محمدحسین صباغی ۹۹۲۳۰۴۴

گزارش پروژه بازیابی اطلاعات

۱)پیش پردازش

عملیات پیش پردازش شامل سه بخش است که در فایل preprocess.py قابل مشاهده است.

۱/۱)tokenizer

در این بخش تمرکز روی جدا کردن کلمات از یکدیگر وتولید توکن ها میباشد. در ابتدا چک میشود که اگر آخر یک کلمه باشیم(به وسیله اسپیس یا خط جدید یا...) کلمه را بعنوان توکن در نظر بگیریم و به سراغ کلمه بعدی برویم.

سپس کلماتی مانند ((می خواهم)) بررسی میشوند که بصورت یک کلمه ذخیره شوند و نه دو کلمه ((می)) و ((خواهم)).

در ادامه به بررسی مخفف ها میپردازیم و آنهارا به کلمه اصلی خود تبدیل میکنیم.

و در آخر ایمیل و آیدی را پردازش میکنیم.

```
42
43  def tokenize(input:str)->list[str]:
44      output = []
45      word = []
46      for i in range(len(input)):
47          if input[i] in space_chars:
48
49              if len(word) < 1:
50                  continue
51              #check verbs
52              if "".join(word) in special_subwords_before:
53                  word.append(special_space)
54                  continue
55              #check abbrevations
56              if "".join(word) in abbrs:
57                  output.append(abbrs["".join(word)])
58                  word = []
59                  continue
60              #check id and email
61              if('@' in word):
62                  if len(word) == 1:
63                      temp = output.pop()
64                      output.append("@" + "".join(temp))
65                      word = []
66                      continue
67                  elif word[-1] == '@':
68                      output.append("@" + "".join(word[:-1]))
69                      word = []
70                      continue
71
72              output.append("".join(word))
73              word = []
74          else:
75              word.append(input[i])
76      output.append("".join(word))
77      return output
```

normalizer(۱٫۲

```python
def normalize(input:list[str])->list[str]:
    output = []
    for i in range(len(input)):
        if '@' not in input[i]:
            k=0
            number_flag = False
            word = []
            for j in range(len(input[i])):
                if input[i][j] in should_change_to:
                    word.append(should_change_to[input[i][j]])
                else:
                    word.append(input[i][j])

                if word[j] in to_be_deleted_chars or (number_flag and word[j] not in numbers):
                    output.append("".join(word[k:j]))
                    k = j + 1
                    number_flag = False
                elif word[j] in numbers and not number_flag:
                    output.append("".join(word[k:j]))
                    k = j + 1
                    number_flag = True

            if len(word[k:]) > 0:
                output.append("".join(word[k:]))

            if input[i] in special_subwords_after:
                temp = output.pop()
                temp2 = output.pop()
                output.append(temp2 + "|" + temp)
                continue

        else:
            output.append(input[i])

    return output
```

در این بخش به پردازش توکن های تولید شده میپردازیم. ابتدا چک میکنیم که اگر ایمیل یا آیدی بود در آن تغییری ایجاد نکنیم.

سپس به وسیله یک دیکشنری از حروفی که باید تغییر کنند حروف یکسان با شکل های متفاوت را یکسان میکنیم.(مثلا آ و ا)

سپس به بررسی حروفی که باید حذف شوند و همچنین پردازش اعداد میپردازیم.

در آخر کلماتی مانند زیباتر را بصورت یک کلمه در نظر میگیریم مه دو کلمه جدا از هم ((زیبا)) و ((تر)).

stemming(۱٫۳

```python
def stemming(input:list[str])->list[str]:
    output = []
    for i in input:
        temp = stem.stem(i)
        if len(temp) < 3:
            temp = i
        output.append(lem.lemmatize(temp, 'V'))
    return output
```

در این بخش توسط دو کتابخانه کلمات ریشه یابی میشوند تا کلمات هم ریشه مانند هم در نظر گرفته شوند.

۲)ساخت ایندکس

در این بخش توسط سه استراکچر به ساخت inverted index میپردازیم.

