1,
'hearing': 6,
'bella': 16,
'cancer': 20,
'battle': 7,
'parents': 13,
'tumors': 6,
'little': 12,
'heat': 5,
'four': 15,
'doctors': 6,
'lebron': 4,
'james': 6,
'nbella': 4,
'rodriguez': 4,
'torres': 4,
'basketball': 5,
'family': 45,
'tuesday': 6,
'sad': 7,
'2013': 18,
'night': 13,
'grueling': 3,
'six': 11,
'loved': 11,
'diagnosed': 6,
'illness': 4,
'despite': 8,
'times': 13,
'raymond': 3,
'mother': 20,
'shannah': 3,
'chemotherapy': 6,
'fan': 6,
'nin': 17,

'surgery': 5,
'16': 2,
'est': 12,
'29': 8,
'gave': 3,
'name': 4,
'sneakers': 2,
'play': 4,
'away': 7,
'fighting': 3,
'disease': 5,
'age': 4,
'returned': 6,
'fight': 4,
'condition': 3,
'peacefully': 2,
'surrounded': 2,
'bottom': 2,
'happy': 7,
'rayna': 2,
'cbs4': 2,
'like': 16,
'miami': 2,
'game': 5,
'dwyaner': 2,
'livelikebella': 2,
'honor': 2,
'brave': 3,
'nheat': 2,
'puts': 2,
'everything': 4,
'perspective': 2,
'news': 5,
'jesus': 2,
'radiation': 4,
'rid': 3,
'remission': 2,
'made': 15,
'brain': 4,
'plus': 10,
'january': 4,
'new': 34,
'abdomen': 2,
'treatment': 8,
'month': 20,
'many': 12,
'hospice': 2,
'care': 7,

'decision': 3,
'peace': 3,
'nhelen': 1,
'pow': 1,
'npublished': 6,
'n09': 1,
'nupdated': 6,
'n10': 2,
'51': 1,
'nmiami': 1,
'superstar': 1,
'touching': 1,
'tribute': 2,
'biggest': 2,
'fans': 4,
'wearing': 6,
'hours': 5,
'10': 20,
'watching': 3,
'passed': 4,
'half': 11,
'young': 3,
'aggressive': 3,
'form': 5,
'astounded': 1,
'recovered': 1,
'something': 10,
'deemed': 1,
'miracle': 1,
'putting': 1,
'another': 6,
'fierce': 1,
'deteriorated': 1,
'slipped': 1,
'friends': 34,
'nscroll': 8,
'video': 9,
'nbrave': 2,
'died': 16,
'nfamily': 2,
'aged': 2,
'lose': 3,
'instead': 5,
'reward': 3,
'eternal': 1,
'younger': 2,
'nsister': 1,
'coined': 1,

'phrase': 2,
'live': 4,
'extensive': 1,
'sessions': 2,
'ntuesday': 1,
'vs': 1,
'pacers': 1,
'wade': 1,
'nwore': 1,
'shoes': 4,
'written': 3,
'came': 9,
'story': 5,
'today': 13,
'brainer': 1,
'according': 9,
'yahoo': 1,
'sport': 1,
'two': 25,
'great': 10,
'teams': 1,
'competing': 1,
'kid': 1,
'loses': 1,
'happened': 6,
'watch': 14,
'copy': 1,
'idol': 1,
'star': 5,
'wore': 4,
'bell': 1,
'nshannah': 1,
'broke': 3,
'daughter': 5,
'facebook': 12,
'page': 3,
'62': 1,
'followers': 1,
'nquoting': 1,
'bible': 1,
'wrote': 4,
'received': 3,
'drink': 2,
'finished': 1,
'nbowed': 1,
'head': 9,
'spirit': 1,
'nbefore': 2,

'adding': 3,
'joined': 5,
'heaven': 1,
'bubbly': 1,
'toddler': 2,
'sponge': 1,
'bob': 1,
'square': 3,
'pants': 1,
'playing': 4,
'lego': 1,
'tender': 1,
'became': 5,
'paralyzed': 1,
'ntumor': 1,
'spine': 2,
'stage': 4,
'alveolar': 1,
'nrhabdomyosarcoma': 1,
'nwas': 1,
'thrown': 1,
'54': 1,
'n16': 1,
'battled': 1,
'body': 19,
'sites': 4,
'soon': 5,
'able': 9,
'walk': 3,
'run': 1,
'jump': 1,
'exclaimed': 1,
'website': 4,
'time': 23,
'nfive': 1,
'five': 7,
'course': 9,
'short': 1,
'nclose': 1,
'side': 7,
'throughout': 3,
'dogged': 2,
'april': 2,
'tumor': 1,
'underwent': 2,
'full': 6,
'nchemotherapy': 1,
'however': 2,

'nseptember': 1,
'2011': 4,
'attacked': 3,
'ovary': 1,
'following': 8,
'discovered': 11,
'later': 10,
'pelvis': 1,
'undergoing': 1,
'grown': 6,
'seriousness': 1,
'led': 3,
'recommend': 1,
'intensive': 1,
'march': 9,
'traveled': 1,
'houston': 1,
'hour': 8,
'remove': 2,
'uterus': 1,
'ovaries': 1,
'entire': 3,
'washed': 2,
'hot': 2,
'chemo': 4,
'nthough': 2,
'treatments': 1,
'organs': 1,
'began': 4,
'failing': 2,
'nher': 3,
'kidney': 1,
'lung': 1,
'functions': 1,
'greatly': 1,
'affected': 6,
'bone': 1,
'marrow': 1,
'exhausted': 2,
'ultrasounds': 1,
'lungs': 3,
'showed': 4,
'appeared': 3,
'lining': 1,
'ntributes': 3,
'tributes': 1,
'piling': 1,
'inspired': 1,

'courage': 1,
'nhospice': 1,
'beginning': 2,
'nat': 1,
'start': 7,
'difficult': 6,
'stop': 4,
'entered': 1,
'words': 5,
'describe': 1,
'feels': 1,
'magnitude': 1,
'painstaking': 1,
'decide': 2,
'longer': 3,
'ensure': 4,
'comfort': 1,
'earlier': 2,
'nthrough': 1,
'chronicled': 1,
'progress': 7,
'highs': 1,
'extremely': 3,
'low': 1,
'lows': 1,
'followed': 1,
'thousands': 4,
'around': 9,
'well': 11,
'wishers': 1,
'donated': 1,
'help': 12,
'prayed': 1,
'heartbroken': 2,
'asked': 9,
'please': 1,
'pray': 1,
'us': 15,
'strength': 1,
'continue': 6,
'nclick': 1,
'xa0here': 1,
'donate': 2,
'charity': 10,
'fund': 2,
'weller': 23,
'nms': 15,
'2014': 12,

'ms': 10,
'zealand': 6,
'europe': 6,
'travel': 5,
'pain': 6,
'felt': 4,
'friend': 25,
'hair': 14,
'italy': 3,
'trip': 8,
'february': 5,
'chest': 4,
'september': 10,
'wig': 3,
'want': 13,
'kate': 25,
'mind': 3,
'travelling': 3,
'front': 5,
'tower': 2,
'paris': 2,
'gelato': 2,
'terre': 2,
'place': 11,
'inside': 4,
'old': 26,
'see': 11,
'months': 11,
'breast': 3,
'home': 25,
'teacher': 2,
'determined': 3,
'even': 14,
'lumpy': 2,
'amazing': 5,
'struggled': 3,
'feel': 7,
'lot': 5,
'knew': 8,
'best': 7,
'giant': 2,
'kirsten': 2,
'way': 15,
'uk': 10,
'week': 8,
'hospital': 17,
'work': 21,
'results': 2,

'return': 4,
'getting': 5,
'going': 11,
'helpless': 2,
'every': 11,
'december': 8,
'finishing': 2,
'still': 8,
'3': 6,
'positive': 3,
'process': 12,
'bid': 6,
'france': 2,
'catherine': 5,
'bridge': 2,
'august': 2,
'kept': 4,
'helped': 8,
'working': 8,
'cap': 2,
'pushed': 2,
'capable': 3,
'mentally': 2,
'public': 9,
'posed': 1,
'eiffel': 1,
'xa0cinque': 1,
'jessica': 1,
'thought': 7,
'photograph': 1,
'taken': 10,
'suite': 1,
'ohakune': 1,
'plans': 1,
'enjoy': 2,
'lifetime': 1,
'holiday': 2,
'nrather': 1,
'school': 26,
'adventure': 1,
'pet': 1,
'treated': 2,
'bad': 1,
'companion': 1,
'njessica': 1,
'outside': 3,
'hampton': 1,
'sits': 1,

'favourite': 4,
'flower': 1,
'daffodils': 1,
'backpacking': 1,
'tossing': 1,
'turning': 1,
'bed': 5,
'sleep': 6,
'sharp': 1,
'unusual': 1,
'coming': 4,
'large': 5,
'ball': 2,
'gristle': 1,
'trapped': 1,
'daily': 1,
'mail': 4,
'veins': 1,
'appearing': 1,
'surface': 1,
'breasts': 1,
'look': 16,
'normal': 3,
'touch': 1,
'wrong': 6,
'remembered': 3,
'early': 7,
'detection': 2,
'protection': 1,
'backpack': 1,
'moved': 2,
'houses': 2,
'met': 5,
'blonde': 1,
'cambridge': 16,
'punting': 1,
'river': 1,
'cam': 1,
'20': 8,
'walked': 2,
'27': 5,
'frantically': 1,
'waited': 2,
'whittington': 1,
'test': 4,
'symptoms': 1,
'checked': 2,
'tried': 5,

'prepare': 1,
'worst': 1,
'prior': 1,
'deep': 1,
'good': 11,
'sat': 5,
'someone': 6,
'screaming': 2,
'ears': 1,
'completely': 2,
'numb': 1,
'single': 1,
'ounce': 1,
'control': 1,
'gone': 5,
'feeling': 2,
'breathe': 1,
'christmas': 3,
'far': 4,
'cancerversary': 1,
'boob': 1,
'cake': 1,
'june': 4,
'pink': 6,
'xa0waitangi': 1,
'pub': 1,
'crawl': 1,
'dressed': 1,
'maori': 1,
'colour': 2,
'song': 6,
'nice': 2,
'freezing': 1,
'endured': 1,
'centimetre': 1,
'tumour': 1,
'grade': 1,
'2': 2,
'her2': 1,
'already': 3,
'starting': 1,
'lymphatic': 1,
'system': 5,
'18': 2,
'surgeries': 1,
'eggs': 1,
'frozen': 2,
'buds': 1,

'nduring': 2,
'penned': 2,
'journey': 1,
'blog': 7,
'xa0in': 2,
'explain': 2,
'update': 3,
'faced': 3,
'skiing': 1,
'straight': 2,
'2015': 2,
'blown': 1,
'charles': 1,
'prague': 1,
'woman': 2,
'nutella': 1,
'ndespite': 1,
'travels': 1,
'keep': 7,
'booking': 1,
'never': 6,
'dominated': 1,
'obviously': 6,
'adjust': 1,
'allowances': 1,
'got': 10,
'wanted': 8,
'focused': 1,
'push': 1,
'also': 39,
'occupied': 1,
'looking': 4,
'wish': 2,
'rested': 1,
'xa0bavarian': 1,
'alps': 1,
'kelhstienhaus': 1,
'hitlers': 1,
'eagles': 1,
'nest': 1,
'diagnosis': 1,
'describes': 4,
'versailles': 1,
'opulent': 1,
'stunning': 1,
'alone': 3,
'cooling': 1,
'first': 14,

'session': 2,
'july': 1,
'freeze': 1,
'follicles': 1,
'trick': 1,
'falling': 2,
'nleft': 1,
'holding': 2,
'leaning': 1,
'pisa': 1,
'xa0quinta': 1,
'da': 1,
'regaleira': 1,
'garden': 2,
'sintra': 1,
'remained': 1,
'arranged': 1,
'medical': 3,
'appointments': 1,
'allowed': 1,
'primary': 2,
'schools': 2,
'whole': 3,
'exploring': 1,
'experiencing': 1,
'different': 7,
'cultures': 1,
'soul': 1,
'ways': 1,
'heal': 1,
'hit': 8,
'portugal': 1,
'behind': 6,
'hill': 6,
'hard': 3,
'know': 13,
'put': 7,
'incredibly': 1,
'things': 13,
'fall': 1,
'elle': 1,
'fourth': 1,
'shine': 2,
'marathon': 1,
'cinque': 1,
'sarah': 2,
'ally': 1,
'xa0neuschwanstein': 1,

```

'castle': 1,
'germany': 1,
'lives': 12,
'auckland': 1,
'goodbye': 1,
'thanks': 4,
'surgeyr': 1,
'clear': 4,
'celebrated': 1,
'style': 11,
'share': 3,
'toast': 1,
'victory': 1,
'currently': 4,
'project': 11,
'conferences': 1,
'usa': 1,
'gather': 1,
'resources': 2,
'information': 11,
'need': 8,
'learn': 3,
'hoping': 3,
'sharing': 2,
'encourage': 1,
'aware': 2,
'bodies': 4,
'nas': 8,
'enjoys': 2,
'free': 15,
'compiled': 1,
'bucket': 2,
...}

