

**sUNIVERSITE YAHIA FARES DE MEDEA
FACULTE DES SCIENCES**



**Département Mathématiques et Informatique
Master 2 Ingénierie systèmes et web technologie**

Mini Projet Recherche d'information

Nom : **AKKACHE**
Prénom : **Lounis**
Groupe : **02**

Année universitaire : 2022/2023

Partie A:

I. Création D'index:

- Pour la création d'index et la recherche j'ai utilisé spécialement une classe nommée **CRINDEXW**

```
Class of whoosh

Table={"File":[],QueryList[0]:[],QueryList[1]:[],QueryList[2]:[]}

class CRINDEXW:
> def __init__(self, IndexfileName, pathCorpus):--
> def Create_index_File(self):--
> def __OpenIndex(self):--
> def __CloseIndex(self):--
> def __TypeOfencoding(self, PathFile):--
> def AddDocumentToIndex(self, NameFile, PathFile):--
> def AddAllDocumentToIndex(self):--
> def Searcher(self, requete, TypeOfScoring):--
> def WriteInIndex(self):--

3) ✓ 0.3s
```

1. Fonction d'initialisation du Moteur de recherche Whoosh:

```
my_analyzer = StemmingAnalyzer(stemfn=stemmer('arabic').stemWord) | StopFilter(set(stopwords.words('arabic')))

self.schema = Schema(title=ID(unique=True, stored=True),
                      path=ID(stored=True),
                      content=TEXT(stored=True, analyzer=my_analyzer))
```

2. Les objets pour : créer, ouvrir et fermer un fichier indexé (Index) par Woosh:

A. Création de fichier indexé par whoosh:

```
def Create_index_File(self):
    if not os.path.exists(self.FileName):
        os.mkdir(self.FileName)
        index = create_in(self.FileName, self.schema)
```

B. L'ouverture d'index

```
def __OpenIndex(self):
    return open_dir(self.FileName)
```

C. Fermeture d'index

```
def __CloseIndex(self):
    self.writer.commit()
```

3. Fonction pour ajouter un fichier à l'index:

```
def AddDocumentToIndex(self, NameFile, PathFile):
    os.chdir("/home/lounis/Documents/M2/RI/TP2/")
    EncType = self.__TypeOfEncoding(PathFile)
    file=open(u""+PathFile+"", encoding=EncType).read()
    self.writer.add_document(title=NameFile,
                             path=PathFile,
                             # content=self.__Normalization_File(file))
                             content=file)
```

II. Recherche dans l'index:

- La préparation du requête j'ai mis au niveau de la fonction Searcher qui me permet de d'initialiser le moteur de recherche index.searcher, récupérer les X premier résultat d'une requête donnée:

```
def Searcher(self, requete, TypeOfScoring, X):
    index = self.__OpenIndex()
    qp = QueryParser("content", schema=index.schema,
                    group=OrGroup).parse(requete)
    # qp = QueryParser("content", schema=index.schema).parse(requete)
    WhooshResult={"Title":[], "content":[]}
    print("-----", qp, "-----")

    with index.searcher(weighting=TypeOfScoring) as s:
        results = s.search(qp, limit=X)
        print(results)
        for c in results:
            WhooshResult['Title'].append(c['title'])
            WhooshResult['content'].append(c['content'])

        found = results.scored_length()
        print(found)

        if results.has_exact_length():
            print("Scored", found, "of exactly", len(results), "documents")
        else:
            low = results.estimated_min_length()
            high = results.estimated_max_length()
            print("Scored", found, "of between", low, "and", high, "documents")

    return WhooshResult, found
```

Création d'une instance de la class CRINDEXW:

Create Schema and add all document to schema

```
objet=CRINDEXW("Index", "/home/lounis/Documents/M2/RI/TP2/Corpus/Arabic")
# objet=CRINDEXW("Index", "/home/lounis/Documents/M2/RI/TP2/Corpus/English")
objet.Create_index_File()
objet.WriteInIndex()
```

✓ 14.9s

Partie B: Evaluation (sur un corpus Arabe)

• Requêtes:

1. الأمن المعلوماتي
2. الفرضيات المعارضة
3. حوادث الطرق

• La recherche manuelle des documents pertinents relativement à 3 requêtes:

le tableau est très long j'ai sauvegardé le dans un Fichier Excel nommé **RechercheManuelle.xlsx**

A. الأمن المعلوماتي (I 34 documents pertinent)

D005 A.txt	D006 A.txt	D009 A.txt	D015 A.txt	D055 A.txt
D091 A.txt	D107 A.txt	D112 A.txt	D115 A.txt	D140 A.txt
D158 A.txt	D172 A.txt	D208 A.txt	D217 A.txt	D241 A.txt
D268 A.txt	D287 A.txt	D329 A.txt	D334 A.txt	D354 A.txt
D368 A.txt	D379 A.txt	D385 A.txt	D422 A.txt	D430 A.txt
D442 A.txt	D482 A.txt	D516 A.txt	D517 A.txt	D538 A.txt
D590 A.txt	D604 A.txt	D608 A.txt		

B. الفرضيات المعارضة (I 25 documents pertinent)

D004 A.txt	D012 A.txt	D054 A.txt	D077 A.txt	D100 A.txt
D114 A.txt	D141 A.txt	D209 A.txt	D327 A.txt	D328 A.txt
D335 A.txt	D343 A.txt	D352 A.txt	D391 A.txt	D424 A.txt
D473 A.txt	D520 A.txt	D530 A.txt	D555 A.txt	D595 A.txt
D601 A.txt	D606 A.txt	D607 A.txt	D612 A.txt	D613 A.txt

C. حوادث الطرق (I 133 documents pertinent)

D003 A.txt	D010 A.txt	D025 A.txt	D032 A.txt	D034 A.txt
D038 A.txt	D043 A.txt	D047 A.txt	D052 A.txt	D059 A.txt
D061 A.txt	D063 A.txt	D064 A.txt	D065 A.txt	D068 A.txt
D069 A.txt	D075 A.txt	D076 A.txt	D085 A.txt	D091 A.txt
D094 A.txt	D100 A.txt	D103 A.txt	D105 A.txt	D110 A.txt
D112 A.txt	D128 A.txt	D134 A.txt	D143 A.txt	D145 A.txt
D152 A.txt	D164 A.txt	D165 A.txt	D167 A.txt	D171 A.txt
D172 A.txt	D178 A.txt	D182 A.txt	D191 A.txt	D199 A.txt
D200 A.txt	D202 A.txt	D205 A.txt	D206 A.txt	D214 A.txt
D217 A.txt	D224 A.txt	D225 A.txt	D231 A.txt	D232 A.txt
D237 A.txt	D238 A.txt	D245 A.txt	D254 A.txt	D261 A.txt
D266 A.txt	D275 A.txt	D277 A.txt	D294 A.txt	D296 A.txt
D312 A.txt	D315 A.txt	D337 A.txt	D346 A.txt	D369 A.txt
D370 A.txt	D380 A.txt	D383 A.txt	D387 A.txt	D388 A.txt
D395 A.txt	D400 A.txt	D402 A.txt	D403 A.txt	D406 A.txt
D408 A.txt	D409 A.txt	D413 A.txt	D417 A.txt	D431 A.txt
D437 A.txt	D439 A.txt	D441 A.txt	D451 A.txt	D454 A.txt
D457 A.txt	D464 A.txt	D475 A.txt	D477 A.txt	D498 A.txt
D499 A.txt	D507 A.txt	D508 A.txt	D517 A.txt	D520 A.txt