۲/۱)positional posting

```
class PositionalPosting:
    def __init__(self, doc_id: int, tf : int, positions):
        self.doc_id = doc_id
        self.tf = tf
        self.positions = positions

    def getDocID(self):
        return self.doc_id

    def getTF(self):
        self.tf = len(self.positions)
        return self.tf

    def getPositions(self):
        return self.positions

class PostingsList:
    def __init__(self):
        self.df = 0
        self.list = []
        self.tf = 0
```

این استراکچر برای نگاه داشتن پستینگ مکانی برای یک توکن درون یک داکیومنت میباشد که آیدی داک مورد نظر و عداد تکرار توکن در داک و همچنین مکان های تکرار توکن در داکیومنت را نگه میدارد.

همچنین شامل سه فانکشن برای گرفتن این اطلاعات میباشد.

PosingList(۲/۲

این استراکچر برای نگه داشتن تمام پستینگ های مکانی یک توکن میباشد که شامل اطلاعاتی مانند تعداد داکیومنت های حاوی آن توکن(df) و تعداد تکرار توکن در کل داکیومنت ها (tf) و لیستی از پستینگ های مکانی آن میباشد.

همچنین شامل فانکشن هایی برای دریافت این اطلاعات و همچنین افزودن پستینگ جدید و افزودن به df و دریافت tfهای هم داکیومنت میشود.

```python
class PostingsList:
    def __init__(self):
        self.df = 0
        self.list = []
        self.tf = 0

    def getPostings(self):
        return self.list

    def getTf(self):
        frequency = 0
        for posting in self.list:
            frequency += posting.getTF()

        self.tf = frequency
        return frequency

    def getDF(self):
        return self.df

    def addPosting(self, p: PositionalPosting):
        self.list.append(p)

    def addDF(self):
        self.df += 1

    def getAllDocIdTF(self):
        res = []
        for p in self.list:
            res.append((p.getDocID(), p.getTF()))
        return res
```

InvertedIndex(۲/۳

در این استراکچر عملیات ساخت و ذخیره سازی inverted index و همچنین تولید بردار های داکیومنت ها و ذخیره آنها و ذخیره لیستchampion انجام می شود.

این بخش شامل سه دیکشنری برای تمام PostingListها و لیست championو همچنین بردار های نشان دهنده داکیومنت ها در فضای برداری میشود. همچنین تعداد تمام داکیومنت ها هم برای محاسبه امتیاز نگه داری میشود.

فانکشن هایی برای دریافت اطلاعات مختلف و همچنین افزودن پستینگ لیست جدید وجود دارند.

علاوه بر این سه فانکشن برای پاک کردن k توکن پرتکرار استفاده شده است.

دو فانکشن هم برای سیو و لود کردن مدل استفاده شده است تا هربار نیاز به تولید دوباره ایندکس نباشد.