```

3.5 Matrices TF_IDF pour chaque document

In [82]: `imp.retourner_calcul_matrice_idf()`

Document n° 1 -----

Matrice TF-IDF HighLight : -----

| | |
|---------|-----|
| (0, 12) | 1.0 |
| (1, 22) | 1.0 |
| (2, 17) | 1.0 |
| (3, 20) | 1.0 |
| (4, 25) | 1.0 |
| (5, 23) | 1.0 |
| (6, 36) | 1.0 |
| (7, 2) | 1.0 |

| | |
|----------|-----|
| (8, 12) | 1.0 |
| (9, 21) | 1.0 |
| (10, 40) | 1.0 |
| (11, 19) | 1.0 |
| (12, 27) | 1.0 |
| (13, 16) | 1.0 |
| (14, 37) | 1.0 |
| (15, 44) | 1.0 |
| (16, 42) | 1.0 |
| (17, 45) | 1.0 |
| (18, 5) | 1.0 |
| (19, 0) | 1.0 |
| (20, 14) | 1.0 |
| (21, 41) | 1.0 |
| (22, 10) | 1.0 |
| (23, 26) | 1.0 |
| (24, 29) | 1.0 |
| (25, 12) | 1.0 |
| (26, 24) | 1.0 |
| (27, 9) | 1.0 |
| (28, 38) | 1.0 |
| (29, 33) | 1.0 |
| (30, 43) | 1.0 |
| (31, 31) | 1.0 |
| (32, 3) | 1.0 |
| (33, 18) | 1.0 |
| (34, 13) | 1.0 |
| (35, 11) | 1.0 |
| (36, 1) | 1.0 |
| (37, 28) | 1.0 |
| (38, 12) | 1.0 |
| (39, 21) | 1.0 |
| (40, 39) | 1.0 |
| (41, 7) | 1.0 |
| (42, 35) | 1.0 |
| (43, 8) | 1.0 |
| (44, 15) | 1.0 |
| (45, 34) | 1.0 |
| (46, 4) | 1.0 |
| (47, 30) | 1.0 |
| (48, 6) | 1.0 |
| (49, 32) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 26) | 1.0 |
| (1, 177) | 1.0 |
| (2, 324) | 1.0 |
| (3, 176) | 1.0 |
| (4, 55) | 1.0 |

| | |
|------------|-----|
| (5, 316) | 1.0 |
| (6, 64) | 1.0 |
| (7, 169) | 1.0 |
| (8, 254) | 1.0 |
| (9, 325) | 1.0 |
| (10, 138) | 1.0 |
| (11, 358) | 1.0 |
| (12, 236) | 1.0 |
| (13, 242) | 1.0 |
| (14, 27) | 1.0 |
| (15, 340) | 1.0 |
| (16, 208) | 1.0 |
| (17, 152) | 1.0 |
| (18, 9) | 1.0 |
| (19, 142) | 1.0 |
| (20, 2) | 1.0 |
| (21, 357) | 1.0 |
| (22, 147) | 1.0 |
| (23, 214) | 1.0 |
| (24, 16) | 1.0 |
| : | : |
| (532, 164) | 1.0 |
| (533, 34) | 1.0 |
| (534, 277) | 1.0 |
| (535, 352) | 1.0 |
| (536, 333) | 1.0 |
| (537, 334) | 1.0 |
| (538, 213) | 1.0 |
| (539, 180) | 1.0 |
| (540, 80) | 1.0 |
| (541, 243) | 1.0 |
| (542, 18) | 1.0 |
| (543, 211) | 1.0 |
| (544, 18) | 1.0 |
| (545, 340) | 1.0 |
| (546, 223) | 1.0 |
| (547, 240) | 1.0 |
| (548, 286) | 1.0 |
| (549, 74) | 1.0 |
| (550, 352) | 1.0 |
| (551, 333) | 1.0 |
| (552, 337) | 1.0 |
| (553, 202) | 1.0 |
| (554, 21) | 1.0 |
| (555, 69) | 1.0 |
| (556, 121) | 1.0 |

Document nř 2 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 21) | 1.0 |
| (1, 29) | 1.0 |
| (2, 41) | 1.0 |
| (3, 46) | 1.0 |
| (4, 11) | 1.0 |
| (5, 43) | 1.0 |
| (6, 17) | 1.0 |
| (7, 3) | 1.0 |
| (8, 32) | 1.0 |
| (9, 2) | 1.0 |
| (10, 17) | 1.0 |
| (11, 21) | 1.0 |
| (12, 34) | 1.0 |
| (13, 6) | 1.0 |
| (14, 0) | 1.0 |
| (15, 52) | 1.0 |
| (16, 35) | 1.0 |
| (17, 16) | 1.0 |
| (18, 23) | 1.0 |
| (19, 7) | 1.0 |
| (20, 5) | 1.0 |
| (21, 17) | 1.0 |
| (22, 45) | 1.0 |
| (23, 21) | 1.0 |
| (24, 30) | 1.0 |
| : | : |
| (36, 33) | 1.0 |
| (37, 38) | 1.0 |
| (38, 4) | 1.0 |
| (39, 39) | 1.0 |
| (40, 48) | 1.0 |
| (41, 44) | 1.0 |
| (42, 13) | 1.0 |
| (43, 18) | 1.0 |
| (44, 21) | 1.0 |
| (45, 28) | 1.0 |
| (46, 48) | 1.0 |
| (47, 31) | 1.0 |
| (48, 27) | 1.0 |
| (49, 20) | 1.0 |
| (50, 49) | 1.0 |
| (51, 36) | 1.0 |
| (52, 19) | 1.0 |
| (53, 25) | 1.0 |
| (54, 24) | 1.0 |
| (55, 51) | 1.0 |
| (56, 26) | 1.0 |

| | |
|----------|-----|
| (57, 42) | 1.0 |
| (58, 22) | 1.0 |
| (59, 53) | 1.0 |
| (60, 14) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 27) | 1.0 |
| (1, 206) | 1.0 |
| (2, 254) | 1.0 |
| (3, 213) | 1.0 |
| (4, 186) | 1.0 |
| (5, 1) | 1.0 |
| (6, 87) | 1.0 |
| (7, 6) | 1.0 |
| (8, 180) | 1.0 |
| (9, 5) | 1.0 |
| (10, 228) | 1.0 |
| (11, 187) | 1.0 |
| (12, 7) | 1.0 |
| (13, 87) | 1.0 |
| (14, 6) | 1.0 |
| (15, 180) | 1.0 |
| (16, 5) | 1.0 |
| (17, 211) | 1.0 |
| (18, 128) | 1.0 |
| (19, 303) | 1.0 |
| (20, 154) | 1.0 |
| (21, 145) | 1.0 |
| (22, 117) | 1.0 |
| (23, 318) | 1.0 |
| (24, 322) | 1.0 |
| : | : |
| (529, 49) | 1.0 |
| (530, 256) | 1.0 |
| (531, 44) | 1.0 |
| (532, 161) | 1.0 |
| (533, 118) | 1.0 |
| (534, 96) | 1.0 |
| (535, 190) | 1.0 |
| (536, 127) | 1.0 |
| (537, 240) | 1.0 |
| (538, 23) | 1.0 |
| (539, 324) | 1.0 |
| (540, 252) | 1.0 |
| (541, 255) | 1.0 |
| (542, 331) | 1.0 |
| (543, 302) | 1.0 |
| (544, 243) | 1.0 |
| (545, 59) | 1.0 |

| | |
|------------|-----|
| (546, 163) | 1.0 |
| (547, 275) | 1.0 |
| (548, 198) | 1.0 |
| (549, 352) | 1.0 |
| (550, 78) | 1.0 |
| (551, 35) | 1.0 |
| (552, 51) | 1.0 |
| (553, 115) | 1.0 |

Document nř 3 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 12) | 1.0 |
| (1, 23) | 1.0 |
| (2, 37) | 1.0 |
| (3, 31) | 1.0 |
| (4, 10) | 1.0 |
| (5, 3) | 1.0 |
| (6, 15) | 1.0 |
| (7, 1) | 1.0 |
| (8, 12) | 1.0 |
| (9, 21) | 1.0 |
| (10, 33) | 1.0 |
| (11, 18) | 1.0 |
| (12, 35) | 1.0 |
| (13, 9) | 1.0 |
| (14, 4) | 1.0 |
| (15, 5) | 1.0 |
| (16, 12) | 1.0 |
| (17, 25) | 1.0 |
| (18, 2) | 1.0 |
| (19, 40) | 1.0 |
| (20, 26) | 1.0 |
| (21, 28) | 1.0 |
| (22, 0) | 1.0 |
| (23, 18) | 1.0 |
| (24, 6) | 1.0 |
| (25, 11) | 1.0 |
| (26, 32) | 1.0 |
| (27, 12) | 1.0 |
| (28, 24) | 1.0 |
| (29, 36) | 1.0 |
| (30, 5) | 1.0 |
| (31, 16) | 1.0 |
| (32, 34) | 1.0 |
| (33, 7) | 1.0 |
| (34, 20) | 1.0 |
| (35, 19) | 1.0 |
| (36, 17) | 1.0 |

| | |
|----------|-----|
| (37, 39) | 1.0 |
| (38, 12) | 1.0 |
| (39, 24) | 1.0 |
| (40, 27) | 1.0 |
| (41, 8) | 1.0 |
| (42, 34) | 1.0 |
| (43, 38) | 1.0 |
| (44, 12) | 1.0 |
| (45, 22) | 1.0 |
| (46, 29) | 1.0 |
| (47, 14) | 1.0 |
| (48, 13) | 1.0 |
| (49, 30) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 31) | 1.0 |
| (1, 302) | 1.0 |
| (2, 160) | 1.0 |
| (3, 118) | 1.0 |
| (4, 395) | 1.0 |
| (5, 289) | 1.0 |
| (6, 166) | 1.0 |
| (7, 444) | 1.0 |
| (8, 383) | 1.0 |
| (9, 202) | 1.0 |
| (10, 220) | 1.0 |
| (11, 298) | 1.0 |
| (12, 206) | 1.0 |
| (13, 427) | 1.0 |
| (14, 387) | 1.0 |
| (15, 439) | 1.0 |
| (16, 293) | 1.0 |
| (17, 380) | 1.0 |
| (18, 201) | 1.0 |
| (19, 74) | 1.0 |
| (20, 372) | 1.0 |
| (21, 232) | 1.0 |
| (22, 277) | 1.0 |
| (23, 8) | 1.0 |
| (24, 449) | 1.0 |
| : | : |
| (679, 55) | 1.0 |
| (680, 54) | 1.0 |
| (681, 266) | 1.0 |
| (682, 450) | 1.0 |
| (683, 99) | 1.0 |
| (684, 5) | 1.0 |
| (685, 421) | 1.0 |
| (686, 32) | 1.0 |

| | |
|------------|-----|
| (687, 266) | 1.0 |
| (688, 451) | 1.0 |
| (689, 276) | 1.0 |
| (690, 392) | 1.0 |
| (691, 404) | 1.0 |
| (692, 218) | 1.0 |
| (693, 109) | 1.0 |
| (694, 245) | 1.0 |
| (695, 5) | 1.0 |
| (696, 225) | 1.0 |
| (697, 329) | 1.0 |
| (698, 259) | 1.0 |
| (699, 210) | 1.0 |
| (700, 334) | 1.0 |
| (701, 205) | 1.0 |
| (702, 68) | 1.0 |
| (703, 222) | 1.0 |