D523 A.txt	D524 A.txt	D526 A.txt	D527 A.txt	D539 A.txt
D540 A.txt	D543 A.txt	D544 A.txt	D545 A.txt	D560 A.txt
D565 A.txt	D569 A.txt	D571 A.txt	D572 A.txt	D574 A.txt
D577 A.txt	D580 A.txt	D581 A.txt	D585 A.txt	D588 A.txt
D589 A.txt	D596 A.txt	D600 A.txt	D601 A.txt	D605 A.txt
D609 A.txt	D610 A.txt	D611 A.txt	D614 A.txt	D615 A.txt
D617 A.txt	D622 A.txt	D627 A.txt	D628 A.txt	D629 A.txt
D630 A.txt	D631 A.txt	D632 A.txt		

I. Phase 1 : Evaluation:

1. La recherche par whoosh de chaque requête en choisi d'un Modèle:

```
QueryList=["حوادث الطرق","الفرضيات المعارضة","الأمن المعلومات"]
NbrDoc={}
score= scoring.BM25F()
# scoring.BM25F()    # scoring.TF_IDF()    # scoring.Frequency()
```

score indique le Modèle qui doit être passé dans la méthode searcher pour faire la recherche de requête

Verification The relevance of the document

```
def Pertinence(TablePertinence,ResultWhoosh,query):
    for File in TablePertinence[query]["File"]:
        if File in ResultWhoosh:
            TablePertinence[query]["pertinence"].append(1)
        else:
            TablePertinence[query]["pertinence"].append(0)
    return TablePertinence
```

Search Query with method scoring

```
for query in QueryList:
    Result,NbrDoc[query]=objet.Searcher(query,score,20)
    ManuallyQueryFind=Pertinence(ManuallyQueryFind,Result,query)
```

```
----- (content:امن OR content:معلومات) -----
<Top 14 Results for Or([Term('content', 'امن'), Term('content', 'معلومات')]) runtime=0.0005900040000597073>
14
Scored 14 of between 14 and 14 documents
----- (content:فرق OR content:معارض) -----
<Top 20 Results for Or([Term('content', 'فرق'), Term('content', 'معارض')]) runtime=0.0009738589997141389>
20
Scored 20 of between 3 and 23 documents
----- (content:حوادث الطرق OR content:طرق) -----
<Top 20 Results for Or([Term('content', 'حوادث'), Term('content', 'طرق')]) runtime=0.0029863400000067406>
20
Scored 20 of between 5 and 124 documents
```

- Cette Fonction me permet de calculer le rappel et la précision de chaque requête

Function For Calculate Both of Precision AND Reminder

```
def CPR(Data,NbrDocPer,query):
    Reminder=[]
    Precision=[]
    cpt=0
    for index in Data.index:
        if(Data[query][index]==0):
            Reminder.append(0)
            Precision.append(0)
        else:
            cpt=cpt+1;
            Reminder.append(round((cpt/NbrDocPer),3))
            Precision.append(round((cpt/index),3))

    return Reminder,Precision,cpt+1
```

- Cette fonction Calcule $F_Measure$ de chaque requête

Function to Calculate Reminder And Precision AND FMesure For Each query

```
def RPPM(ManuallyQueryFind,QueryList):
    FMesure={QueryList[0]:0,QueryList[1]:0,QueryList[2]:0}
    print("F_Mesure")
    for query in QueryList:
        ManuallyQueryFind[query]['Reminder'],ManuallyQueryFind[query]['Precision'],cpt=CPR(ManuallyQueryFind,NbrDoc[query],query)
        F_Mes=0
        if(NbrDoc[query]!=0):
            P=round((cpt/len(ManuallyQueryFind[query]['File'])),3)
            R=round((cpt/NbrDoc[query]))
            F_Mes=round(((2*P*R)/(P+R)),5)
        FMesure[query]=F_Mes
        print(query,"=====> ",F_Mes)
    return FMesure
```

- Cette Fonction permet de calculer AVG Prec de chaque requête donnée:

Calculate AVGPrec

```
def Calcul_AVGPrec(Data,query,NbrDoc):
    AVG_Precision=0
    if (NbrDoc!=0):
        for index in range(len(Data[query]["File"])):
            if(Data[query]["pertinane"][index]!=0):
                AVG_Precision=AVG_Precision+Data[query]["Precision"][index]
            AVG_Precision=round((AVG_Precision/NbrDoc),3)

    print(query," -----> ",AVG_Precision)
    return AVG_Precision
```

- La précision interpolée aux 11 points de rappel se fait par la fonction ci-dessus:

Interpolated Precisions

```
def InterpolatedPrecision(Rappel,Data,query):
    PrecInter=[0,0,0,0,0,0,0,0,0,0,0]
    index=0
    pos=0

    for i in range(len(Rappel)):
        index=0
        while index <= len(Data[query]["Precision"])-1:
            if (Rappel[i] <= Data[query]["Reminder"][index]):
                pos=index
                break
            else :
                pos= index
                index=index+1
        Rapp=max(list(Data[query]["Precision"][pos:(len(Data[query]["Precision"]))]))
        PrecInter[i]=Rapp
    return PrecInter
```

- La trace du courbe se fait par la fonction :

Function to Plot Interpolated Precision

```
def Plot(Dict_Inter,QueryList):
    fig, ax = plt.subplots()
    for query in QueryList:
        ax.plot(Dict_Inter["Rappel"]+[1.0,1.2], Dict_Inter["PreInt "+query]+[0,0],"o-",label=(query[:-1]))
        ax.legend()

    ax.set_xlabel("Reminder")
    ax.set_ylabel("Precision")
    plt.title("Plot Of Interpolated Precisions")
    plt.show()
```

Les Résultat D'Évaluation:

1. Model Frequency:

A. الأمن المعلوماتي (I 34 documents pertinent)

“./DFQuery/DFQ1Frequency.xlsx”

	File	pertinane	Reminder	Precision
1	D005 A.txt	0	0	0
2	D006 A.txt	1	0.03	0.50
3	D009 A.txt	0	0	0
4	D015 A.txt	1	0.06	0.50
5	D055 A.txt	0	0	0
6	D091 A.txt	1	0.09	0.50
7	D107 A.txt	0	0	0
8	D112 A.txt	0	0	0
9	D115 A.txt	0	0	0
10	D140 A.txt	0	0	0
11	D158 A.txt	0	0	0
12	D172 A.txt	1	0.12	0.33
13	D208 A.txt	0	0	0
14	D217 A.txt	0	0	0
15	D241 A.txt	0	0	0
16	D268 A.txt	1	0.15	0.31
17	D287 A.txt	0	0	0
18	D329 A.txt	0	0	0