در آخر هم دو فانکشن addScore و CreateVers برای تولید بردار های داکیومنت ها تعریف شده اند.

```python
class InvertedIndex:
    def __init__(self, docs_num : int):
        self.dict = {}
        self.docs_num= docs_num
        self.champions = {}
        self.vertices = {}

    def getPostingList(self, term: str) -> PostingsList:
        if term not in self.dict:
            return None
        return self.dict[term]

    def getChampionPostingList(self, term: str) -> PostingsList:
        if term not in self.champions:
            return None
        return self.champions[term]

    def getdocsNum(self):
        return self.docs_num

    def addPosting(self, term: str, doc_id: int, tf: int, positions = None):
        if term not in self.dict:
            self.dict[term] = PostingsList()

        self.dict[term].addDF()
        self.dict[term].addPosting(PositionalPosting(doc_id, tf, positions))

        if tf >= 3:
            if term not in self.champions:
                self.champions[term] = PostingsList()
            self.champions[term].addDF()
            self.champions[term].addPosting(PositionalPosting(doc_id, tf, positions))
```

```python
def deleteRepeatedWords(self, k: int):
    heap = []
    for key in self.dict:
        heapq.heappush(heap, (-self.dict[key].getTf(), key))
    remove_list = []
    for _ in range(k):
        item = heapq.heappop(heap)
        self.delete(item[1])
        remove_list.append(item)

    heap = []
    for key in self.champions:
        heapq.heappush(heap, (-self.champions[key].getTf(), key))
    champion_remove_list = []
    for _ in range(k):
        item = heapq.heappop(heap)
        self.deleteChamp(item[1])
        champion_remove_list.append(item)

    return [(-key, value) for key, value in remove_list] ,[(-key, value) for key, value in champion_remove_list]

def delete(self, term):
    self.dict.pop(term)

def deleteChamp(self, term):
    self.champions.pop(term)

def save(obj, path, mode):
    with open(path, mode) as file:
        pickle.dump(obj, file)
    print(f"{obj} saved.")
```

```
def load(path, mode):
    with open(path, mode) as file:
        obj = pickle.load(file)
    return obj

def addScore(self,token: str):
    postings = self.getPostingList(token)
    id_tf_pairs = postings.getAllDocIdTF()
    df = postings.getDF()
    for (id, tf) in id_tf_pairs:
        try:
            self.vertices[str(id)][token] = (1 + log10(tf)) * log10(self.getdocsNum() / df)
        except:
            pass

def createVers(self, data):
    for i in range(len(data)):
        doc_id = str(i)
        self.vertices[doc_id] = {}
    for i in self.dict:
        self.addScore(i)
```

```
f create_index(data, delete):
  inverted = InvertedIndex(len(data))