Document nr 4 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 10) | 1.0 |
| (1, 18) | 1.0 |
| (2, 8) | 1.0 |
| (3, 28) | 1.0 |
| (4, 12) | 1.0 |
| (5, 30) | 1.0 |
| (6, 24) | 1.0 |
| (7, 39) | 1.0 |
| (8, 36) | 1.0 |
| (9, 37) | 1.0 |
| (10, 27) | 1.0 |
| (11, 32) | 1.0 |
| (12, 7) | 1.0 |
| (13, 14) | 1.0 |
| (14, 11) | 1.0 |
| (15, 34) | 1.0 |
| (16, 19) | 1.0 |
| (17, 25) | 1.0 |
| (18, 23) | 1.0 |
| (19, 20) | 1.0 |
| (20, 10) | 1.0 |
| (21, 18) | 1.0 |
| (22, 38) | 1.0 |
| (23, 12) | 1.0 |
| (24, 3) | 1.0 |
| (25, 33) | 1.0 |
| (26, 2) | 1.0 |
| (27, 4) | 1.0 |

| | |
|----------|-----|
| (28, 5) | 1.0 |
| (29, 16) | 1.0 |
| (30, 6) | 1.0 |
| (31, 40) | 1.0 |
| (32, 13) | 1.0 |
| (33, 29) | 1.0 |
| (34, 10) | 1.0 |
| (35, 17) | 1.0 |
| (36, 35) | 1.0 |
| (37, 0) | 1.0 |
| (38, 31) | 1.0 |
| (39, 22) | 1.0 |
| (40, 21) | 1.0 |
| (41, 15) | 1.0 |
| (42, 1) | 1.0 |
| (43, 26) | 1.0 |
| (44, 9) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 16) | 1.0 |
| (1, 133) | 1.0 |
| (2, 76) | 1.0 |
| (3, 126) | 1.0 |
| (4, 17) | 1.0 |
| (5, 170) | 1.0 |
| (6, 39) | 1.0 |
| (7, 27) | 1.0 |
| (8, 18) | 1.0 |
| (9, 83) | 1.0 |
| (10, 166) | 1.0 |
| (11, 193) | 1.0 |
| (12, 45) | 1.0 |
| (13, 35) | 1.0 |
| (14, 172) | 1.0 |
| (15, 90) | 1.0 |
| (16, 136) | 1.0 |
| (17, 71) | 1.0 |
| (18, 177) | 1.0 |
| (19, 90) | 1.0 |
| (20, 186) | 1.0 |
| (21, 164) | 1.0 |
| (22, 231) | 1.0 |
| (23, 221) | 1.0 |
| (24, 228) | 1.0 |
| : | : |
| (448, 230) | 1.0 |
| (449, 229) | 1.0 |
| (450, 45) | 1.0 |
| (451, 59) | 1.0 |

| | |
|------------|-----|
| (452, 213) | 1.0 |
| (453, 189) | 1.0 |
| (454, 49) | 1.0 |
| (455, 54) | 1.0 |
| (456, 130) | 1.0 |
| (457, 214) | 1.0 |
| (458, 64) | 1.0 |
| (459, 89) | 1.0 |
| (460, 209) | 1.0 |
| (461, 167) | 1.0 |
| (462, 51) | 1.0 |
| (463, 0) | 1.0 |
| (464, 145) | 1.0 |
| (465, 28) | 1.0 |
| (466, 152) | 1.0 |
| (467, 90) | 1.0 |
| (468, 163) | 1.0 |
| (469, 109) | 1.0 |
| (470, 29) | 1.0 |
| (471, 198) | 1.0 |
| (472, 66) | 1.0 |

Document nr 5 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 26) | 1.0 |
| (1, 40) | 1.0 |
| (2, 68) | 1.0 |
| (3, 54) | 1.0 |
| (4, 57) | 1.0 |
| (5, 70) | 1.0 |
| (6, 33) | 1.0 |
| (7, 17) | 1.0 |
| (8, 21) | 1.0 |
| (9, 19) | 1.0 |
| (10, 5) | 1.0 |
| (11, 11) | 1.0 |
| (12, 53) | 1.0 |
| (13, 26) | 1.0 |
| (14, 43) | 1.0 |
| (15, 49) | 1.0 |
| (16, 35) | 1.0 |
| (17, 1) | 1.0 |
| (18, 58) | 1.0 |
| (19, 29) | 1.0 |
| (20, 60) | 1.0 |
| (21, 39) | 1.0 |
| (22, 72) | 1.0 |
| (23, 12) | 1.0 |

| | |
|----------|-----|
| (24, 26) | 1.0 |
| : | : |
| (64, 18) | 1.0 |
| (65, 26) | 1.0 |
| (66, 47) | 1.0 |
| (67, 20) | 1.0 |
| (68, 25) | 1.0 |
| (69, 30) | 1.0 |
| (70, 62) | 1.0 |
| (71, 23) | 1.0 |
| (72, 46) | 1.0 |
| (73, 71) | 1.0 |
| (74, 53) | 1.0 |
| (75, 6) | 1.0 |
| (76, 26) | 1.0 |
| (77, 38) | 1.0 |
| (78, 53) | 1.0 |
| (79, 13) | 1.0 |
| (80, 3) | 1.0 |
| (81, 70) | 1.0 |
| (82, 14) | 1.0 |
| (83, 10) | 1.0 |
| (84, 8) | 1.0 |
| (85, 15) | 1.0 |
| (86, 7) | 1.0 |
| (87, 24) | 1.0 |
| (88, 22) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 36) | 1.0 |
| (1, 185) | 1.0 |
| (2, 97) | 1.0 |
| (3, 89) | 1.0 |
| (4, 220) | 1.0 |
| (5, 292) | 1.0 |
| (6, 297) | 1.0 |
| (7, 314) | 1.0 |
| (8, 278) | 1.0 |
| (9, 71) | 1.0 |
| (10, 185) | 1.0 |
| (11, 65) | 1.0 |
| (12, 105) | 1.0 |
| (13, 96) | 1.0 |
| (14, 17) | 1.0 |
| (15, 60) | 1.0 |
| (16, 212) | 1.0 |
| (17, 176) | 1.0 |
| (18, 4) | 1.0 |
| (19, 253) | 1.0 |

| | |
|------------|-----|
| (20, 244) | 1.0 |
| (21, 212) | 1.0 |
| (22, 92) | 1.0 |
| (23, 144) | 1.0 |
| (24, 71) | 1.0 |
| : | : |
| (467, 43) | 1.0 |
| (468, 119) | 1.0 |
| (469, 120) | 1.0 |
| (470, 14) | 1.0 |
| (471, 149) | 1.0 |
| (472, 104) | 1.0 |
| (473, 277) | 1.0 |
| (474, 220) | 1.0 |
| (475, 296) | 1.0 |
| (476, 286) | 1.0 |
| (477, 279) | 1.0 |
| (478, 167) | 1.0 |
| (479, 193) | 1.0 |
| (480, 233) | 1.0 |
| (481, 55) | 1.0 |
| (482, 144) | 1.0 |
| (483, 141) | 1.0 |
| (484, 266) | 1.0 |
| (485, 143) | 1.0 |
| (486, 214) | 1.0 |
| (487, 220) | 1.0 |
| (488, 152) | 1.0 |
| (489, 151) | 1.0 |
| (490, 294) | 1.0 |
| (491, 139) | 1.0 |

Document nř 6 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 15) | 1.0 |
| (1, 22) | 1.0 |
| (2, 18) | 1.0 |
| (3, 21) | 1.0 |
| (4, 36) | 1.0 |
| (5, 25) | 1.0 |
| (6, 35) | 1.0 |
| (7, 2) | 1.0 |
| (8, 38) | 1.0 |
| (9, 15) | 1.0 |
| (10, 27) | 1.0 |
| (11, 8) | 1.0 |
| (12, 40) | 1.0 |
| (13, 49) | 1.0 |

| | |
|----------|-----|
| (14, 37) | 1.0 |
| (15, 4) | 1.0 |
| (16, 39) | 1.0 |
| (17, 15) | 1.0 |
| (18, 26) | 1.0 |
| (19, 13) | 1.0 |
| (20, 47) | 1.0 |
| (21, 45) | 1.0 |
| (22, 44) | 1.0 |
| (23, 42) | 1.0 |
| (24, 15) | 1.0 |
| : | : |
| (34, 1) | 1.0 |
| (35, 51) | 1.0 |
| (36, 30) | 1.0 |
| (37, 46) | 1.0 |
| (38, 13) | 1.0 |
| (39, 34) | 1.0 |
| (40, 16) | 1.0 |
| (41, 15) | 1.0 |
| (42, 24) | 1.0 |
| (43, 19) | 1.0 |
| (44, 20) | 1.0 |
| (45, 17) | 1.0 |
| (46, 7) | 1.0 |
| (47, 0) | 1.0 |
| (48, 11) | 1.0 |
| (49, 43) | 1.0 |
| (50, 6) | 1.0 |
| (51, 15) | 1.0 |
| (52, 28) | 1.0 |
| (53, 41) | 1.0 |
| (54, 32) | 1.0 |
| (55, 33) | 1.0 |
| (56, 3) | 1.0 |
| (57, 14) | 1.0 |
| (58, 50) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 37) | 1.0 |
| (1, 61) | 1.0 |
| (2, 118) | 1.0 |
| (3, 92) | 1.0 |
| (4, 63) | 1.0 |
| (5, 68) | 1.0 |
| (6, 243) | 1.0 |
| (7, 94) | 1.0 |
| (8, 252) | 1.0 |
| (9, 100) | 1.0 |

| | |
|------------|-----|
| (10, 44) | 1.0 |
| (11, 123) | 1.0 |
| (12, 240) | 1.0 |
| (13, 158) | 1.0 |
| (14, 40) | 1.0 |
| (15, 81) | 1.0 |
| (16, 196) | 1.0 |
| (17, 161) | 1.0 |
| (18, 9) | 1.0 |
| (19, 281) | 1.0 |
| (20, 164) | 1.0 |
| (21, 140) | 1.0 |
| (22, 253) | 1.0 |
| (23, 118) | 1.0 |
| (24, 233) | 1.0 |
| : | : |
| (498, 138) | 1.0 |
| (499, 184) | 1.0 |
| (500, 143) | 1.0 |
| (501, 234) | 1.0 |
| (502, 275) | 1.0 |
| (503, 175) | 1.0 |
| (504, 38) | 1.0 |
| (505, 19) | 1.0 |
| (506, 213) | 1.0 |
| (507, 133) | 1.0 |
| (508, 20) | 1.0 |
| (509, 211) | 1.0 |
| (510, 260) | 1.0 |
| (511, 138) | 1.0 |
| (512, 64) | 1.0 |
| (513, 148) | 1.0 |
| (514, 231) | 1.0 |
| (515, 139) | 1.0 |
| (516, 97) | 1.0 |
| (517, 47) | 1.0 |
| (518, 232) | 1.0 |
| (519, 113) | 1.0 |
| (520, 77) | 1.0 |
| (521, 217) | 1.0 |
| (522, 136) | 1.0 |

Document nř 7 -----

Matrice TF-IDF HighLight : -----

| | |
|---------|-----|
| (0, 14) | 1.0 |
| (1, 21) | 1.0 |
| (2, 31) | 1.0 |
| (3, 17) | 1.0 |

| | |
|----------|-----|
| (4, 39) | 1.0 |
| (5, 9) | 1.0 |
| (6, 10) | 1.0 |
| (7, 4) | 1.0 |
| (8, 6) | 1.0 |
| (9, 14) | 1.0 |
| (10, 24) | 1.0 |
| (11, 34) | 1.0 |
| (12, 27) | 1.0 |
| (13, 12) | 1.0 |
| (14, 11) | 1.0 |
| (15, 20) | 1.0 |
| (16, 13) | 1.0 |
| (17, 30) | 1.0 |
| (18, 14) | 1.0 |
| (19, 26) | 1.0 |
| (20, 15) | 1.0 |
| (21, 7) | 1.0 |
| (22, 18) | 1.0 |
| (23, 35) | 1.0 |
| (24, 16) | 1.0 |
| (25, 33) | 1.0 |
| (26, 14) | 1.0 |
| (27, 22) | 1.0 |
| (28, 41) | 1.0 |
| (29, 2) | 1.0 |
| (30, 36) | 1.0 |
| (31, 18) | 1.0 |
| (32, 32) | 1.0 |
| (33, 3) | 1.0 |
| (34, 14) | 1.0 |
| (35, 25) | 1.0 |
| (36, 0) | 1.0 |
| (37, 1) | 1.0 |
| (38, 19) | 1.0 |
| (39, 5) | 1.0 |
| (40, 28) | 1.0 |
| (41, 40) | 1.0 |
| (42, 31) | 1.0 |
| (43, 14) | 1.0 |
| (44, 23) | 1.0 |
| (45, 37) | 1.0 |
| (46, 29) | 1.0 |
| (47, 41) | 1.0 |
| (48, 8) | 1.0 |
| (49, 38) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|---------|-----|
| (0, 20) | 1.0 |
|---------|-----|