19	D334 A.txt	0	0	0
20	D354 A.txt	0	0	0
21	D368 A.txt	0	0	0
22	D379 A.txt	0	0	0
23	D385 A.txt	1	0.18	0.26
24	D422 A.txt	1	0.21	0.29
25	D430 A.txt	1	0.24	0.32
26	D442 A.txt	1	0.27	0.35
27	D482 A.txt	1	0.29	0.37
28	D516 A.txt	1	0.32	0.39
29	D517 A.txt	0	0	0
30	D538 A.txt	1	0.35	0.40
31	D590 A.txt	0	0	0
32	D604 A.txt	1	0.38	0.41
33	D608 A.txt	0	0	0
34	D611 A.txt	1	0.41	0.41

B. الفرضيات المعارضة (I 25 documents pertinent)

“./DFQuery/DFQ2Frequency.xlsx”

	File	pertinane	Reminder	Precision
1	D004 A.txt	1	0.04	1
2	D012 A.txt	1	0.08	1
3	D054 A.txt	1	0.12	1
4	D077 A.txt	1	0.16	1
5	D100 A.txt	1	0.20	1
6	D114 A.txt	1	0.24	1
7	D141 A.txt	1	0.28	1
8	D209 A.txt	1	0.32	1
9	D327 A.txt	1	0.36	1
10	D328 A.txt	1	0.40	1
11	D335 A.txt	1	0.44	1
12	D343 A.txt	1	0.48	1
13	D352 A.txt	1	0.52	1
14	D391 A.txt	0	0	0
15	D424 A.txt	1	0.56	0.93
16	D473 A.txt	1	0.60	0.94
17	D520 A.txt	1	0.64	0.94
18	D530 A.txt	1	0.68	0.94

19	D555 A.txt	0	0	0
20	D595 A.txt	0	0	0
21	D601 A.txt	0	0	0
22	D606 A.txt	1	0.72	0.82
23	D607 A.txt	1	0.76	0.83
24	D612 A.txt	0	0	0
25	D613 A.txt	1	0.80	0.80

C. حوادث الطرق (I 133 documents pertinent) :

Le tableau est très long tu trouves dans répertoire

“./DFQuery/DFQ3Frequency.xlsx” après l'exécution du code

- *F_Mesure de chaque requête:*

Calculate Reminder And Precision AND FMeasure

```
C_RPFM(ManuallyQueryFind,QueryList)
```

✓ 0.8s

F_Mesure

0.441	<=====	الأمن المعلوماتي
0.84	<=====	الفرضيات المعارضة
0.158	<=====	حوادث الطرق

- *AVG de chaque requête avec le MAP:*

Calculate AVG_Precision For each query

```
AVGQueryList={}
MAP=0
print("AVGPrec")
for query in QueryList:
    AVGQueryList[query]=Calcul_AVGPrec(ManuallyQueryFind,query,NbrDoc[query])
    MAP=MAP+AVGQueryList[query]
MAP=MAP/len(QueryList)
print("MAP = ",MAP)
```

✓ 0.6s

AVGPrec

0.382	<-----	الأمن المعلوماتي
0.96	<-----	الفرضيات المعارضة
0.154	<-----	حوادث الطرق

MAP = 0.49866666666666665

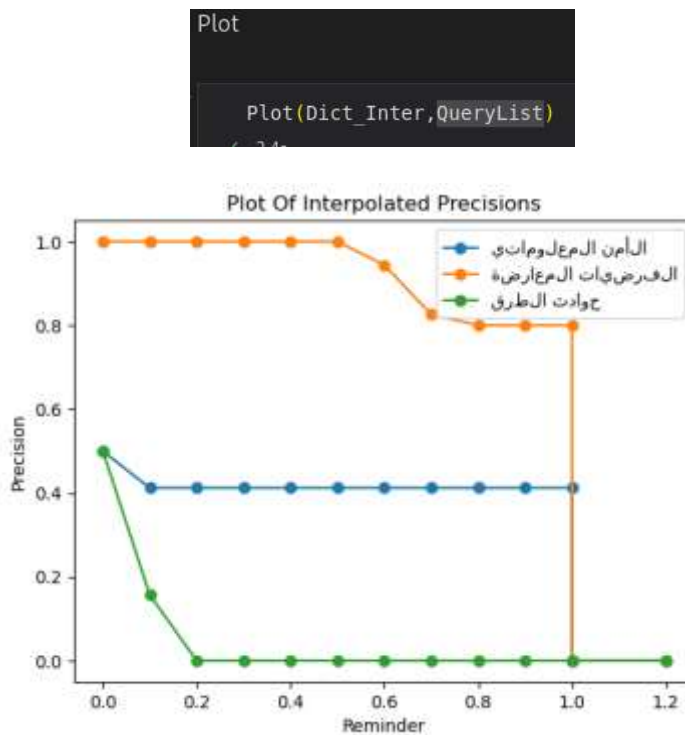
- *Précision interpolée*

Calculate Interpolated For Each Query

```
Rappel=[0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0]
Dict_Inter={"Rappel":Rappel}
for query in QueryList:
    Dict_Inter["PreIr"+query]=InterpoledPrecision(Rappel,ManuallyQueryFind,query)
```

Rappel	PreInt الأمن المعلوماتي	PreInt الفرضيات المعارضة	PreInt حوادث الطرق
0	0.50	1	0.50
0.10	0.41	1	0.16
0.20	0.41	1	0
0.30	0.41	1	0
0.40	0.41	1	0
0.50	0.41	1	0
0.60	0.41	0.94	0
0.70	0.41	0.83	0
0.80	0.41	0.80	0
0.90	0.41	0.80	0
1	0.41	0.80	0

- Traçage de courbe



Calcul précision au P5,P10,P15,P20

```
P5=Prec_Pos(ManuallyQueryFind,5)
P10=Prec_Pos(ManuallyQueryFind,10)
P15=Prec_Pos(ManuallyQueryFind,15)
P20=Prec_Pos(ManuallyQueryFind,20)

print("P5",P5)
print("P10",P10)
print("P15",P15)
print("P20",P20)
```

P5 {'الأمن المعلوماتي': [0.2], 'الفرضيات المعارضة': [1.0], 'حوادث الطرق': [0.1]}
P10 {'الأمن المعلوماتي': [0.15], 'الفرضيات المعارضة': [1.0], 'حوادث الطرق': [0.083]}
P15 {'الأمن المعلوماتي': [0.122], 'الفرضيات المعارضة': [0.929], 'حوادث الطرق': [0.056]}
P20 {'الأمن المعلوماتي': [0.107], 'الفرضيات المعارضة': [0.838], 'حوادث الطرق': [0.05]}

2. Model BM25F()

- A. الأمن المعلوماتي (I 34 documents pertinent)
Tu trouves le tableau dans “./DFQuery/DFQ1BM25F.xlsx”
- B. الفرضيات المعارضة (I 25 documents pertinent)
Tu trouves le tableau dans “./DFQuery/DFQ2BM25F.xlsx”
- C. حوادث الطرق (I 133 documents pertinent) :
Tu trouves le tableau dans “./DFQuery/DFQ3BM25F.xlsx”