  for i in range(len(data)):
      content = data[str(i)]['content']
      tokens = preprocess.preproccess(content)
      terms = {}
      for j in range(len(tokens)):
          if tokens[j] not in terms:
              terms[tokens[j]] = [0, []]

          terms[tokens[j]][0] += 1
          terms[tokens[j]][1].append(j + 1)

      for key in terms:
          inverted.addPosting(key, i + 1, terms[key][0], terms[key][1])

      if i % 1000 == 0:
          print("docs processed: ", i)
  print("processing finished\ndeleting for inverted...")
  inverted_deletes = inverted.deleteRepeatedWords(delete)
  print("normal deletes = ", inverted_deletes[0])
  print("champion deletes = ", inverted_deletes[1])
  print("creating vers...")
  inverted.createVers(data)
  print("vers finished")

  return inverted
```

فانکشن create_index برای تولید کردن ایندکس تعریف شده است که در فایل main با فراخوانی این تابع ایندکس ساخته و سپس ذخیره میشود.

۲/۴)کلمات پرتکرار حذف شده

```
normal deletes =  [(219489, 'و'), (164431, 'رد'), (133643, 'ه‌ب'), (131956, '#تسه'), (92960, 'زا'), (838
38, 'ی‌ا'), (76187, 'ه‌ک'), (70005, 'ن‌ک#درک'), (69065, 'اب'), (68895, 'ار'), (59098, 'وش#دش'), (58166, '
#تسا'), (43344, 'ش‌اب#دوب'), (41085, 'راد#تشاد'), (30993, 'ارب'), (23648, 'کی'), (22371, 'مه'), (22261,
'روشک'), (22237, 'ه‌داد#ه‌د'), (22020, 'مه‌ت'), (19767, 'ام'), (18866, 'دوخ'), (17525, 'وگ#تفگ'), (16969, '
ان'), (16876, 'زاب'), (16681, 'رب'), (16203, 'دی‌اب'), (15881, 'ات'), (15616, 'الس‌ا'), (15486, 'یو'), (1
5235, 'ام‌ا'), (15189, 'ری‌ا'), (15052, 'ریم#درم'), (14555, 'سلچم'), (14232, 'سراف'), (13790, 'رازگ'), (1
3375, 'لود'), (12804, 'راک'), (12779, 'سی‌ر'), (12413, 'رازگربخ'), (12250, 'ی‌اه‌تن‌ا'), (11715, 'م‌ای‌پ'),
 (11373, 'ن‌کی‌زاب'), (11201, 'ل‌اس'), (10691, 'ور#تفر'), (10151, 'یلم'), (10055, 'ه‌کن‌ی‌ا'), (9497, 'رارق')
, (9416, 'ه‌اوخ#تساوخ'), (9124, 'ود')]
champion deletes =  [(217244, 'و'), (162665, 'رد'), (131464, 'ه‌ب'), (128105, '#تسه'), (88577, 'زا'), (7
8933, 'ی‌ا'), (71306, 'ه‌ک'), (64697, 'ن‌ک#درک'), (63655, 'اب'), (63399, 'ار'), (52989, 'وش#دش'), (52792,
 '#تسا'), (37646, 'ش‌اب#دوب'), (35296, 'راد#تشاد'), (24695, 'ارب'), (18404, 'مه‌ت'), (18063, 'مه'), (1793
8, 'کی'), (17859, 'روشک'), (17070, 'ام'), (16074, 'ه‌داد#ه‌د'), (13290, 'زاب'), (12923, 'دوخ'), (12629, 'ا
ن'), (12531, 'سلچم'), (12301, 'دی‌اب'), (12098, 'ریم#درم'), (11936, 'الس‌ا'), (11877, 'وگ#تفگ'), (11732,
'ری‌ا'), (11335, 'لود'), (10938, 'ام‌ا'), (10622, 'رب'), (10035, 'سی‌ر'), (10024, 'ات'), (9323, 'یو'), (8
598, 'ن‌کی‌زاب'), (8572, 'راک'), (7747, 'ب‌الق‌ن‌ا'), (7378, 'ل‌اس'), (7002, 'یلم'), (6475, 'ه‌کن‌ی‌ا'), (6082,
'ور#تفر'), (6049, 'بوخ'), (6004, 'روه‌مچ'), (5812, 'من'), (5773, 'ه‌اگ‌ش‌اب'), (5667, 'اروش'), (5529, 'ون‌اق
ن'), (5147, 'ل‌ابتوف')]
```