| | |
|------------|-----|
| (1, 187) | 1.0 |
| (2, 53) | 1.0 |
| (3, 129) | 1.0 |
| (4, 191) | 1.0 |
| (5, 163) | 1.0 |
| (6, 7) | 1.0 |
| (7, 74) | 1.0 |
| (8, 5) | 1.0 |
| (9, 242) | 1.0 |
| (10, 4) | 1.0 |
| (11, 201) | 1.0 |
| (12, 164) | 1.0 |
| (13, 9) | 1.0 |
| (14, 74) | 1.0 |
| (15, 6) | 1.0 |
| (16, 242) | 1.0 |
| (17, 4) | 1.0 |
| (18, 200) | 1.0 |
| (19, 145) | 1.0 |
| (20, 125) | 1.0 |
| (21, 267) | 1.0 |
| (22, 89) | 1.0 |
| (23, 94) | 1.0 |
| (24, 22) | 1.0 |
| : | : |
| (407, 112) | 1.0 |
| (408, 209) | 1.0 |
| (409, 280) | 1.0 |
| (410, 165) | 1.0 |
| (411, 131) | 1.0 |
| (412, 185) | 1.0 |
| (413, 36) | 1.0 |
| (414, 32) | 1.0 |
| (415, 83) | 1.0 |
| (416, 254) | 1.0 |
| (417, 244) | 1.0 |
| (418, 176) | 1.0 |
| (419, 12) | 1.0 |
| (420, 257) | 1.0 |
| (421, 278) | 1.0 |
| (422, 165) | 1.0 |
| (423, 106) | 1.0 |
| (424, 291) | 1.0 |
| (425, 214) | 1.0 |
| (426, 97) | 1.0 |
| (427, 266) | 1.0 |
| (428, 64) | 1.0 |
| (429, 31) | 1.0 |

| | |
|------------|-----|
| (430, 246) | 1.0 |
| (431, 147) | 1.0 |

Document nř 8 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 18) | 1.0 |
| (1, 27) | 1.0 |
| (2, 3) | 1.0 |
| (3, 11) | 1.0 |
| (4, 26) | 1.0 |
| (5, 21) | 1.0 |
| (6, 16) | 1.0 |
| (7, 22) | 1.0 |
| (8, 12) | 1.0 |
| (9, 17) | 1.0 |
| (10, 36) | 1.0 |
| (11, 33) | 1.0 |
| (12, 6) | 1.0 |
| (13, 14) | 1.0 |
| (14, 18) | 1.0 |
| (15, 28) | 1.0 |
| (16, 25) | 1.0 |
| (17, 20) | 1.0 |
| (18, 7) | 1.0 |
| (19, 24) | 1.0 |
| (20, 1) | 1.0 |
| (21, 2) | 1.0 |
| (22, 23) | 1.0 |
| (23, 32) | 1.0 |
| (24, 19) | 1.0 |
| (25, 15) | 1.0 |
| (26, 30) | 1.0 |
| (27, 0) | 1.0 |
| (28, 2) | 1.0 |
| (29, 18) | 1.0 |
| (30, 29) | 1.0 |
| (31, 39) | 1.0 |
| (32, 35) | 1.0 |
| (33, 34) | 1.0 |
| (34, 37) | 1.0 |
| (35, 9) | 1.0 |
| (36, 4) | 1.0 |
| (37, 31) | 1.0 |
| (38, 13) | 1.0 |
| (39, 10) | 1.0 |
| (40, 5) | 1.0 |
| (41, 8) | 1.0 |
| (42, 38) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 28) | 1.0 |
| (1, 21) | 1.0 |
| (2, 97) | 1.0 |
| (3, 22) | 1.0 |
| (4, 67) | 1.0 |
| (5, 141) | 1.0 |
| (6, 98) | 1.0 |
| (7, 41) | 1.0 |
| (8, 31) | 1.0 |
| (9, 79) | 1.0 |
| (10, 123) | 1.0 |
| (11, 70) | 1.0 |
| (12, 217) | 1.0 |
| (13, 72) | 1.0 |
| (14, 63) | 1.0 |
| (15, 140) | 1.0 |
| (16, 231) | 1.0 |
| (17, 33) | 1.0 |
| (18, 11) | 1.0 |
| (19, 147) | 1.0 |
| (20, 203) | 1.0 |
| (21, 160) | 1.0 |
| (22, 52) | 1.0 |
| (23, 237) | 1.0 |
| (24, 79) | 1.0 |
| : | : |
| (462, 226) | 1.0 |
| (463, 210) | 1.0 |
| (464, 256) | 1.0 |
| (465, 90) | 1.0 |
| (466, 250) | 1.0 |
| (467, 85) | 1.0 |
| (468, 111) | 1.0 |
| (469, 220) | 1.0 |
| (470, 115) | 1.0 |
| (471, 102) | 1.0 |
| (472, 72) | 1.0 |
| (473, 208) | 1.0 |
| (474, 245) | 1.0 |
| (475, 135) | 1.0 |
| (476, 65) | 1.0 |
| (477, 254) | 1.0 |
| (478, 183) | 1.0 |
| (479, 179) | 1.0 |
| (480, 57) | 1.0 |
| (481, 158) | 1.0 |
| (482, 19) | 1.0 |

| | |
|------------|-----|
| (483, 260) | 1.0 |
| (484, 260) | 1.0 |
| (485, 260) | 1.0 |
| (486, 260) | 1.0 |

Document nř 9 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 11) | 1.0 |
| (1, 17) | 1.0 |
| (2, 10) | 1.0 |
| (3, 26) | 1.0 |
| (4, 23) | 1.0 |
| (5, 33) | 1.0 |
| (6, 32) | 1.0 |
| (7, 5) | 1.0 |
| (8, 13) | 1.0 |
| (9, 11) | 1.0 |
| (10, 19) | 1.0 |
| (11, 6) | 1.0 |
| (12, 9) | 1.0 |
| (13, 29) | 1.0 |
| (14, 4) | 1.0 |
| (15, 25) | 1.0 |
| (16, 24) | 1.0 |
| (17, 20) | 1.0 |
| (18, 14) | 1.0 |
| (19, 12) | 1.0 |
| (20, 0) | 1.0 |
| (21, 21) | 1.0 |
| (22, 11) | 1.0 |
| (23, 18) | 1.0 |
| (24, 31) | 1.0 |
| (25, 25) | 1.0 |
| (26, 2) | 1.0 |
| (27, 22) | 1.0 |
| (28, 15) | 1.0 |
| (29, 16) | 1.0 |
| (30, 23) | 1.0 |
| (31, 34) | 1.0 |
| (32, 30) | 1.0 |
| (33, 11) | 1.0 |
| (34, 19) | 1.0 |
| (35, 3) | 1.0 |
| (36, 8) | 1.0 |
| (37, 1) | 1.0 |
| (38, 28) | 1.0 |
| (39, 7) | 1.0 |
| (40, 27) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 27) | 1.0 |
| (1, 179) | 1.0 |
| (2, 162) | 1.0 |
| (3, 198) | 1.0 |
| (4, 79) | 1.0 |
| (5, 221) | 1.0 |
| (6, 128) | 1.0 |
| (7, 125) | 1.0 |
| (8, 183) | 1.0 |
| (9, 5) | 1.0 |
| (10, 267) | 1.0 |
| (11, 84) | 1.0 |
| (12, 234) | 1.0 |
| (13, 95) | 1.0 |
| (14, 32) | 1.0 |
| (15, 153) | 1.0 |
| (16, 278) | 1.0 |
| (17, 34) | 1.0 |
| (18, 76) | 1.0 |
| (19, 102) | 1.0 |
| (20, 148) | 1.0 |
| (21, 231) | 1.0 |
| (22, 287) | 1.0 |
| (23, 175) | 1.0 |
| (24, 20) | 1.0 |
| : | : |
| (456, 109) | 1.0 |
| (457, 18) | 1.0 |
| (458, 133) | 1.0 |
| (459, 15) | 1.0 |
| (460, 296) | 1.0 |
| (461, 270) | 1.0 |
| (462, 8) | 1.0 |
| (463, 312) | 1.0 |
| (464, 173) | 1.0 |
| (465, 185) | 1.0 |
| (466, 146) | 1.0 |
| (467, 23) | 1.0 |
| (468, 137) | 1.0 |
| (469, 45) | 1.0 |
| (470, 134) | 1.0 |
| (471, 216) | 1.0 |
| (472, 5) | 1.0 |
| (473, 267) | 1.0 |
| (474, 147) | 1.0 |
| (475, 240) | 1.0 |
| (476, 262) | 1.0 |

| | |
|------------|-----|
| (477, 40) | 1.0 |
| (478, 122) | 1.0 |
| (479, 99) | 1.0 |
| (480, 2) | 1.0 |

Document nř 10 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 5) | 1.0 |
| (1, 9) | 1.0 |
| (2, 11) | 1.0 |
| (3, 6) | 1.0 |
| (4, 8) | 1.0 |
| (5, 0) | 1.0 |
| (6, 3) | 1.0 |
| (7, 14) | 1.0 |
| (8, 4) | 1.0 |
| (9, 5) | 1.0 |
| (10, 10) | 1.0 |
| (11, 7) | 1.0 |
| (12, 15) | 1.0 |
| (13, 2) | 1.0 |
| (14, 16) | 1.0 |
| (15, 12) | 1.0 |
| (16, 13) | 1.0 |
| (17, 1) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 19) | 1.0 |
| (1, 111) | 1.0 |
| (2, 94) | 1.0 |
| (3, 146) | 1.0 |
| (4, 123) | 1.0 |
| (5, 31) | 1.0 |
| (6, 49) | 1.0 |
| (7, 102) | 1.0 |
| (8, 74) | 1.0 |
| (9, 167) | 1.0 |
| (10, 152) | 1.0 |
| (11, 133) | 1.0 |
| (12, 36) | 1.0 |
| (13, 62) | 1.0 |
| (14, 62) | 1.0 |
| (15, 130) | 1.0 |
| (16, 46) | 1.0 |
| (17, 40) | 1.0 |
| (18, 57) | 1.0 |
| (19, 117) | 1.0 |
| (20, 177) | 1.0 |
| (21, 27) | 1.0 |

| | |
|------------|-----|
| (22, 45) | 1.0 |
| (23, 93) | 1.0 |
| (24, 3) | 1.0 |
| : | : |
| (226, 10) | 1.0 |
| (227, 12) | 1.0 |
| (228, 181) | 1.0 |
| (229, 142) | 1.0 |
| (230, 125) | 1.0 |
| (231, 51) | 1.0 |
| (232, 39) | 1.0 |
| (233, 105) | 1.0 |
| (234, 52) | 1.0 |
| (235, 156) | 1.0 |
| (236, 67) | 1.0 |
| (237, 147) | 1.0 |
| (238, 24) | 1.0 |
| (239, 95) | 1.0 |
| (240, 11) | 1.0 |
| (241, 135) | 1.0 |
| (242, 5) | 1.0 |
| (243, 174) | 1.0 |
| (244, 80) | 1.0 |
| (245, 47) | 1.0 |
| (246, 60) | 1.0 |
| (247, 178) | 1.0 |
| (248, 162) | 1.0 |
| (249, 103) | 1.0 |
| (250, 17) | 1.0 |

Document nř 11 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 4) | 1.0 |
| (1, 9) | 1.0 |
| (2, 2) | 1.0 |
| (3, 13) | 1.0 |
| (4, 14) | 1.0 |
| (5, 11) | 1.0 |
| (6, 3) | 1.0 |
| (7, 7) | 1.0 |
| (8, 5) | 1.0 |
| (9, 8) | 1.0 |
| (10, 0) | 1.0 |
| (11, 12) | 1.0 |
| (12, 10) | 1.0 |
| (13, 1) | 1.0 |
| (14, 6) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 25) | 1.0 |
| (1, 202) | 1.0 |
| (2, 77) | 1.0 |
| (3, 200) | 1.0 |
| (4, 179) | 1.0 |
| (5, 5) | 1.0 |
| (6, 82) | 1.0 |
| (7, 2) | 1.0 |
| (8, 169) | 1.0 |
| (9, 1) | 1.0 |
| (10, 206) | 1.0 |
| (11, 178) | 1.0 |
| (12, 4) | 1.0 |
| (13, 82) | 1.0 |
| (14, 3) | 1.0 |
| (15, 169) | 1.0 |
| (16, 1) | 1.0 |
| (17, 199) | 1.0 |
| (18, 300) | 1.0 |
| (19, 286) | 1.0 |
| (20, 112) | 1.0 |
| (21, 118) | 1.0 |
| (22, 259) | 1.0 |
| (23, 193) | 1.0 |
| (24, 296) | 1.0 |
| : | : |
| (489, 234) | 1.0 |
| (490, 47) | 1.0 |
| (491, 195) | 1.0 |
| (492, 73) | 1.0 |
| (493, 250) | 1.0 |
| (494, 92) | 1.0 |
| (495, 114) | 1.0 |
| (496, 39) | 1.0 |
| (497, 251) | 1.0 |
| (498, 201) | 1.0 |
| (499, 283) | 1.0 |
| (500, 252) | 1.0 |
| (501, 289) | 1.0 |
| (502, 3) | 1.0 |
| (503, 169) | 1.0 |
| (504, 1) | 1.0 |
| (505, 74) | 1.0 |
| (506, 234) | 1.0 |
| (507, 299) | 1.0 |
| (508, 176) | 1.0 |
| (509, 130) | 1.0 |
| (510, 101) | 1.0 |

| | |
|------------|-----|
| (511, 133) | 1.0 |
| (512, 0) | 1.0 |
| (513, 304) | 1.0 |