Même Étapes Avec Frequency () juste changement mode de scoring

- *Résultat de recherche*

```
----- (content:امن OR content:معلومات) -----  
<Top 14 Results for Or([Term('content', 'امن'), Term('content', 'معلومات')]) runtime=8.800788596003477521>  
14  
Scored 14 of between 14 and 14 documents  
----- (content:فرض OR content:معارضة) -----  
<Top 20 Results for Or([Term('content', 'فرض'), Term('content', 'معارضة')]) runtime=8.8008511039995937608>  
20  
Scored 20 of between 3 and 23 documents  
----- (content:حوادث OR content:طرق) -----  
<Top 20 Results for Or([Term('content', 'حوادث'), Term('content', 'طرق')]) runtime=8.8020796670028878688>  
20  
Scored 20 of between 5 and 124 documents
```

- *F_Mesure:*

```
✓ C_RPFM(ManuallyQueryFind,QueryList)  
  
F_Mesure  
0.441 <===== الأمن المعلوماتي  
0.84 <===== الفرضيات المعارضة  
0.158 <===== حوادث الطرق
```

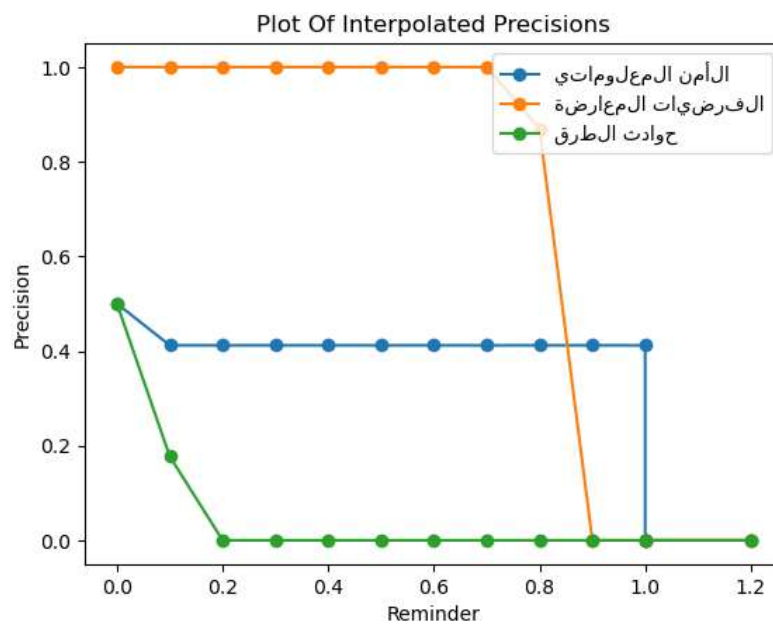
- *AVG et MAP:*

```
✓ AVGQueryList={} ...  
  
AVGPrec  
0.382 <----- الأمن المعلوماتي  
0.987 <----- الفرضيات المعارضة  
0.205 <----- حوادث الطرق  
MAP = 0.5246666666666667
```

- *Précision interpolée*

Rappel	الأمن المعلوماتي PreInt	الفرضيات المعارضة PreInt	حوادث الطرق PreInt
0	0.50	1	0.50
0.10	0.41	1	0.18
0.20	0.41	1	0
0.30	0.41	1	0
0.40	0.41	1	0
0.50	0.41	1	0
0.60	0.41	1	0
0.70	0.41	1	0
0.80	0.41	0.87	0
0.90	0.41	0	0
1	0.41	0	0

- *Courbe:*



Calcul précision au P5,10, P15, P20

```

P5=Prec_Pos(ManuallyQueryFind,5)
P10=Prec_Pos(ManuallyQueryFind,10)
P15=Prec_Pos(ManuallyQueryFind,15)
P20=Prec_Pos(ManuallyQueryFind,20)

✓ 0.7%

print("P5",P5)
print("P10",P10)
print("P15",P15)
print("P20",P20)

✓ 0.1%

P5 {'[0.1]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.2]': 'الأمن المعلوماتي'}
P10 {'[0.117]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.15]': 'الأمن المعلوماتي'}
P15 {'[0.078]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.122]': 'الأمن المعلوماتي'}
P20 {'[0.083]': 'حوادث الطرق', '[0.9]': 'الفرضيات المعارضة', '[0.107]': 'الأمن المعلوماتي'}

```


Les Résultat de Reformulation Automatique:

1. Model Frequency():

J'ai Télécharger l'ArabicWordNet (AWN)

J'ai fait la recherche de tous les mots existants dans mes requêtes

Existing Words in All Queries

```

words={}
for query in QueryList:
    for word in query.split(" "):
        words[word]=[]
        #synset WordNet
words['الطرق']=['طَرِيقٌ","سَبِيلٌ","إِتِّحَاةٌ","جِهَةٌ","وَجْهَةٌ","مدخل","سَارِع","خُذُوتٌ","خُشُولٌ","حَدَثٌ","خَاِدِيَّةٌ","وَقْعٌ","جَرِيٌ","دَارٌ","حِصْلٌ"]
words['المعارضة']=['عَرَضٌ","مَقْوَِرٌ","عَرَضٌ","أُطْهَرٌ","قَدَمٌ","جَلَّتْ","عَرَضٌ"]
words['الفرصيات']=['وَجْهَةٌ","أَصَابٌ","قَرَصٌ","أَكْدٌ"]
words['المعلوما تي']=['مَعْلُومَةٌ","حَقِيقَةٌ"]
words['الآ من']=['كَقَلٌ","مَانَ","صَمِينٌ","أَكْدٌ","أَمِنْ"]

for key,value in words.items():
    words[key]=[stemmer("arabic").stemWord(word) for word in value]
QueryAWN=[
    ' '.join(words['الآ من']+words['المعلوما تي']),
    ' '.join(words['الفرصيات']+words['المعارضة']),
    ' '.join(words['حوادث']+words['الطرق'])]

```

A. الأمن المعلوماتي (134 documents pertinent)

tu trouve le tableaux dans “./DFQuery/DFRQ1Frequency.xlsx”

B. الفرضيات المعارضة (I 25 documents pertinent)

tu trouve le tableaux dans “./DFQuery/DFRQ2Frequency.xlsx”

C. حوادث الطرق (1133 documents pertinent) :

Tu trouves le tableaux dans “./DFQuery/DFRQ3Frequency.xlsx”La

- Recherche par whoosh

Searcher

```
i=0;
for query in QueryList:
    print(query)
    Result2,NbrDocAWN[QueryAWN[i]]=objet.Searcher(QueryAWN[i],score,20)
    ManuallyQueryFindAWN=Pertinence(ManuallyQueryFindAWN,Result2,query)
    i=i+1
```

[illegible]

- Table de précision et Rappel de chaque requête:

```
Calculate Reminder And Precision For Each query IN model ArabicWordNet

for query in QueryList:
    ManuallyQueryFindAWN[query]["Reminder"],ManuallyQueryFindAWN[query]["Precision"],cpt=CPR(ManuallyQueryFindAWN,NbrDoc[query],query)
```