کلمات حذف شده لیست اصلی:

و-><۲۱۹۴۸۹ در-><۱۶۴۴۳۱ به-><۱۳۳۶۴۳ هست(و انواع آن)-><۱۳۱۹۵۶
از-><۹۲۹۶۰ این-><۸۳۸۳۸ که-><۷۶۱۸۷

کرد(و انواع آن)-><۷۰۰۵ با-><۶۹۰۶۵ را-><۶۸۸۹۵
شد(و انواع آن)-><۵۹۰۹۸ است-><۵۸۱۶۶ بود-><۴۳۳۴۴ داشت-><۴۱۰۸۵

و...

کلمات حذف شده لیست champion:

مانند حالت اصلی است بغیر از اواخر که فرق میکند.

۳) QueryProcessor

این بخش که در فایل querryrunner.py قرار گرفته است وظیفه پردازش کویری کاربر و برگرداندن نتیجه را دارد که توسط فایل runner.py فراخوانی میشود.

دو فانکشن اصلی آن findNormal و FindChampion هستند. که ابتدا از لیست Champion برای جستجو استفاده میشود و در صورتی که به وسیله آن تعداد داکیومنت دلخواه به دست نیاید سراغ normal میرویم.

در هر دو فانکشن ابتدا کلماتی از کویری که در دیکشنری نیستند حذف میشوند و امتیاز باقی کلمات محاسبه میشود. و سپس داکیومنت ها به صورت برداری دریافت میشوند و به وسیله ی فانکشن دیگری شباهت کوسینوسی بین کویریو بردار های داکیومنت ها به دست می آید و داکبونت ها با امتیاز بیشتر برگردانده میشوند.

```python
class QueryProcessor:
    def __init__(self, index: index.InvertedIndex):
        self.index = index

    def findNormal(self, query: str, k: int):
        tokens = preprocess.preproccess(query.strip())
        remove_list = []
        for i in range(len(tokens)):
            pl = self.index.getPostingList(tokens[i])
            if pl is None:
                remove_list.append(i)
        for i in remove_list:
            tokens.pop(i)

        query_scores = {}
        query_frequency = {}
        for i in tokens:
            if i not in query_frequency:
                query_frequency[i] = 0
            query_frequency[i] += 1

        for i in query_frequency:
            query_scores[i] = 1 + log10(query_frequency[i])

        doc_scores = {}
        for i in tokens:
            scores = self.getScoreList(i)
            for id in scores:
                if id not in doc_scores:
                    doc_scores[id] = {}
                doc_scores[id][i] = scores[id]

        heap = []
        for docID, doc_vector in doc_scores.items():
            doc_score = self.getSimilarity(query_scores, doc_vector)
            heapq.heappush(heap, (-doc_score, docID))

        if len(heap) < k:
            k = len(heap)

        result = []
        for _ in range(k):
            neg_score, id = heapq.heappop(heap)
            result.append((id, -neg_score))

        return result

    def findChampion(self, query: str, k: int):
        tokens = preprocess.preproccess(query.strip())
        remove_list = []
        for i in range(len(tokens)):
            pl = self.index.getChampionPostingList(tokens[i])
            if pl is None:
                remove_list.append(tokens[i])
        for i in remove_list:
            tokens.remove(i)