Document nř 12 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 19) | 1.0 |
| (1, 31) | 1.0 |
| (2, 54) | 1.0 |
| (3, 51) | 1.0 |
| (4, 52) | 1.0 |
| (5, 24) | 1.0 |
| (6, 23) | 1.0 |
| (7, 16) | 1.0 |
| (8, 54) | 1.0 |
| (9, 51) | 1.0 |
| (10, 52) | 1.0 |
| (11, 4) | 1.0 |
| (12, 22) | 1.0 |
| (13, 38) | 1.0 |
| (14, 49) | 1.0 |
| (15, 6) | 1.0 |
| (16, 11) | 1.0 |
| (17, 19) | 1.0 |
| (18, 28) | 1.0 |
| (19, 10) | 1.0 |
| (20, 8) | 1.0 |
| (21, 18) | 1.0 |
| (22, 47) | 1.0 |
| (23, 29) | 1.0 |
| (24, 5) | 1.0 |
| : | : |
| (43, 19) | 1.0 |
| (44, 35) | 1.0 |
| (45, 43) | 1.0 |
| (46, 12) | 1.0 |
| (47, 7) | 1.0 |
| (48, 21) | 1.0 |
| (49, 0) | 1.0 |
| (50, 19) | 1.0 |
| (51, 32) | 1.0 |
| (52, 3) | 1.0 |
| (53, 40) | 1.0 |
| (54, 30) | 1.0 |
| (55, 55) | 1.0 |
| (56, 15) | 1.0 |
| (57, 42) | 1.0 |
| (58, 20) | 1.0 |

| | |
|----------|-----|
| (59, 19) | 1.0 |
| (60, 34) | 1.0 |
| (61, 14) | 1.0 |
| (62, 43) | 1.0 |
| (63, 53) | 1.0 |
| (64, 45) | 1.0 |
| (65, 36) | 1.0 |
| (66, 13) | 1.0 |
| (67, 46) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|-------------|-----|
| (0, 42) | 1.0 |
| (1, 471) | 1.0 |
| (2, 444) | 1.0 |
| (3, 466) | 1.0 |
| (4, 137) | 1.0 |
| (5, 432) | 1.0 |
| (6, 52) | 1.0 |
| (7, 298) | 1.0 |
| (8, 471) | 1.0 |
| (9, 444) | 1.0 |
| (10, 464) | 1.0 |
| (11, 127) | 1.0 |
| (12, 6) | 1.0 |
| (13, 161) | 1.0 |
| (14, 471) | 1.0 |
| (15, 444) | 1.0 |
| (16, 464) | 1.0 |
| (17, 149) | 1.0 |
| (18, 8) | 1.0 |
| (19, 299) | 1.0 |
| (20, 171) | 1.0 |
| (21, 247) | 1.0 |
| (22, 90) | 1.0 |
| (23, 435) | 1.0 |
| (24, 251) | 1.0 |
| : | : |
| (1123, 57) | 1.0 |
| (1124, 174) | 1.0 |
| (1125, 111) | 1.0 |
| (1126, 37) | 1.0 |
| (1127, 187) | 1.0 |
| (1128, 158) | 1.0 |
| (1129, 83) | 1.0 |
| (1130, 380) | 1.0 |
| (1131, 44) | 1.0 |
| (1132, 270) | 1.0 |
| (1133, 86) | 1.0 |
| (1134, 171) | 1.0 |

| | |
|-------------|-----|
| (1135, 471) | 1.0 |
| (1136, 444) | 1.0 |
| (1137, 464) | 1.0 |
| (1138, 223) | 1.0 |
| (1139, 198) | 1.0 |
| (1140, 297) | 1.0 |
| (1141, 225) | 1.0 |
| (1142, 423) | 1.0 |
| (1143, 426) | 1.0 |
| (1144, 380) | 1.0 |
| (1145, 215) | 1.0 |
| (1146, 390) | 1.0 |
| (1147, 467) | 1.0 |

Document nř 13 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 7) | 1.0 |
| (1, 15) | 1.0 |
| (2, 18) | 1.0 |
| (3, 3) | 1.0 |
| (4, 5) | 1.0 |
| (5, 11) | 1.0 |
| (6, 1) | 1.0 |
| (7, 12) | 1.0 |
| (8, 7) | 1.0 |
| (9, 13) | 1.0 |
| (10, 6) | 1.0 |
| (11, 10) | 1.0 |
| (12, 18) | 1.0 |
| (13, 4) | 1.0 |
| (14, 23) | 1.0 |
| (15, 0) | 1.0 |
| (16, 17) | 1.0 |
| (17, 16) | 1.0 |
| (18, 7) | 1.0 |
| (19, 14) | 1.0 |
| (20, 19) | 1.0 |
| (21, 20) | 1.0 |
| (22, 8) | 1.0 |
| (23, 9) | 1.0 |
| (24, 21) | 1.0 |
| (25, 22) | 1.0 |
| (26, 2) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 26) | 1.0 |
| (1, 195) | 1.0 |
| (2, 17) | 1.0 |
| (3, 202) | 1.0 |

| | |
|------------|-----|
| (4, 307) | 1.0 |
| (5, 2) | 1.0 |
| (6, 5) | 1.0 |
| (7, 238) | 1.0 |
| (8, 7) | 1.0 |
| (9, 75) | 1.0 |
| (10, 4) | 1.0 |
| (11, 214) | 1.0 |
| (12, 77) | 1.0 |
| (13, 220) | 1.0 |
| (14, 162) | 1.0 |
| (15, 271) | 1.0 |
| (16, 6) | 1.0 |
| (17, 149) | 1.0 |
| (18, 0) | 1.0 |
| (19, 336) | 1.0 |
| (20, 191) | 1.0 |
| (21, 77) | 1.0 |
| (22, 220) | 1.0 |
| (23, 149) | 1.0 |
| (24, 0) | 1.0 |
| : | : |
| (559, 101) | 1.0 |
| (560, 262) | 1.0 |
| (561, 196) | 1.0 |
| (562, 114) | 1.0 |
| (563, 275) | 1.0 |
| (564, 125) | 1.0 |
| (565, 299) | 1.0 |
| (566, 239) | 1.0 |
| (567, 305) | 1.0 |
| (568, 336) | 1.0 |
| (569, 74) | 1.0 |
| (570, 324) | 1.0 |
| (571, 271) | 1.0 |
| (572, 225) | 1.0 |
| (573, 141) | 1.0 |
| (574, 138) | 1.0 |
| (575, 42) | 1.0 |
| (576, 323) | 1.0 |
| (577, 58) | 1.0 |
| (578, 216) | 1.0 |
| (579, 256) | 1.0 |
| (580, 260) | 1.0 |
| (581, 271) | 1.0 |
| (582, 152) | 1.0 |
| (583, 84) | 1.0 |

Document nř 14 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 18) | 1.0 |
| (1, 23) | 1.0 |
| (2, 28) | 1.0 |
| (3, 13) | 1.0 |
| (4, 16) | 1.0 |
| (5, 6) | 1.0 |
| (6, 14) | 1.0 |
| (7, 30) | 1.0 |
| (8, 2) | 1.0 |
| (9, 10) | 1.0 |
| (10, 8) | 1.0 |
| (11, 19) | 1.0 |
| (12, 4) | 1.0 |
| (13, 31) | 1.0 |
| (14, 18) | 1.0 |
| (15, 25) | 1.0 |
| (16, 26) | 1.0 |
| (17, 5) | 1.0 |
| (18, 29) | 1.0 |
| (19, 27) | 1.0 |
| (20, 22) | 1.0 |
| (21, 18) | 1.0 |
| (22, 24) | 1.0 |
| (23, 3) | 1.0 |
| (24, 11) | 1.0 |
| (25, 7) | 1.0 |
| (26, 1) | 1.0 |
| (27, 15) | 1.0 |
| (28, 21) | 1.0 |
| (29, 12) | 1.0 |
| (30, 17) | 1.0 |
| (31, 9) | 1.0 |
| (32, 32) | 1.0 |
| (33, 20) | 1.0 |
| (34, 0) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 30) | 1.0 |
| (1, 145) | 1.0 |
| (2, 218) | 1.0 |
| (3, 150) | 1.0 |
| (4, 132) | 1.0 |
| (5, 12) | 1.0 |
| (6, 77) | 1.0 |
| (7, 8) | 1.0 |
| (8, 182) | 1.0 |
| (9, 7) | 1.0 |

| | |
|------------|-----|
| (10, 155) | 1.0 |
| (11, 133) | 1.0 |
| (12, 1) | 1.0 |
| (13, 77) | 1.0 |
| (14, 8) | 1.0 |
| (15, 182) | 1.0 |
| (16, 7) | 1.0 |
| (17, 134) | 1.0 |
| (18, 89) | 1.0 |
| (19, 164) | 1.0 |
| (20, 38) | 1.0 |
| (21, 26) | 1.0 |
| (22, 16) | 1.0 |
| (23, 131) | 1.0 |
| (24, 121) | 1.0 |
| : | : |
| (323, 24) | 1.0 |
| (324, 206) | 1.0 |
| (325, 15) | 1.0 |
| (326, 40) | 1.0 |
| (327, 59) | 1.0 |
| (328, 98) | 1.0 |
| (329, 0) | 1.0 |
| (330, 129) | 1.0 |
| (331, 36) | 1.0 |
| (332, 152) | 1.0 |
| (333, 169) | 1.0 |
| (334, 225) | 1.0 |
| (335, 180) | 1.0 |
| (336, 142) | 1.0 |
| (337, 215) | 1.0 |
| (338, 32) | 1.0 |
| (339, 124) | 1.0 |
| (340, 51) | 1.0 |
| (341, 159) | 1.0 |
| (342, 80) | 1.0 |
| (343, 158) | 1.0 |
| (344, 183) | 1.0 |
| (345, 154) | 1.0 |
| (346, 102) | 1.0 |
| (347, 160) | 1.0 |

Document nř 15 -----

Matrice TF-IDF HighLight : -----

| | |
|---------|-----|
| (0, 9) | 1.0 |
| (1, 12) | 1.0 |
| (2, 4) | 1.0 |
| (3, 23) | 1.0 |

| | |
|----------|-----|
| (4, 22) | 1.0 |
| (5, 10) | 1.0 |
| (6, 5) | 1.0 |
| (7, 8) | 1.0 |
| (8, 1) | 1.0 |
| (9, 16) | 1.0 |
| (10, 23) | 1.0 |
| (11, 25) | 1.0 |
| (12, 9) | 1.0 |
| (13, 13) | 1.0 |
| (14, 18) | 1.0 |
| (15, 19) | 1.0 |
| (16, 17) | 1.0 |
| (17, 20) | 1.0 |
| (18, 0) | 1.0 |
| (19, 2) | 1.0 |
| (20, 14) | 1.0 |
| (21, 9) | 1.0 |
| (22, 3) | 1.0 |
| (23, 24) | 1.0 |
| (24, 11) | 1.0 |
| (25, 6) | 1.0 |
| (26, 15) | 1.0 |
| (27, 21) | 1.0 |
| (28, 7) | 1.0 |
| (29, 25) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 21) | 1.0 |
| (1, 31) | 1.0 |
| (2, 24) | 1.0 |
| (3, 253) | 1.0 |
| (4, 242) | 1.0 |
| (5, 145) | 1.0 |
| (6, 30) | 1.0 |
| (7, 3) | 1.0 |
| (8, 91) | 1.0 |
| (9, 18) | 1.0 |
| (10, 191) | 1.0 |
| (11, 223) | 1.0 |
| (12, 192) | 1.0 |
| (13, 234) | 1.0 |
| (14, 26) | 1.0 |
| (15, 53) | 1.0 |
| (16, 46) | 1.0 |
| (17, 183) | 1.0 |
| (18, 238) | 1.0 |
| (19, 71) | 1.0 |
| (20, 175) | 1.0 |

| | |
|------------|-----|
| (21, 298) | 1.0 |
| (22, 288) | 1.0 |
| (23, 295) | 1.0 |
| (24, 131) | 1.0 |
| : | : |
| (437, 298) | 1.0 |
| (438, 288) | 1.0 |
| (439, 295) | 1.0 |
| (440, 6) | 1.0 |
| (441, 213) | 1.0 |
| (442, 189) | 1.0 |
| (443, 203) | 1.0 |
| (444, 119) | 1.0 |
| (445, 287) | 1.0 |
| (446, 17) | 1.0 |
| (447, 297) | 1.0 |
| (448, 173) | 1.0 |
| (449, 228) | 1.0 |
| (450, 273) | 1.0 |
| (451, 89) | 1.0 |
| (452, 263) | 1.0 |
| (453, 113) | 1.0 |
| (454, 267) | 1.0 |
| (455, 165) | 1.0 |
| (456, 119) | 1.0 |
| (457, 38) | 1.0 |
| (458, 276) | 1.0 |
| (459, 251) | 1.0 |
| (460, 188) | 1.0 |
| (461, 297) | 1.0 |