- F_Mesure (Reformulation):

```
Calculate FMesure

FMes2=C_RPFM(ManuallyQueryFindAWN,QueryList)
```

```
F_Mesure
0.294 <===== الأمن المعلوماتي
0.16 <===== الفرضيات المعارضة
0.098 <===== حوادث الطرق
```

```
Calculate Average AND MAP

AVGQueryListAWN={}
MAP_AWN=0
for query in QueryList:
    AVGQueryListAWN[query]=Calcul_AVGPrec(ManuallyQueryFindAWN,query,NbrDoc[query])
    MAP_AWN=MAP_AWN+AVGQueryListAWN[query]
MAP_AWN=MAP_AWN/len(QueryList)
print("MAP_AWN = ",MAP_AWN)
```

```
AVG
0.211 <----- الأمن المعلوماتي
0.107 <----- الفرضيات المعارضة
0.045 <----- حوادث الطرق
MAP_AWN = 0.121
```

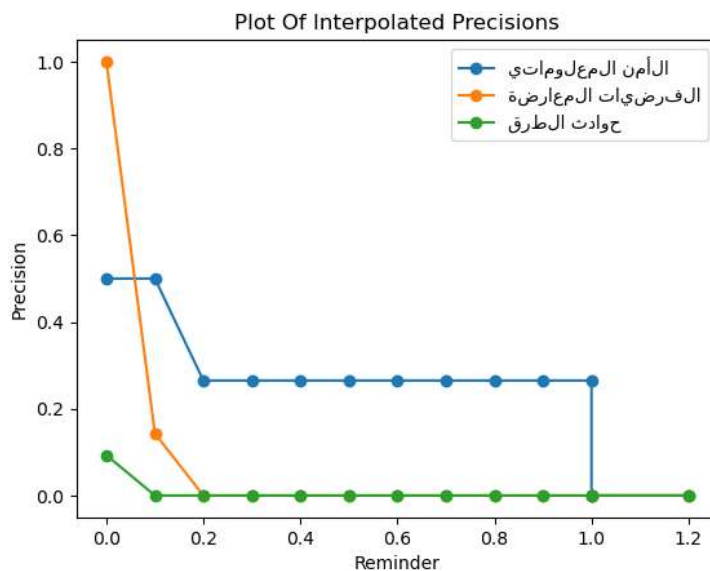
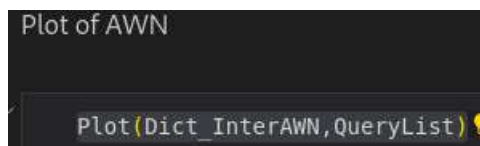
- *Précision interpolée*

```
Calculate Interpolated For Each Query

Rappel=[0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0]
Dict_InterAWN={"Rappel":Rappel}
for query in QueryList:
    Dict_InterAWN["PreInt "+query]=InterpolatedPrecision(Rappel,ManuallyQueryFindAWN,query)
```

Rappel	الأمن المعلوماتي PreInt	الفرضيات المعارضة PreInt	حوادث الطرق PreInt
0	0.50	1	0.09
0.10	0.50	0.14	0
0.20	0.27	0	0
0.30	0.27	0	0
0.40	0.27	0	0
0.50	0.27	0	0
0.60	0.27	0	0
0.70	0.27	0	0
0.80	0.27	0	0
0.90	0.27	0	0
1	0.27	0	0

- *Tracage de courbe*



```

P5AWN=Prec_Pos(ManuallyQueryFindAWN,5)
P10AWN=Prec_Pos(ManuallyQueryFindAWN,10)
P15AWN=Prec_Pos(ManuallyQueryFindAWN,15)
P20AWN=Prec_Pos(ManuallyQueryFindAWN,20)

✓ 0.2s

print("P5",P5AWN)
print("P10",P10AWN)
print("P15",P15AWN)
print("P20",P20AWN)

✓ 0.2s

P5 {'[0.0]': 'حوادث الطرق', '[0.4]': 'الفرضيات المعارضة', '[0.05]': 'الأمن المعلوماتي'}
P10 {'[0.0]': 'حوادث الطرق', '[0.2]': 'الفرضيات المعارضة', '[0.196]': 'الأمن المعلوماتي'}
P15 {'[0.0]': 'حوادث الطرق', '[0.133]': 'الفرضيات المعارضة', '[0.13]': 'الأمن المعلوماتي'}
P20 {'[0.003]': 'حوادث الطرق', '[0.1]': 'الفرضيات المعارضة', '[0.098]': 'الأمن المعلوماتي'}

```

2. Model BM25F():

- A. الأمين المعلوماتي (I 34 documents pertinent)
tu trouve le tableaux dans “./DFQuery/DFRQ1BM25F.xlsx”
- B. الفرضيات المعارضة (I 25 documents pertinent)
tu trouve le tableaux dans “./DFQuery/DFRQ2BM25F.xlsx”
- C. حوادث الطرق (I 133 documents pertinent):
tu trouve le tableaux dans “./DFQuery/DFRQ3BM25F.xlsx”

Même travail avec Frequency() et le Même synset

- recherche

```

الأمين
----- (content:أمين OR content:أكذ OR content:سمن OR content:ساز OR content:كفل OR content:حقيق OR content:معلوم) -----
>Top 20 Results for Or([Term('content','أمين'), Term('content','أكذ'), Term('content','سمن'), Term('content','ساز'), Term('content','كفل'), Term('content','حقيق'), Term('content','معلوم')]) runtime=0.004365246000816114>
20
Scored 20 of between 8 and 116 documents
الفرضيات المعارضة
----- (content:أكذ OR content:فريم OR content:أصاب OR content:وجه OR content:فريم OR content:جلب OR content:قدم OR content:الظهر OR content:اسور) -----
>Top 20 Results for Or([Term('content','أكذ'), Term('content','فريم'), Term('content','أصاب'), Term('content','وجه'), Term('content','فريم'), Term('content','جلب'), Term('content','قدم'), Term('content','الظهر'), Term('content','اسور')]) runtime=0.008850838999922639>
20
Scored 20 of between 4 and 274 documents
حوادث الطرق
----- (content:جمل OR content:دار OR content:جر OR content:وقع OR content:حادث OR content:جند OR content:جسول OR content:حدود OR content:طريق OR content:سبل OR content:ايح OR content:وجه OR content:مدخل OR content:إطارع) -----
>Top 20 Results for Or([Term('content','جمل'), Term('content','دار'), Term('content','جر'), Term('content','وقع'), Term('content','حادث'), Term('content','جند'), Term('content','جسول'), Term('content','حدود'), Term('content','طريق'), Term('content','سبل'), Term('content','ايح'), Term('content','وجه'), Term('content','مدخل'), Term('content','إطارع')]) runtime=0.0092476539999489522>
20
Scored 20 of between 1 and 443 documents

```

- F_Mesure:

```

Calculate FMeasure

✓ FMes2=C_RPFM(ManuallyQueryFindAWN,QueryList)

F_Mesure
0.235 <===== الأمن المعلوماتي
0.4 <===== الفرضيات المعارضة
0.038 <===== حوادث الطرق