        query_scores = {}
        query_frequency = {}
        for i in tokens:
            if i not in query_frequency:
                query_frequency[i] = 0
            query_frequency[i] += 1

        for i in query_frequency:
            query_scores[i] = 1.0 + log10(query_frequency[i])

        doc_scores = {}
        for i in tokens:
            scores = self.getScoreListChampion(i)
            for id in scores:
                if id not in doc_scores:
                    doc_scores[id] = {}
                doc_scores[id][i] = scores[id]
        heap = []
        for docID, doc_vector in doc_scores.items():
            doc_score = self.getSimilarity(query_scores, doc_vector)
            heapq.heappush(heap, (-doc_score, docID))

        if len(heap) < k:
            result = []
            for _ in range(len(heap)):
                neg_score, id = heapq.heappop(heap)
                result.append((id, -neg_score))
```

```
                result.append((id, -neg_score))
            normal_res = self.findNormal(query, 2 * k)
            for i in range(len(normal_res)):
                if normal_res[i] not in result:
                    result.append(normal_res[i])
        else:
            result = []
            for _ in range(k):
                neg_score, id = heapq.heappop(heap)
                result.append((id, -neg_score))
        return result

    def getScoreList(self, token: str):
        scores = self.index.vertices
        postings = self.index.getPostingList(token)
        result = {}
        for i in postings.list:
            result[str(i.getDocID())] = scores[str(i.getDocID())][token]
        return result

    def getScoreListChampion(self, token: str):
        scores = self.index.vertices
        postings = self.index.getChampionPostingList(token)
        result = {}
        for i in postings.list:
            result[str(i.getDocID())] = scores[str(i.getDocID())][token]
        return result

    def getSimilarity(self, vector1: dict, vector2: dict):
        squared_sum1 = 0.0
        squared_sum2 = 0.0

        for t in vector1:
            squared_sum1 += vector1[t] ** 2
        for t in vector2:
            squared_sum2 += vector2[t] ** 2

        squared_sum1 = sqrt(squared_sum1)
        squared_sum2 = sqrt(squared_sum2)

        score = 0.0
        for i in vector1:
            if i in vector2:
                score += (vector1[i] / squared_sum1) * (vector2[i] / squared_sum2)
```

```
    for i in vector1:
        if i in vector2:
            score += (vector1[i] / squared_sum1) * (vector2[i] / squared_sum2)

    return score
```

۴)کویری ها

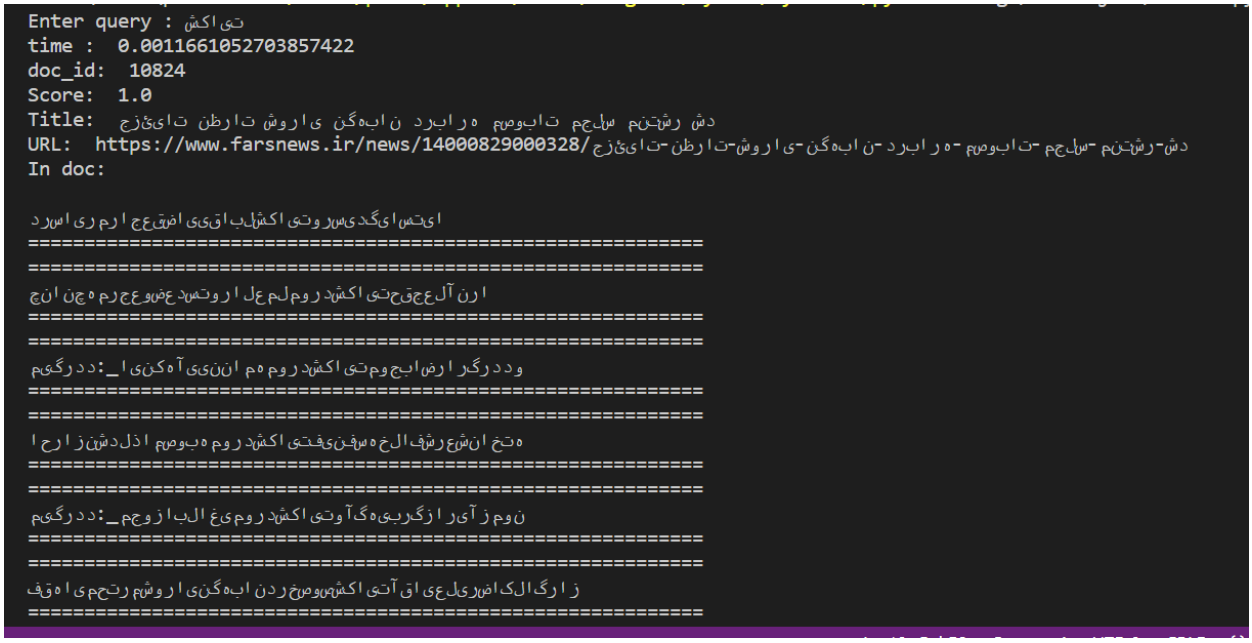
با اجرای فایل runner.py میتوان کویری زد و نتیجه را دریافت کرد.