Document nř 16 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 6) | 1.0 |
| (1, 11) | 1.0 |
| (2, 5) | 1.0 |
| (3, 17) | 1.0 |
| (4, 7) | 1.0 |
| (5, 3) | 1.0 |
| (6, 4) | 1.0 |
| (7, 10) | 1.0 |
| (8, 20) | 1.0 |
| (9, 6) | 1.0 |
| (10, 13) | 1.0 |
| (11, 15) | 1.0 |
| (12, 14) | 1.0 |
| (13, 18) | 1.0 |
| (14, 14) | 1.0 |

| | |
|----------|-----|
| (15, 8) | 1.0 |
| (16, 23) | 1.0 |
| (17, 6) | 1.0 |
| (18, 12) | 1.0 |
| (19, 16) | 1.0 |
| (20, 19) | 1.0 |
| (21, 2) | 1.0 |
| (22, 22) | 1.0 |
| (23, 9) | 1.0 |
| (24, 20) | 1.0 |
| (25, 1) | 1.0 |
| (26, 0) | 1.0 |
| (27, 21) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 10) | 1.0 |
| (1, 54) | 1.0 |
| (2, 99) | 1.0 |
| (3, 57) | 1.0 |
| (4, 7) | 1.0 |
| (5, 39) | 1.0 |
| (6, 29) | 1.0 |
| (7, 23) | 1.0 |
| (8, 70) | 1.0 |
| (9, 63) | 1.0 |
| (10, 97) | 1.0 |
| (11, 109) | 1.0 |
| (12, 17) | 1.0 |
| (13, 22) | 1.0 |
| (14, 68) | 1.0 |
| (15, 24) | 1.0 |
| (16, 73) | 1.0 |
| (17, 92) | 1.0 |
| (18, 73) | 1.0 |
| (19, 44) | 1.0 |
| (20, 62) | 1.0 |
| (21, 86) | 1.0 |
| (22, 17) | 1.0 |
| (23, 93) | 1.0 |
| (24, 31) | 1.0 |
| : | : |
| (143, 40) | 1.0 |
| (144, 30) | 1.0 |
| (145, 67) | 1.0 |
| (146, 36) | 1.0 |
| (147, 99) | 1.0 |
| (148, 41) | 1.0 |
| (149, 88) | 1.0 |
| (150, 48) | 1.0 |

| | |
|------------|-----|
| (151, 95) | 1.0 |
| (152, 53) | 1.0 |
| (153, 20) | 1.0 |
| (154, 64) | 1.0 |
| (155, 71) | 1.0 |
| (156, 6) | 1.0 |
| (157, 21) | 1.0 |
| (158, 52) | 1.0 |
| (159, 53) | 1.0 |
| (160, 104) | 1.0 |
| (161, 12) | 1.0 |
| (162, 92) | 1.0 |
| (163, 12) | 1.0 |
| (164, 100) | 1.0 |
| (165, 32) | 1.0 |
| (166, 6) | 1.0 |
| (167, 106) | 1.0 |

Document nř 17 -----
Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 10) | 1.0 |
| (1, 15) | 1.0 |
| (2, 9) | 1.0 |
| (3, 25) | 1.0 |
| (4, 28) | 1.0 |
| (5, 21) | 1.0 |
| (6, 26) | 1.0 |
| (7, 6) | 1.0 |
| (8, 4) | 1.0 |
| (9, 10) | 1.0 |
| (10, 18) | 1.0 |
| (11, 17) | 1.0 |
| (12, 23) | 1.0 |
| (13, 7) | 1.0 |
| (14, 11) | 1.0 |
| (15, 22) | 1.0 |
| (16, 10) | 1.0 |
| (17, 16) | 1.0 |
| (18, 21) | 1.0 |
| (19, 8) | 1.0 |
| (20, 20) | 1.0 |
| (21, 13) | 1.0 |
| (22, 27) | 1.0 |
| (23, 2) | 1.0 |
| (24, 5) | 1.0 |
| (25, 3) | 1.0 |
| (26, 0) | 1.0 |
| (27, 10) | 1.0 |

| | |
|----------|-----|
| (28, 19) | 1.0 |
| (29, 14) | 1.0 |
| (30, 1) | 1.0 |
| (31, 24) | 1.0 |
| (32, 12) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 31) | 1.0 |
| (1, 200) | 1.0 |
| (2, 276) | 1.0 |
| (3, 193) | 1.0 |
| (4, 153) | 1.0 |
| (5, 9) | 1.0 |
| (6, 77) | 1.0 |
| (7, 6) | 1.0 |
| (8, 145) | 1.0 |
| (9, 4) | 1.0 |
| (10, 203) | 1.0 |
| (11, 154) | 1.0 |
| (12, 8) | 1.0 |
| (13, 77) | 1.0 |
| (14, 7) | 1.0 |
| (15, 145) | 1.0 |
| (16, 4) | 1.0 |
| (17, 155) | 1.0 |
| (18, 96) | 1.0 |
| (19, 44) | 1.0 |
| (20, 51) | 1.0 |
| (21, 56) | 1.0 |
| (22, 16) | 1.0 |
| (23, 106) | 1.0 |
| (24, 69) | 1.0 |
| : | : |
| (469, 278) | 1.0 |
| (470, 282) | 1.0 |
| (471, 216) | 1.0 |
| (472, 292) | 1.0 |
| (473, 278) | 1.0 |
| (474, 287) | 1.0 |
| (475, 171) | 1.0 |
| (476, 236) | 1.0 |
| (477, 76) | 1.0 |
| (478, 23) | 1.0 |
| (479, 211) | 1.0 |
| (480, 291) | 1.0 |
| (481, 191) | 1.0 |
| (482, 5) | 1.0 |
| (483, 22) | 1.0 |
| (484, 148) | 1.0 |

| | |
|------------|-----|
| (485, 88) | 1.0 |
| (486, 107) | 1.0 |
| (487, 214) | 1.0 |
| (488, 250) | 1.0 |
| (489, 264) | 1.0 |
| (490, 146) | 1.0 |
| (491, 137) | 1.0 |
| (492, 143) | 1.0 |
| (493, 265) | 1.0 |

Document nř 18 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 14) | 1.0 |
| (1, 21) | 1.0 |
| (2, 16) | 1.0 |
| (3, 15) | 1.0 |
| (4, 17) | 1.0 |
| (5, 42) | 1.0 |
| (6, 3) | 1.0 |
| (7, 14) | 1.0 |
| (8, 24) | 1.0 |
| (9, 17) | 1.0 |
| (10, 26) | 1.0 |
| (11, 32) | 1.0 |
| (12, 18) | 1.0 |
| (13, 1) | 1.0 |
| (14, 0) | 1.0 |
| (15, 14) | 1.0 |
| (16, 22) | 1.0 |
| (17, 31) | 1.0 |
| (18, 46) | 1.0 |
| (19, 20) | 1.0 |
| (20, 5) | 1.0 |
| (21, 2) | 1.0 |
| (22, 5) | 1.0 |
| (23, 37) | 1.0 |
| (24, 7) | 1.0 |
| : | : |
| (36, 11) | 1.0 |
| (37, 38) | 1.0 |
| (38, 28) | 1.0 |
| (39, 34) | 1.0 |
| (40, 44) | 1.0 |
| (41, 14) | 1.0 |
| (42, 25) | 1.0 |
| (43, 33) | 1.0 |
| (44, 9) | 1.0 |
| (45, 19) | 1.0 |

| | |
|----------|-----|
| (46, 10) | 1.0 |
| (47, 39) | 1.0 |
| (48, 6) | 1.0 |
| (49, 40) | 1.0 |
| (50, 32) | 1.0 |
| (51, 14) | 1.0 |
| (52, 24) | 1.0 |
| (53, 30) | 1.0 |
| (54, 45) | 1.0 |
| (55, 4) | 1.0 |
| (56, 36) | 1.0 |
| (57, 8) | 1.0 |
| (58, 19) | 1.0 |
| (59, 43) | 1.0 |
| (60, 43) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 15) | 1.0 |
| (1, 158) | 1.0 |
| (2, 251) | 1.0 |
| (3, 11) | 1.0 |
| (4, 159) | 1.0 |
| (5, 210) | 1.0 |
| (6, 145) | 1.0 |
| (7, 38) | 1.0 |
| (8, 20) | 1.0 |
| (9, 27) | 1.0 |
| (10, 198) | 1.0 |
| (11, 73) | 1.0 |
| (12, 220) | 1.0 |
| (13, 111) | 1.0 |
| (14, 245) | 1.0 |
| (15, 260) | 1.0 |
| (16, 132) | 1.0 |
| (17, 104) | 1.0 |
| (18, 147) | 1.0 |
| (19, 1) | 1.0 |
| (20, 99) | 1.0 |
| (21, 72) | 1.0 |
| (22, 53) | 1.0 |
| (23, 42) | 1.0 |
| (24, 20) | 1.0 |
| : | : |
| (507, 146) | 1.0 |
| (508, 198) | 1.0 |
| (509, 167) | 1.0 |
| (510, 38) | 1.0 |
| (511, 19) | 1.0 |
| (512, 219) | 1.0 |

| | |
|------------|-----|
| (513, 85) | 1.0 |
| (514, 12) | 1.0 |
| (515, 269) | 1.0 |
| (516, 122) | 1.0 |
| (517, 130) | 1.0 |
| (518, 27) | 1.0 |
| (519, 108) | 1.0 |
| (520, 62) | 1.0 |
| (521, 35) | 1.0 |
| (522, 4) | 1.0 |
| (523, 0) | 1.0 |
| (524, 55) | 1.0 |
| (525, 5) | 1.0 |
| (526, 34) | 1.0 |
| (527, 57) | 1.0 |
| (528, 260) | 1.0 |
| (529, 260) | 1.0 |
| (530, 260) | 1.0 |
| (531, 260) | 1.0 |

Document nř 19 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 20) | 1.0 |
| (1, 29) | 1.0 |
| (2, 8) | 1.0 |
| (3, 24) | 1.0 |
| (4, 0) | 1.0 |
| (5, 44) | 1.0 |
| (6, 14) | 1.0 |
| (7, 52) | 1.0 |
| (8, 11) | 1.0 |
| (9, 39) | 1.0 |
| (10, 34) | 1.0 |
| (11, 12) | 1.0 |
| (12, 53) | 1.0 |
| (13, 16) | 1.0 |
| (14, 54) | 1.0 |
| (15, 46) | 1.0 |
| (16, 21) | 1.0 |
| (17, 20) | 1.0 |
| (18, 33) | 1.0 |
| (19, 44) | 1.0 |
| (20, 2) | 1.0 |
| (21, 5) | 1.0 |
| (22, 53) | 1.0 |
| (23, 16) | 1.0 |
| (24, 10) | 1.0 |