```

- AVG and MAP

```

Calculate Average AND MAP

✓ AVGQueryListAWN={} ...

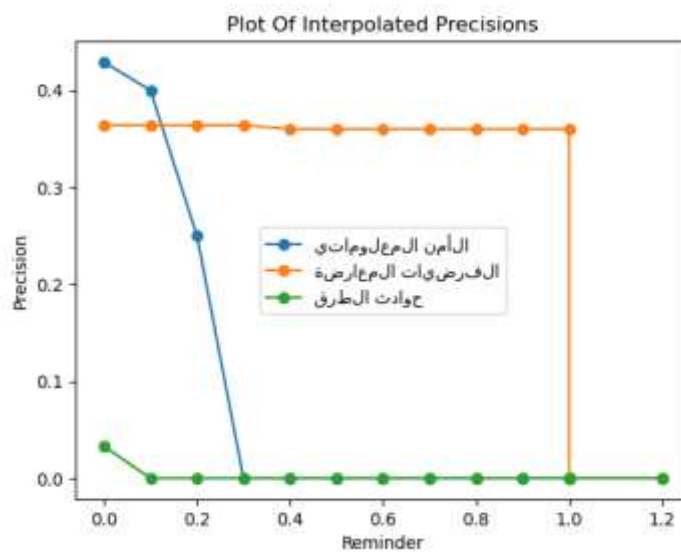
AVG
0.151 <----- الأمن المعلوماتي
0.135 <----- الفرضيات المعارضة
0.006 <----- حوادث الطرق
MAP Awn = 0.097333333333333334

```

- *Précision interpolée*

Rappel	الأمن المعلوماتي PreInt	الفرضيات المعارضة PreInt	حوادث الطرق PreInt
0	0.43	0.36	0.03
0.10	0.40	0.36	0
0.20	0.25	0.36	0
0.30	0	0.36	0
0.40	0	0.36	0
0.50	0	0.36	0
0.60	0	0.36	0
0.70	0	0.36	0
0.80	0	0.36	0
0.90	0	0.36	0
1	0	0.36	0

- *Courbe:*



Calcul précision au P5,10,P15,P20

```

P5AWN=Prec_Pos(ManuallyQueryFindAWN,5)
P10AWN=Prec_Pos(ManuallyQueryFindAWN,10)
P15AWN=Prec_Pos(ManuallyQueryFindAWN,15)
P20AWN=Prec_Pos(ManuallyQueryFindAWN,20)

print("P5",P5AWN)
print("P10",P10AWN)
print("P15",P15AWN)
print("P20",P20AWN)

P5 {'الأمن المعلوماتي': [0.05], 'الفرضيات المعارضة': [0.04], 'حوادث الطرق': [0.0]}
P10 {'الأمن المعلوماتي': [0.1412], 'الفرضيات المعارضة': [0.0533], 'حوادث الطرق': [0.0]}
P15 {'الأمن المعلوماتي': [0.09413333333333333], 'الفرضيات المعارضة': [0.05093333333333334], 'حوادث الطرق': [0.0]}
P20 {'الأمن المعلوماتي': [0.0706], 'الفرضيات المعارضة': [0.06205], 'حوادث الطرق': [0.0]}

```


Les Résultat de Reformulation Automatique+Validation utilisateur:

1. Model Frequency ():

- *élimination du mots et lemmatisation:*

```
wordsv={}
wordsv['الطرق']=["طريق","طرق"]
wordsv['حوادث']=["حمل","دار","جرى","وقع","خادثه","خادث","خمول","خدون","خوات"]
wordsv['المعارضة']=["المعارضة"]
wordsv['الفرضيات']=["الفرضيات"]
wordsv['المعلوما تي']=["معلومة","معلوما تي"]
wordsv['الامن']=["امن","مان","مين"]

✓ 0.6s

for key,value in wordsv.items():
    wordsv[key]=[stemmer("arabic").stemWord(word) for word in value]
QueryAWN=[
    '.join(wordsv['الامن']+wordsv['المعلوما تي']),
    '.join(wordsv['الفرضيات']+wordsv['المعارضة']),
    '.join(wordsv['حوادث']+wordsv['الطرق'])]
```

- *Résultat de Recherche:*

```
Searcher

i=0;
for query in QueryList:
    Result,NbrDocAWN[QueryAWN[i]]=objet.Searcher(QueryAWN[i],score,20)
    ManuallyQueryFindAWN=Pertinance(ManuallyQueryFindAWN,Result,query)
    i=i+1
```

```
----- (content:امن OR content:مان OR content:مين OR content:معلوما OR content:معلوما) -----
<Top 20 Results for Or([Term('content','امن'), Term('content','مان'), Term('content','مين'), Term('content','معلوما'), Term('content','معلوما')]) runtime=0.003083027608865045>
20
Scored 20 of between 14 and 96 documents
----- (content:معارضة OR content:معارضة) -----
<Top 20 Results for Or([Term('content','معارضة'), Term('content','معارضة')]) runtime=0.0006848379994374234>
20
Scored 20 of between 3 and 23 documents
----- (content:حمل OR content:دار OR content:جرى OR content:وقع OR content:خادثه OR content:خادث OR content:خمول OR content:خوات OR content:حوادث OR content:حوادث OR content:طريق OR content:طرق) -----
<Top 20 Results for Or([Term('content','حمل'), Term('content','دار'), Term('content','جرى'), Term('content','وقع'), Term('content','خادثه'), Term('content','خادث'), Term('content','خمول'), Term('content','خوات'), Term('content','حوادث'), Term('content','حوادث'), Term('content','طريق'), Term('content','طرق')]) runtime=0.005177814800489889>
20
Scored 20 of between 1 and 496 documents
```

A. الامن المعلوماتي (134 documents pertinent)

Tu trouves le tableau dans “./DFQuery/DFRVQ1Frequency.xlsx”

B. الفرضيات المعارضة (25 documents pertinent)

Tu trouves le tableau dans “./DFQuery/DFRVQ2Frequency.xlsx”

C. حوادث الطرق (113 documents pertinent) :

Tu trouves le tableau dans “./DFQuery/DFRVQ3Frequency.xlsx”

- *F_Mesure*

```
Calculate Reminder And Precision AND FMesure

C_RPFM(ManuallyQueryFindAWNv,QueryList)
✓ 0.5s

F_Mesure
0.265 <===== الأمن المعلوماتي
0.84 <===== الفرضيات المعارضة
0.135 <===== حوادث الطرق
```

- *MAP et AVG*

```
Calculate Average AND MAP

AVGQueryListAWNv={}
MAP_AWNv=0
for query in QueryList:
    AVGQueryListAWNv[query]=Calcul_AVGPrec(ManuallyQueryFindAWNv,query,NbrDoc[query])
    MAP_AWNv=MAP_AWNv+AVGQueryListAWNv[query]
MAP_AWNv=MAP_AWNv/len(QueryList)
print("MAP_AWNv = ",MAP_AWNv)
✓ 0.1s

0.191 <----- الأمن المعلوماتي
0.96 <----- الفرضيات المعارضة
0.11 <----- حوادث الطرق
MAP_AWNv = 0.4203333333333334
```