در زیر چند کویری نشان داده شده است.

۴/۱) شکایت:

نتایج کویری یک کلمه ای ساده ((شکایت)) را میتوان در زیر مشاهده کرد. به دلیل اینکه کویری یک کلمه ای است امتیاز tfidf آن به وسیله نرمالسازی محاسبه شده است، برای اسنادی که حداقل یکبار آنرا دارند ۱ میشود. همچنین برای هر داک مکان هایی که شکایت آمده است نشان داده شده است.

نتایج بدست آمده مرتبط هستند و در حوزه موارد قضایی و شکایت میباشند.



```
------------------------------------------------------------

doc_id:  11406
Score:  1.0
Title:  وزارت نفت شکایت خود از یک خبرنگار را متوقف کند
URL:  https://www.farsnews.ir/news/14000816000080/وزارت-نفت-شکایت-خود-از-یک-خبرنگار-را-متوقف-کند
In doc:

بهوزیرنفتخواستارتوقفشکایتوزارتنفتازیک
============================================================
============================================================
ازسویوزارتنفتموردشکایتقضاییقرارگرفتهاند.
============================================================
============================================================
باافکارعمومی،دستورتوقفشکایتیادشدهرادر
============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  11524
Score:  1.0
Title:  ماجرای تذکر رهبر انقلاب به گزارش مجلس چه بود؟/ حمایت پدرانه از دانشجویان تا حفظ حرمت قانون
URL:  https://www.farsnews.ir/news/14000812000656/ماجرای-تذکر-رهبر-انقلاب-به-گزارش-مجلس-چه-بود-حمایت-پدرانه-از
In doc:

کهازاقداماینفردشکایتخواهدکرد،اماهیچوقت
============================================================
```

```
doc_id:  11741
Score:  1.0
Title:  تاریخ آزمون دوره 49 دستیاری تخصصی پزشکی تصحیح شود
URL:  https://www.farsnews.ir/news/14000809000238/تاریخ-آزمون-دوره-49-دستیاری-تخصصی-پزشکی-تصحیح-شود
In doc:

شدهازشعبه45درشکایتمرتبطباآزمون48
============================================================
============================================================
شده،نهتنهابرایطرفشکایتلازمالاجراشدهبلکهمکلف
============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  11830
Score:  1.0
Title:  روحانی و جهانگیری متخلفان بزرگ یک پرونده/ ماجرای بورسیه‌ها چه بود؟
URL:  https://www.farsnews.ir/news/14000726000151/روحانی-و-جهانگیری-متخلفان-بزرگ-یک-پرونده-ماجرای-بورسیه-ها-چه-بود
In doc:

ومشکلدارخبرداد.*شکایتدانشجویانبورسیهازوزارت
============================================================
============================================================
قضائیبخواهندازاینوزارتخانهشکایتکنند،امامصوبه630
============================================================
============================================================
شورایعالیانقلابفرهنگیاجازهشکایتبورسیه‌هادرمحاکمقضائی
============================================================
============================================================
```

```
Score:  1.0
Title:  عزیزی: صندلی ریاست بیشتر از همه برای عزیزیخادم لرزان بود/ اسکوچیچ می تواند به فوتبال ما کمک کند
URL:  https://www.farsnews.ir/news/14001207000863/عزیزی-صندلی-ریاست-بیشتر-از-همه-برای-عزیزیخادم-لرزان-بود-اسکوچیچ-می
In doc:

رگیدکیزااهمیترمتسمتیکاشونایرالابتوفریخایشاوح
============================================================
============================================================
دنتفگگیمیاهرودرد.دنکیمتیاکش،دروخیمتسکشهکیمیتره
============================================================
============================================================
لبجنمیارب،دنرادتیاکشمهزاهمهکناریا
============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  1342
Score:  1.0
Title:  جنجال در انتقال بازیکن ماشین سازی به خیبرخرم آباد/ ماجرای عجیب مُهر باشگاه تبریزی و سرپرست برکنار شده
URL:  https://www.farsnews.ir/news/14001206000543/جنجال-در-انتقال-بازیکن-ماشین-سازی-به-خیبرخرم-آباد-ماجرای-عجیب-مُهر
In doc:

هبرگاامادسرنتیاکشهبراکاتددرگربنکیزاب
============================================================
============================================================
لاورقبطارنامدوختیاکشادرفاعطق،دنشابهتشادرارصا
============================================================
============================================================
```

```
doc_id:  139
Score:  1.0
Title:  تقدیر مربی تیم فوتبال خلیج فارس از ماجدی بابت رسیدگی به شائبه تبانی در لیگ جوانان
URL:  https://www.farsnews.ir/news/14001222000297/تقدیر-مربی-تیم-فوتبال-خلیج-فارس-از-ماجدی-بابت-رسیدگی-به-شائبه-تبانی-در
In doc:

یدجامداشریمروتسدابتیاکشنیایپرد.درکضارتعا
============================================================
============================================================
جیلخناوجلابتوفمیتتیاکشدروموردسرافیشزرورگانربخ
============================================================
============================================================
تبثزاسپ.مدیدرکتیاکشلابتوفنویسارفدهبیزابهجیتن
============================================================
============================================================
نویساردفتسرپرسیدجامداشریمتیاکشتبثزاسپ.مدیدرکتیاکش
============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  1584
Score:  1.0
Title:  عکس+شد قطری ه‌نارسر هژوس لالقتسا تساوخرد هب CAS در تسد
URL:  https://www.farsnews.ir/news/14001203000704/تسد-در-CAS-هب-تساوخرد-لالقتسا-لوس-هژ-هنارسر-یرطق-دش-سکع
In doc:
```

۴،۲)کریسمس

نتایج کریسمس به عنوان کویری تک کلمه ای سخت را در زیر مشاهده میکنید. مانند بخش قبل بدلیل استفاده از tfidf و نرمالسازی امتیاز همه ۱ شده است. همچنین زمان آن از کویری ((شکایت)) کمتر شده است زیرا کلمه نادرتری است و هنگام index elimination داک های بیشتری حذف میشوند.