: :

| | |
|----------|-----|
| (43, 40) | 1.0 |
| (44, 51) | 1.0 |
| (45, 18) | 1.0 |
| (46, 9) | 1.0 |
| (47, 48) | 1.0 |
| (48, 15) | 1.0 |
| (49, 7) | 1.0 |
| (50, 6) | 1.0 |
| (51, 20) | 1.0 |
| (52, 28) | 1.0 |
| (53, 49) | 1.0 |
| (54, 31) | 1.0 |
| (55, 36) | 1.0 |
| (56, 1) | 1.0 |
| (57, 47) | 1.0 |
| (58, 40) | 1.0 |
| (59, 35) | 1.0 |
| (60, 22) | 1.0 |
| (61, 17) | 1.0 |
| (62, 3) | 1.0 |
| (63, 55) | 1.0 |
| (64, 37) | 1.0 |
| (65, 50) | 1.0 |
| (66, 30) | 1.0 |
| (67, 42) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 24) | 1.0 |
| (1, 267) | 1.0 |
| (2, 73) | 1.0 |
| (3, 222) | 1.0 |
| (4, 66) | 1.0 |
| (5, 206) | 1.0 |
| (6, 211) | 1.0 |
| (7, 105) | 1.0 |
| (8, 314) | 1.0 |
| (9, 196) | 1.0 |
| (10, 42) | 1.0 |
| (11, 141) | 1.0 |
| (12, 150) | 1.0 |
| (13, 129) | 1.0 |
| (14, 31) | 1.0 |
| (15, 288) | 1.0 |
| (16, 72) | 1.0 |
| (17, 91) | 1.0 |
| (18, 244) | 1.0 |
| (19, 32) | 1.0 |
| (20, 64) | 1.0 |
| (21, 126) | 1.0 |

| | |
|------------|-----|
| (22, 3) | 1.0 |
| (23, 280) | 1.0 |
| (24, 150) | 1.0 |
| : | : |
| (500, 216) | 1.0 |
| (501, 14) | 1.0 |
| (502, 297) | 1.0 |
| (503, 92) | 1.0 |
| (504, 25) | 1.0 |
| (505, 304) | 1.0 |
| (506, 201) | 1.0 |
| (507, 160) | 1.0 |
| (508, 236) | 1.0 |
| (509, 33) | 1.0 |
| (510, 189) | 1.0 |
| (511, 172) | 1.0 |
| (512, 143) | 1.0 |
| (513, 172) | 1.0 |
| (514, 84) | 1.0 |
| (515, 278) | 1.0 |
| (516, 225) | 1.0 |
| (517, 176) | 1.0 |
| (518, 294) | 1.0 |
| (519, 152) | 1.0 |
| (520, 291) | 1.0 |
| (521, 159) | 1.0 |
| (522, 5) | 1.0 |
| (523, 119) | 1.0 |
| (524, 159) | 1.0 |

Document nř 20 -----

Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 10) | 1.0 |
| (1, 16) | 1.0 |
| (2, 29) | 1.0 |
| (3, 4) | 1.0 |
| (4, 25) | 1.0 |
| (5, 27) | 1.0 |
| (6, 11) | 1.0 |
| (7, 17) | 1.0 |
| (8, 14) | 1.0 |
| (9, 10) | 1.0 |
| (10, 19) | 1.0 |
| (11, 23) | 1.0 |
| (12, 20) | 1.0 |
| (13, 22) | 1.0 |
| (14, 24) | 1.0 |
| (15, 1) | 1.0 |

| | |
|----------|-----|
| (16, 0) | 1.0 |
| (17, 8) | 1.0 |
| (18, 10) | 1.0 |
| (19, 18) | 1.0 |
| (20, 13) | 1.0 |
| (21, 6) | 1.0 |
| (22, 12) | 1.0 |
| (23, 3) | 1.0 |
| (24, 2) | 1.0 |
| (25, 28) | 1.0 |
| (26, 5) | 1.0 |
| (27, 10) | 1.0 |
| (28, 15) | 1.0 |
| (29, 30) | 1.0 |
| (30, 26) | 1.0 |
| (31, 21) | 1.0 |
| (32, 7) | 1.0 |
| (33, 9) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 34) | 1.0 |
| (1, 131) | 1.0 |
| (2, 63) | 1.0 |
| (3, 274) | 1.0 |
| (4, 45) | 1.0 |
| (5, 189) | 1.0 |
| (6, 155) | 1.0 |
| (7, 532) | 1.0 |
| (8, 80) | 1.0 |
| (9, 25) | 1.0 |
| (10, 267) | 1.0 |
| (11, 381) | 1.0 |
| (12, 429) | 1.0 |
| (13, 292) | 1.0 |
| (14, 341) | 1.0 |
| (15, 442) | 1.0 |
| (16, 421) | 1.0 |
| (17, 370) | 1.0 |
| (18, 126) | 1.0 |
| (19, 304) | 1.0 |
| (20, 152) | 1.0 |
| (21, 248) | 1.0 |
| (22, 181) | 1.0 |
| (23, 241) | 1.0 |
| (24, 423) | 1.0 |
| : | : |
| (983, 151) | 1.0 |
| (984, 298) | 1.0 |
| (985, 256) | 1.0 |

| | |
|-------------|-----|
| (986, 418) | 1.0 |
| (987, 380) | 1.0 |
| (988, 545) | 1.0 |
| (989, 26) | 1.0 |
| (990, 396) | 1.0 |
| (991, 351) | 1.0 |
| (992, 367) | 1.0 |
| (993, 476) | 1.0 |
| (994, 377) | 1.0 |
| (995, 131) | 1.0 |
| (996, 104) | 1.0 |
| (997, 35) | 1.0 |
| (998, 382) | 1.0 |
| (999, 168) | 1.0 |
| (1000, 334) | 1.0 |
| (1001, 20) | 1.0 |
| (1002, 112) | 1.0 |
| (1003, 170) | 1.0 |
| (1004, 67) | 1.0 |
| (1005, 36) | 1.0 |
| (1006, 31) | 1.0 |
| (1007, 65) | 1.0 |

Document nř 21 -----
Matrice TF-IDF HighLight : -----

| | |
|----------|-----|
| (0, 15) | 1.0 |
| (1, 24) | 1.0 |
| (2, 4) | 1.0 |
| (3, 8) | 1.0 |
| (4, 1) | 1.0 |
| (5, 17) | 1.0 |
| (6, 5) | 1.0 |
| (7, 11) | 1.0 |
| (8, 0) | 1.0 |
| (9, 15) | 1.0 |
| (10, 22) | 1.0 |
| (11, 20) | 1.0 |
| (12, 10) | 1.0 |
| (13, 30) | 1.0 |
| (14, 3) | 1.0 |
| (15, 14) | 1.0 |
| (16, 28) | 1.0 |
| (17, 2) | 1.0 |
| (18, 21) | 1.0 |
| (19, 19) | 1.0 |
| (20, 15) | 1.0 |
| (21, 23) | 1.0 |
| (22, 18) | 1.0 |

| | |
|----------|-----|
| (23, 6) | 1.0 |
| (24, 29) | 1.0 |
| (25, 12) | 1.0 |
| (26, 13) | 1.0 |
| (27, 16) | 1.0 |
| (28, 15) | 1.0 |
| (29, 25) | 1.0 |
| (30, 27) | 1.0 |
| (31, 7) | 1.0 |
| (32, 20) | 1.0 |
| (33, 9) | 1.0 |
| (34, 26) | 1.0 |
| (35, 31) | 1.0 |

Matrice TF-IDF Article : -----

| | |
|------------|-----|
| (0, 37) | 1.0 |
| (1, 301) | 1.0 |
| (2, 294) | 1.0 |
| (3, 296) | 1.0 |
| (4, 262) | 1.0 |
| (5, 13) | 1.0 |
| (6, 128) | 1.0 |
| (7, 10) | 1.0 |
| (8, 375) | 1.0 |
| (9, 9) | 1.0 |
| (10, 308) | 1.0 |
| (11, 263) | 1.0 |
| (12, 11) | 1.0 |
| (13, 128) | 1.0 |
| (14, 10) | 1.0 |
| (15, 375) | 1.0 |
| (16, 9) | 1.0 |
| (17, 264) | 1.0 |
| (18, 90) | 1.0 |
| (19, 451) | 1.0 |
| (20, 103) | 1.0 |
| (21, 312) | 1.0 |
| (22, 39) | 1.0 |
| (23, 115) | 1.0 |
| (24, 371) | 1.0 |
| : | : |
| (782, 426) | 1.0 |
| (783, 50) | 1.0 |
| (784, 328) | 1.0 |
| (785, 77) | 1.0 |
| (786, 211) | 1.0 |
| (787, 70) | 1.0 |
| (788, 445) | 1.0 |
| (789, 315) | 1.0 |

| | |
|------------|-----|
| (790, 281) | 1.0 |
| (791, 76) | 1.0 |
| (792, 210) | 1.0 |
| (793, 408) | 1.0 |
| (794, 234) | 1.0 |
| (795, 377) | 1.0 |
| (796, 75) | 1.0 |
| (797, 443) | 1.0 |
| (798, 340) | 1.0 |
| (799, 366) | 1.0 |
| (800, 131) | 1.0 |
| (801, 61) | 1.0 |
| (802, 239) | 1.0 |
| (803, 276) | 1.0 |
| (804, 132) | 1.0 |
| (805, 255) | 1.0 |
| (806, 38) | 1.0 |

Document nř 22 -----

Matrice TF-IDF HighLight : -----

[[1]]

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 0) | 1.0 |
| (1, 544) | 1.0 |
| (2, 490) | 1.0 |
| (3, 2) | 1.0 |
| (4, 7) | 1.0 |
| (5, 6) | 1.0 |
| (6, 486) | 1.0 |
| (7, 7) | 1.0 |
| (8, 7) | 1.0 |
| (9, 60) | 1.0 |
| (10, 487) | 1.0 |
| (11, 494) | 1.0 |
| (12, 7) | 1.0 |
| (13, 7) | 1.0 |
| (14, 241) | 1.0 |
| (15, 489) | 1.0 |
| (16, 492) | 1.0 |
| (17, 7) | 1.0 |
| (18, 7) | 1.0 |
| (19, 431) | 1.0 |
| (20, 491) | 1.0 |
| (21, 488) | 1.0 |
| (22, 7) | 1.0 |
| (23, 7) | 1.0 |
| (24, 244) | 1.0 |

: :

| | |
|-------------|-----|
| (3445, 271) | 1.0 |
| (3446, 488) | 1.0 |
| (3447, 7) | 1.0 |
| (3448, 443) | 1.0 |
| (3449, 496) | 1.0 |
| (3450, 7) | 1.0 |
| (3451, 7) | 1.0 |
| (3452, 31) | 1.0 |
| (3453, 488) | 1.0 |
| (3454, 7) | 1.0 |
| (3455, 443) | 1.0 |
| (3456, 494) | 1.0 |
| (3457, 7) | 1.0 |
| (3458, 7) | 1.0 |
| (3459, 99) | 1.0 |
| (3460, 488) | 1.0 |
| (3461, 7) | 1.0 |
| (3462, 443) | 1.0 |
| (3463, 498) | 1.0 |
| (3464, 7) | 1.0 |
| (3465, 7) | 1.0 |
| (3466, 168) | 1.0 |
| (3467, 488) | 1.0 |
| (3468, 7) | 1.0 |
| (3469, 116) | 1.0 |

Document nř 23 -----

Matrice TF-IDF HighLight : -----

[[1]]

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 0) | 1.0 |
| (1, 88) | 1.0 |
| (2, 55) | 1.0 |
| (3, 2) | 1.0 |
| (4, 6) | 1.0 |
| (5, 5) | 1.0 |
| (6, 6) | 1.0 |
| (7, 6) | 1.0 |
| (8, 19) | 1.0 |
| (9, 53) | 1.0 |
| (10, 6) | 1.0 |
| (11, 6) | 1.0 |
| (12, 29) | 1.0 |
| (13, 54) | 1.0 |
| (14, 61) | 1.0 |
| (15, 6) | 1.0 |
| (16, 6) | 1.0 |
| (17, 24) | 1.0 |

| | |
|-----------|-----|
| (18, 56) | 1.0 |
| (19, 57) | 1.0 |
| (20, 6) | 1.0 |
| (21, 6) | 1.0 |
| (22, 27) | 1.0 |
| (23, 58) | 1.0 |
| (24, 55) | 1.0 |
| : | : |
| (220, 6) | 1.0 |
| (221, 6) | 1.0 |
| (222, 22) | 1.0 |
| (223, 5) | 1.0 |
| (224, 6) | 1.0 |
| (225, 6) | 1.0 |
| (226, 41) | 1.0 |
| (227, 5) | 1.0 |
| (228, 57) | 1.0 |
| (229, 6) | 1.0 |
| (230, 6) | 1.0 |
| (231, 11) | 1.0 |
| (232, 5) | 1.0 |
| (233, 55) | 1.0 |
| (234, 6) | 1.0 |
| (235, 6) | 1.0 |
| (236, 37) | 1.0 |
| (237, 59) | 1.0 |
| (238, 6) | 1.0 |
| (239, 6) | 1.0 |
| (240, 13) | 1.0 |
| (241, 61) | 1.0 |
| (242, 6) | 1.0 |
| (243, 6) | 1.0 |
| (244, 39) | 1.0 |

Document nř 24 -----

Matrice TF-IDF HighLight : -----

[[1]]