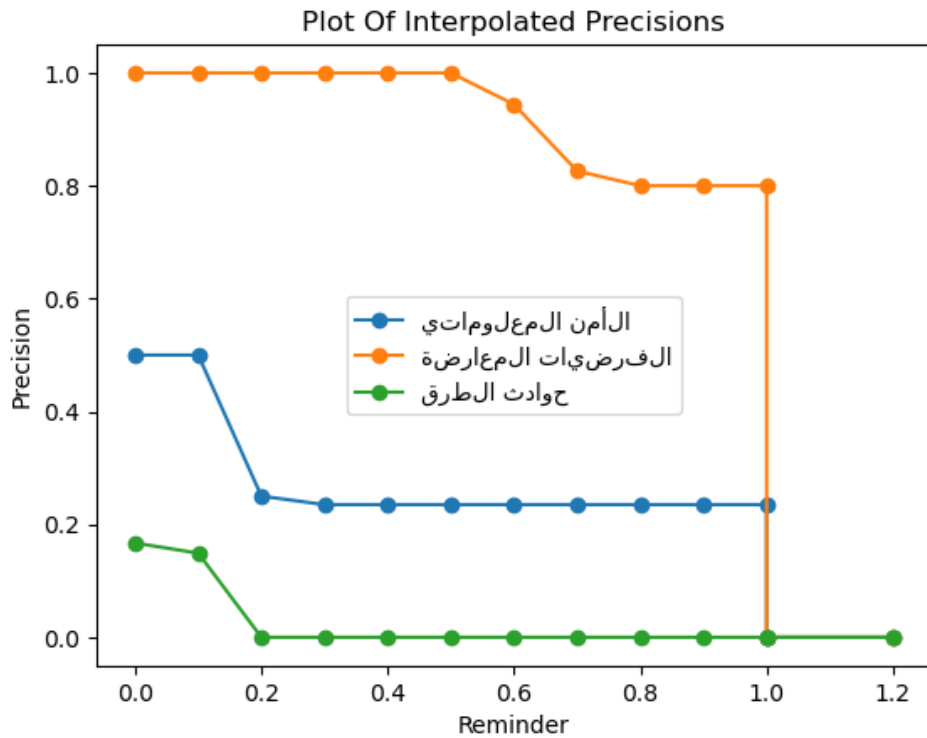
- *Précision interpolée*

```
Calculate Interpolated For Each Query

Dict_InterAWNv={"Rappel":Rappel}
for query in QueryList:
    Dict_InterAWNv["PreInt "+query]=InterpolatedPrecision(Rappel,TablePertinanceAWNv,query)
Dict_InterAWNv=pd.DataFrame(Dict_InterAWNv)
✓ 0.2s
```

Rappel	PreInt الأمن المعلوماتي	PreInt الفرضيات المعارضة	PreInt حوادث الطرق
0	0.50	1	0.17
0.10	0.50	1	0.15
0.20	0.25	1	0
0.30	0.24	1	0
0.40	0.24	1	0
0.50	0.24	1	0
0.60	0.24	0.94	0
0.70	0.24	0.83	0
0.80	0.24	0.80	0
0.90	0.24	0.80	0
1	0.24	0.80	0

- *Traçage de courbe:*



Calcul précision au P5, 10, P15, P20

```
P5AWNVPrec_Pos(ManuallyQueryFindAWNVP5)
P10AWNVPrec_Pos(ManuallyQueryFindAWNVP10)
P15AWNVPrec_Pos(ManuallyQueryFindAWNVP15)
P20AWNVPrec_Pos(ManuallyQueryFindAWNVP20)
✓ 0.1s

print("P5",P5AWNVP)
print("P10",P10AWNVP)
print("P15",P15AWNVP)
print("P20",P20AWNVP)
✓ 0.1s

P5 {'[0.0]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.05]': 'الأمن المعلوماتي'}
P10 {'[0.017]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.196]': 'الأمن المعلوماتي'}
P15 {'[0.011]': 'حوادث الطرق', '[0.929]': 'الفرضيات المعارضة', '[0.13]': 'الأمن المعلوماتي'}
P20 {'[0.014]': 'حوادث الطرق', '[0.838]': 'الفرضيات المعارضة', '[0.098]': 'الأمن المعلوماتي'}
```

2. Model BM25 F ():

- Résultat de recherche:

```

----- (content:معلومات OR content:معلومات OR content:معلومات OR content:معلومات OR content:معلومات) -----
<Top 20 Results for Or([Term('content', 'معلومات'), Term('content', 'معلومات'), Term('content', 'معلومات'), Term('content', 'معلومات'), Term('content', 'معلومات')]) runtime=0.00446085600196966>
28
Scored 28 of between 14 and 96 documents
----- (content:معارضة OR content:معارضة) -----
<Top 20 Results for Or([Term('content', 'معارضة'), Term('content', 'معارضة')]) runtime=0.0010487959989404771>
28
Scored 28 of between 3 and 23 documents
----- (content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق OR content:حوادث الطرق) -----
<Top 20 Results for Or([Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق'), Term('content', 'حوادث الطرق')]) runtime=0.009072430082561305>
28
Scored 28 of between 1 and 496 documents

```

- A. الأمن المعلوماتي (1 34 documents pertinent)**
Tu trouves le tableau dans “./DFQuery/DFRVQ1BM25F.xlsx”
 - B. الفرضيات المعارضة (1 25 documents pertinent)**
Tu trouves le tableau dans “./DFQuery/DFRVQ2BM25F.xlsx”
 - C. حوادث الطرق (1 133 documents pertinent) :**
Tu trouves le tableau dans “./DFQuery/DFRVQ3BM25F.xlsx”
- *F_mesure*

```

Calculate Reminder And Precision AND FMeasure

C_RPFM(ManuallyQueryFindAWN,QueryList)
✓ 0.3s

F_Mesure
0.265 <===== الأمن المعلوماتي
0.84 <===== الفرضيات المعارضة
0.135 <===== حوادث الطرق

```

- *AVG ET Map*

```

Calculate Average AND MAP

AVGQueryListAWN={}
MAP_AWNV=0
for query in QueryList:
    AVGQueryListAWN[query]=Calcul_AVGPrec(ManuallyQueryFindAWN,query,NbrDoc[query])
    MAP_AWNV=MAP_AWNV+AVGQueryListAWN[query]
MAP_AWNV=MAP_AWNV/len(QueryList)
print("MAP_AWNV = ",MAP_AWNV)
✓ 0.6s

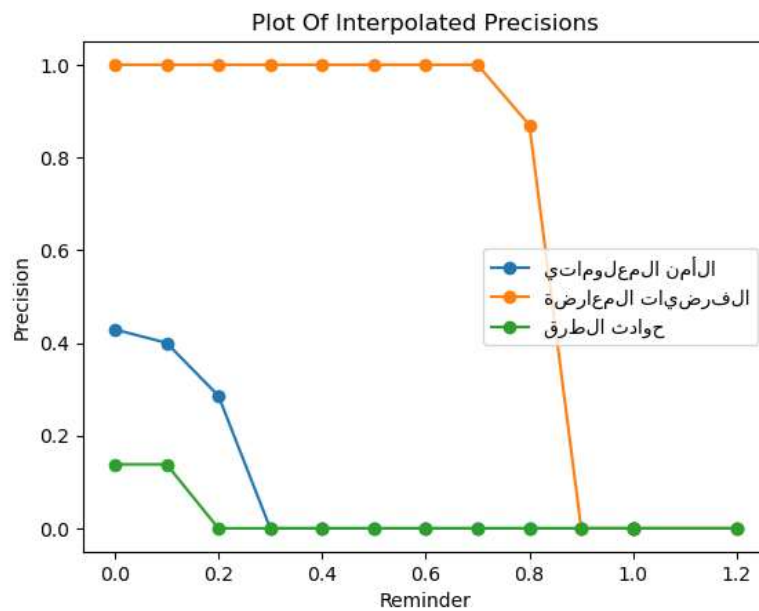
0.173 <----- الأمن المعلوماتي
0.987 <----- الفرضيات المعارضة
0.103 <----- حوادث الطرق
MAP_AWNV = 0.421

```

- *Précision interpolée*

Rappel	الأمن المعلوماتي PreInt	الفرضيات المعارضة PreInt	حوادث الطرق PreInt
0	0.43	1	0.14
0.10	0.40	1	0.14
0.20	0.29	1	0
0.30	0	1	0
0.40	0	1	0
0.50	0	1	0
0.60	0	1	0
0.70	0	1	0
0.80	0	0.87	0
0.90	0	0	0
1	0	0	0

- *Courbe*



Calcul précision au points

```

P5AWN=Prec_Pos(ManuallyQueryFindAWN,5)
P10AWN=Prec_Pos(ManuallyQueryFindAWN,10)
P15AWN=Prec_Pos(ManuallyQueryFindAWN,15)
P20AWN=Prec_Pos(ManuallyQueryFindAWN,20)
✓ 0.3s

print("P5",P5AWN)
print("P10",P10AWN)
print("P15",P15AWN)
print("P20",P20AWN)
✓ 0.3s

P5 {'[0.0]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.05]': 'الأمن المعلوماتي'}
P10 {'[0.0]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.141]': 'الأمن المعلوماتي'}
P15 {'[0.004]': 'حوادث الطرق', '[1.0]': 'الفرضيات المعارضة', '[0.094]': 'الأمن المعلوماتي'}
P20 {'[0.008]': 'حوادث الطرق', '[0.9]': 'الفرضيات المعارضة', '[0.071]': 'الأمن المعلوماتي'}

```


Conclusion -Tableaux récapitulatifs:-

Tableau1: "Table1.xlsx"

	Mots de requête1	Mots de requête2	Mots de requête3
Requête originale	الأمن المعلوماتي	الفرضيات المعارضة	حوادث الطرق
Requête Reformulée automatique	أمن أكد ضمن ضمان كفل تحقيق معلوم	أكد فرض أصاب وجه عرض جلب قدم أظهر عرض صور مثل	حصل دأر جري وقع حادث حدث حصول حدوث طريق سبيل أبح جهة وجه مدخل شارع
Requête Reformulée automatique +	ضمن ضمان أمن معلوم معلومات	فرض معارض	حصل دأر جري وقع حادث حدث حصول حدوث حوادث طريق طرق

Tableau 2: "Table2.xlsx"

A ▼	B ▼
Mots	Existe En AWN
الطرق	Oui
حوادث	Oui
المعارضة	Non
الفرضيات	Non
المعلوماتي	Oui
الأمن	Oui

Tableau 3

Rappel	الأمن المعلوماتي Précision Interpolée					
	Sans reformulation		Reformulation automatique		Reformulation automatique+ validation de l'utilisateur	
	BM25F	Frequency	BM25F	Frequency	BM25F	Frequency
0	0.50	0.50	0.43	0.50	0.43	0.50
0.10	0.41	0.41	0.40	0.50	0.40	0.50
0.20	0.41	0.41	0.25	0.27	0.29	0.25
0.30	0.41	0.41	0	0.27	0	0.24
0.40	0.41	0.41	0	0.27	0	0.24
0.50	0.41	0.41	0	0.27	0	0.24
0.60	0.41	0.41	0	0.27	0	0.24
0.70	0.41	0.41	0	0.27	0	0.24
0.80	0.41	0.41	0	0.27	0	0.24
0.90	0.41	0.41	0	0.27	0	0.24
1	0.41	0.41	0	0.27	0	0.24

Rappel	حوادث الطرق Précision Interpolée					
	Sans reformulation		Reformulation automatique		Reformulation automatique+ validation de l'utilisateur	
	BM25F	Frequency	BM25F	Frequency	BM25F	Frequency
0	0.50	0.50	0.03	0.09	0.14	0.17
0.10	0.18	0.16	0	0	0.14	0.15
0.20	0	0	0	0	0	0
0.30	0	0	0	0	0	0
0.40	0	0	0	0	0	0
0.50	0	0	0	0	0	0
0.60	0	0	0	0	0	0
0.70	0	0	0	0	0	0
0.80	0	0	0	0	0	0
0.90	0	0	0	0	0	0
1	0	0	0	0	0	0

Rappel	الفرصيات المعارضة Précision Interpolée					
	Sans reformulation		Reformulation automatique		Reformulation automatique+ validation de l'utilisateur	
	BM25F	Frequency	BM25F	Frequency	BM25F	Frequency
0	1	1	0.36	1	1	1
0.10	1	1	0.36	0.14	1	1
0.20	1	1	0.36	0	1	1
0.30	1	1	0.36	0	1	1
0.40	1	1	0.36	0	1	1
0.50	1	1	0.36	0	1	1
0.60	1	0.94	0.36	0	1	0.94
0.70	1	0.83	0.36	0	1	0.83
0.80	0.87	0.80	0.36	0	0.87	0.80
0.90	0	0.80	0.36	0	0	0.80
1	0	0.80	0.36	0	0	0.80

Rappel	Précision Interpolée Q1+Q2+Q3					
	Sans reformulation		Reformulation automatique		Reformulation automatique+ validation de l'utilisateur	
	BM25F	Frequency	BM25F	Frequency	BM25F	Frequency
0	0.67	0.67	0.27	0.53	0.52	0.56
0.10	0.53	0.52	0.25	0.21	0.51	0.55
0.20	0.47	0.47	0.20	0.09	0.43	0.42
0.30	0.47	0.47	0.12	0.09	0.33	0.41
0.40	0.47	0.47	0.12	0.09	0.33	0.41
0.50	0.47	0.47	0.12	0.09	0.33	0.41
0.60	0.47	0.45	0.12	0.09	0.33	0.39
0.70	0.47	0.41	0.12	0.09	0.33	0.36
0.80	0.43	0.40	0.12	0.09	0.29	0.35
0.90	0.14	0.40	0.12	0.09	0	0.35
1	0.14	0.40	0.12	0.09	0	0.35

Tableau Final

Precision	Sans reformulation		Reformulation automatique		Reformulation automatique+ validation de l'utilisateur	
	BM25F	Frequency	BM25F	Frequency	BM25F	Frequency
P@ 5	0,433	0,433	0,03	0,15	0,35	0,35
P@ 10	0,422	0,411	0,065	0,132	0,38	0,404
P@ 15	0,4	0,369	0,048	0,265	0,366	0,36
P@ 20	0,348	0,332	0,051	0,067	0,326	0,317
AVGPre	0,525	0,499	0,097	0,121	0,421	0,42
MAP	0,525	0,499	0,097	0,121	0,421	0,42
RelevantDoc	64	64	64	64	64	64
returned sys	54	161	278	278	205	205
Returned real Doc	30,00	50,00	50,00	50,00	104,00	105,00