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 0) | 1.0 |
| (1, 533) | 1.0 |
| (2, 479) | 1.0 |
| (3, 2) | 1.0 |
| (4, 7) | 1.0 |
| (5, 6) | 1.0 |
| (6, 475) | 1.0 |
| (7, 7) | 1.0 |
| (8, 7) | 1.0 |
| (9, 63) | 1.0 |

| | |
|-------------|-----|
| (10, 476) | 1.0 |
| (11, 485) | 1.0 |
| (12, 7) | 1.0 |
| (13, 7) | 1.0 |
| (14, 280) | 1.0 |
| (15, 478) | 1.0 |
| (16, 479) | 1.0 |
| (17, 7) | 1.0 |
| (18, 7) | 1.0 |
| (19, 343) | 1.0 |
| (20, 480) | 1.0 |
| (21, 7) | 1.0 |
| (22, 7) | 1.0 |
| (23, 290) | 1.0 |
| (24, 482) | 1.0 |
| : | : |
| (3417, 115) | 1.0 |
| (3418, 477) | 1.0 |
| (3419, 7) | 1.0 |
| (3420, 464) | 1.0 |
| (3421, 483) | 1.0 |
| (3422, 7) | 1.0 |
| (3423, 7) | 1.0 |
| (3424, 52) | 1.0 |
| (3425, 477) | 1.0 |
| (3426, 7) | 1.0 |
| (3427, 464) | 1.0 |
| (3428, 487) | 1.0 |
| (3429, 7) | 1.0 |
| (3430, 7) | 1.0 |
| (3431, 79) | 1.0 |
| (3432, 477) | 1.0 |
| (3433, 7) | 1.0 |
| (3434, 464) | 1.0 |
| (3435, 481) | 1.0 |
| (3436, 7) | 1.0 |
| (3437, 7) | 1.0 |
| (3438, 162) | 1.0 |
| (3439, 477) | 1.0 |
| (3440, 7) | 1.0 |
| (3441, 118) | 1.0 |

Document nř 25 -----

Matrice TF-IDF HighLight : -----

[[1]]

Matrice TF-IDF Article : -----

| | |
|----------|-----|
| (0, 0) | 1.0 |
| (1, 100) | 1.0 |

| | |
|-----------|-----|
| (2, 66) | 1.0 |
| (3, 3) | 1.0 |
| (4, 8) | 1.0 |
| (5, 7) | 1.0 |
| (6, 8) | 1.0 |
| (7, 8) | 1.0 |
| (8, 30) | 1.0 |
| (9, 63) | 1.0 |
| (10, 72) | 1.0 |
| (11, 8) | 1.0 |
| (12, 8) | 1.0 |
| (13, 38) | 1.0 |
| (14, 65) | 1.0 |
| (15, 8) | 1.0 |
| (16, 8) | 1.0 |
| (17, 50) | 1.0 |
| (18, 67) | 1.0 |
| (19, 72) | 1.0 |
| (20, 8) | 1.0 |
| (21, 8) | 1.0 |
| (22, 55) | 1.0 |
| (23, 69) | 1.0 |
| (24, 8) | 1.0 |
| : | : |
| (262, 60) | 1.0 |
| (263, 4) | 1.0 |
| (264, 8) | 1.0 |
| (265, 8) | 1.0 |
| (266, 35) | 1.0 |
| (267, 76) | 1.0 |
| (268, 8) | 1.0 |
| (269, 8) | 1.0 |
| (270, 51) | 1.0 |
| (271, 7) | 1.0 |
| (272, 70) | 1.0 |
| (273, 8) | 1.0 |
| (274, 8) | 1.0 |
| (275, 31) | 1.0 |
| (276, 7) | 1.0 |
| (277, 70) | 1.0 |
| (278, 8) | 1.0 |
| (279, 8) | 1.0 |
| (280, 62) | 1.0 |
| (281, 7) | 1.0 |
| (282, 66) | 1.0 |
| (283, 8) | 1.0 |
| (284, 8) | 1.0 |
| (285, 23) | 1.0 |

```

(286, 1)          1.0
-----
Document nř 26 -----
Matrice TF-IDF HighLight : -----
[[1]]
Matrice TF-IDF Article : -----
(0, 0)           1.0
(1, 655)         1.0
(2, 601)         1.0
(3, 2)           1.0
(4, 7)           1.0
(5, 6)           1.0
(6, 597)         1.0
(7, 7)           1.0
(8, 7)           1.0
(9, 70)          1.0
(10, 598)        1.0
(11, 605)        1.0
(12, 7)          1.0
(13, 7)          1.0
(14, 407)        1.0
(15, 600)        1.0
(16, 605)        1.0
(17, 7)          1.0
(18, 7)          1.0
(19, 217)        1.0
(20, 602)        1.0
(21, 607)        1.0
(22, 7)          1.0
(23, 7)          1.0
(24, 162)        1.0
:               :
(4519, 7)        1.0
(4520, 582)      1.0
(4521, 603)      1.0
(4522, 7)        1.0
(4523, 7)        1.0
(4524, 273)      1.0
(4525, 777)      1.0
(4526, 599)      1.0
(4527, 7)        1.0
(4528, 582)      1.0
(4529, 7)        1.0
(4530, 7)        1.0
(4531, 97)       1.0
(4532, 779)      1.0
(4533, 599)      1.0
(4534, 7)        1.0

```

| | |
|-------------|-----|
| (4535, 582) | 1.0 |
| (4536, 607) | 1.0 |
| (4537, 7) | 1.0 |
| (4538, 7) | 1.0 |
| (4539, 291) | 1.0 |
| (4540, 781) | 1.0 |
| (4541, 599) | 1.0 |
| (4542, 7) | 1.0 |
| (4543, 156) | 1.0 |

Document nř 27 -----

Matrice TF-IDF HighLight : -----

[[1]]

Matrice TF-IDF Article : -----

| | |
|-----------|-----|
| (0, 0) | 1.0 |
| (1, 84) | 1.0 |
| (2, 51) | 1.0 |
| (3, 2) | 1.0 |
| (4, 6) | 1.0 |
| (5, 5) | 1.0 |
| (6, 6) | 1.0 |
| (7, 6) | 1.0 |
| (8, 19) | 1.0 |
| (9, 48) | 1.0 |
| (10, 61) | 1.0 |
| (11, 6) | 1.0 |
| (12, 6) | 1.0 |
| (13, 30) | 1.0 |
| (14, 50) | 1.0 |
| (15, 57) | 1.0 |
| (16, 6) | 1.0 |
| (17, 6) | 1.0 |
| (18, 44) | 1.0 |
| (19, 52) | 1.0 |
| (20, 51) | 1.0 |
| (21, 6) | 1.0 |
| (22, 6) | 1.0 |
| (23, 38) | 1.0 |
| (24, 54) | 1.0 |
| : | : |
| (216, 6) | 1.0 |
| (217, 19) | 1.0 |
| (218, 5) | 1.0 |
| (219, 59) | 1.0 |
| (220, 6) | 1.0 |
| (221, 6) | 1.0 |
| (222, 29) | 1.0 |
| (223, 5) | 1.0 |

| | |
|-----------|-----|
| (224, 59) | 1.0 |
| (225, 6) | 1.0 |
| (226, 6) | 1.0 |
| (227, 36) | 1.0 |
| (228, 5) | 1.0 |
| (229, 53) | 1.0 |
| (230, 6) | 1.0 |
| (231, 6) | 1.0 |
| (232, 21) | 1.0 |
| (233, 59) | 1.0 |
| (234, 6) | 1.0 |
| (235, 6) | 1.0 |
| (236, 20) | 1.0 |
| (237, 61) | 1.0 |
| (238, 6) | 1.0 |
| (239, 6) | 1.0 |
| (240, 37) | 1.0 |

3.6 Calcul de la matrice TF-IDF globale pour les documents de type HighLights

In [119]: `print(imp.retourner_calcul_matrice_idf_global_Highlight())`

| | |
|----------|-----------------|
| (0, 275) | 0.155020238177 |
| (0, 387) | 0.387550595442 |
| (0, 439) | 0.113628903429 |
| (0, 333) | 0.113628903429 |
| (0, 385) | 0.113628903429 |
| (0, 498) | 0.113628903429 |
| (0, 648) | 0.125193307735 |
| (0, 441) | 0.113628903429 |
| (0, 626) | 0.113628903429 |
| (0, 78) | 0.100968329154 |
| (0, 265) | 0.0745116075272 |
| (0, 302) | 0.0703421140724 |
| (0, 73) | 0.113628903429 |
| (0, 421) | 0.183971017502 |
| (0, 698) | 0.113628903429 |
| (0, 374) | 0.100968329154 |
| (0, 228) | 0.0603984542225 |
| (0, 519) | 0.113628903429 |
| (0, 309) | 0.113628903429 |
| (0, 627) | 0.113628903429 |
| (0, 755) | 0.113628903429 |
| (0, 27) | 0.140417179804 |
| (0, 753) | 0.19012349177 |
| (0, 749) | 0.113628903429 |

```

(0, 0)          0.302904987462
:
(20, 181)       0.121267051785
(20, 169)       0.121267051785
(20, 413)       0.13647291415
(20, 84)        0.13647291415
(20, 285)       0.13647291415
(20, 88)        0.13647291415
(20, 214)       0.13647291415
(20, 388)       0.13647291415
(20, 369)       0.272945828299
(20, 207)       0.13647291415
(20, 638)       0.13647291415
(20, 268)       0.13647291415
(20, 531)       0.13647291415
(20, 18)        0.13647291415
(20, 372)       0.13647291415
(20, 324)       0.13647291415
(20, 120)       0.13647291415
(20, 241)       0.13647291415
(20, 244)       0.13647291415
(20, 520)       0.13647291415
(20, 154)       0.13647291415
(20, 187)       0.13647291415
(20, 723)       0.13647291415
(20, 637)       0.13647291415
(20, 666)       0.13647291415

```

3.7 Calcul de la matrice TF_IDF globale pour les documents de type Article

```
In [122]: print(imp.retourner_calcul_matrice_idf_global_Article())
```

```

(0, 357)        0.00691244687711
(0, 565)        0.0534062035226
(0, 2254)       0.0251547904737
(0, 3976)       0.0251547904737
(0, 2239)       0.269585428207
(0, 2249)       0.0133708098473
(0, 701)        0.0670561033602
(0, 3790)       0.0503095809474
(0, 4042)       0.0171589335673
(0, 814)        0.100619161895
(0, 2175)       0.0251547904737
(0, 2680)       0.275932051533
(0, 2928)       0.0223520344534
(0, 1448)       0.0445051696021
(0, 4002)       0.0503095809474

```

| | |
|------------|-----------------|
| (0, 1802) | 0.0754643714211 |
| (0, 3982) | 0.073906346707 |
| (0, 6017) | 0.0407268948202 |
| (0, 2724) | 0.0503095809474 |
| (0, 3757) | 0.120112534971 |
| (0, 2809) | 0.109004730425 |
| (0, 360) | 0.0267416196945 |
| (0, 5885) | 0.249164807645 |
| (0, 5800) | 0.111760172267 |
| (0, 2504) | 0.0251547904737 |
| : | : |
| (26, 5389) | 0.0129320890968 |
| (26, 4888) | 0.0258641781936 |
| (26, 4282) | 0.0129320890968 |
| (26, 4842) | 0.0129320890968 |
| (26, 5432) | 0.0129320890968 |
| (26, 5428) | 0.0258641781936 |
| (26, 4551) | 0.0129320890968 |
| (26, 5525) | 0.0129320890968 |
| (26, 4979) | 0.014553663663 |
| (26, 5267) | 0.014553663663 |
| (26, 4482) | 0.014553663663 |
| (26, 4152) | 0.014553663663 |
| (26, 5426) | 0.014553663663 |
| (26, 5190) | 0.014553663663 |
| (26, 4108) | 0.014553663663 |
| (26, 5363) | 0.014553663663 |
| (26, 4368) | 0.014553663663 |
| (26, 4907) | 0.014553663663 |
| (26, 5170) | 0.014553663663 |
| (26, 4381) | 0.014553663663 |
| (26, 4932) | 0.014553663663 |
| (26, 5202) | 0.014553663663 |
| (26, 4771) | 0.014553663663 |
| (26, 4738) | 0.014553663663 |
| (26, 5211) | 0.014553663663 |

3.8 Test de l'enregistrement de documents

```
In [14]: imp.enregistrerListeFichiers()
```

3.9 Test du chargement du texte de documents

```
In [15]: imp.Charger_Document(2)
```

ARTICLE N° 2

['b', 'nhelen', 'pow', 'npublished', 'n09', '16', 'est', '29', 'may', '2013', 'nupdated', 'n10

HIGHLIGHTS Nř 2

['highlight', 'nbella', 'rodriguez', 'torres', 'diagnosed', 'stage', 'four', 'alveolar', 'nrha