بیشتر کویری های برگردانده شده مرتبط هستند ولی برخی از آنها ارتباط کمی دارند.

```
------------------------------------------------------------
------------------------------------------------------------

doc_id:  5483
Score:  1.0
Title:  مس چه زمانی به پارس برمی گردد؟
URL:  https://www.farsnews.ir/news/14001014000228/مس-چه-زمانی-به-پارس-برمی-گردد
In doc:

بهترین بازیکن جهان در ایام کریسمس به این ویروس مبتلا

============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  5921
Score:  1.0
Title:  الهامی: مقابل نفت مسجد سلیمان محکوم به پیروزی هستیم/استان سرسبز مازندران کی زمین تمرین هم ندارد
URL:  https://www.farsnews.ir/news/14001007000852/الهامی-مقابل-نفت-مسجد-سلیمان-محکوم-به-پیروزی-هستیم-استان-سرسبز
In doc:

نفت مسجد سلیمان اظهار داشت: کریسمس را به همه پیروان

============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  5926
Score:  1.0
Title:  مهاجم خارجی مس رفسنجان،4 کودک را به محل تحویل برگرداند +عکس
```

```
doc_id:  5926
Score:  1.0
Title:  مهاجم خارجی مس رفسنجان،4 کودک را به محل تحویل برگرداند +عکس
URL:  https://www.farsnews.ir/news/14001007000809/مهاجم-خارجی-مس-رفسنجان4-کودک-را-به-محل-تحویل-برگرداند-عکس
In doc:


------------------------------------------------------------
------------------------------------------------------------

doc_id:  5933
Score:  1.0
Title:  ستاره اسپانیایی: دیده کریسمس گوارادویال به واژی+عکس
URL:  https://www.farsnews.ir/news/14001007000739/ستاره-اسپانیایی-دیده-کریسمس-گوارادویال-به-واژی-عکس
In doc:

این طرح، تورس به دیده کریسمس پپ گوارادویال به واژی

============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  6120
Score:  1.0
Title:  کشتار در زرنگشاه فوتبال در آستانه سال جدید
URL:  https://www.farsnews.ir/news/14001005000143/کشتار-در-زرنگشاه-فوتبال-در-آستانه-سال-جدید
In doc:

پنج نفر در هنگام جشن کریسمس در کی زمین فوتبال

============================================================
============================================================
```

```
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  6148
Score:  1.0
Title:  ضربه تیم انگلیسی به لیون برای جذب آزمون/اورتون ای نیوکاسل؛ سردار راهی لیگ برتر می شود؟
URL:  https://www.farsnews.ir/news/14001004000708/ضربه-تیم-انگلیسی-به-لیون-برای-جذب-آزمون-اورتون-ای-نیوکاسل-سردار-راهی
In doc:

بودند فهرست بسیار طولان برای کریسمس بنویسند. به نظر می
============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  6697
Score:  1.0
Title:  جهانبخش: طارمی گفت برگردان من به چلسی بهتر بود/ کیروش روی ذهنیت ما خیلی تاثیر گذاشت/ نتایج اسکوچیچ کیفیت او را نشان میدهد
URL:  https://www.farsnews.ir/news/14000927000153/جهانبخش-طارمی-گفت-برگردان-من-به-چلسی-بهتر-بود-کیروش-روی-ذهنیت-ما
In doc:

که یک اتفاق جدید است. کریسمس در انگلیس متفاوت است
============================================================
============================================================

------------------------------------------------------------
------------------------------------------------------------

doc_id:  9413
Score:  1.0
Title:  معاون رئیس‌جمهور: جلسه نمایندگان اقلیت‌های دینی با آیت‌الله رئیسی برگزار می‌شود
```

۴/۳) لیگ برتر:

نتایج کویری ((لیگ برتر)) به عنوان کویری چند کلمه ای ساده را در زیر مشاهده میکنید.
مشاهده میشود که دیگر تمام امتیاز ها ۱ نیشتند زیرا کویری چند کلمه ای است و داک ها بر اساس امتیاز و شباهت به کویری مرتب شده اند.

کویری های برگردانده شده مرتبط هستند و در ارتباط با لیگ های برتر انگلیس یا ایران هستند.

```
doc_id:  2453
Score:  0.9999680009022884
Title:  تصمیمات سازمان لیگ فوتسال راهگشا یا بلای جان؟
URL:  https://www.farsnews.ir/news/14001119001101/تصمیمات-سازمان-لیگ-فوتسال-راهگشا-یا-بلای-جان
In doc:

لیگ، انتخابات هیات رئیسه سازمان لیگ فوتسال در سال ۱۳۹۸
============================================================
============================================================
رئیسه از بدو تاسیس سازمان لیگ فوتسال بود. با این
============================================================
============================================================
به آرش جابری دبیر سازمان لیگ سپرده شد. با انتخاب
============================================================
============================================================
فوتبال، جابری هم سکاندار سازمان لیگ فوتسال و هم دبیر
============================================================
============================================================
شد. در قانون و اساسنامه سازمان لیگ و فدراسیون فوتبال هم
============================================================
============================================================
```

```
Score:  0.9999420218152351
Title:  ربیعی: هوادار تیمی جسور است/ در 9 هفته آینده برای جاطلبی خودمان میجنگیم
URL:  https://www.farsnews.ir/news/14001220000217/ربیعی--هوادار-تیمی-جسور-است-در-9-هفته-آینده-برای-جاطلبی-خودمان
In doc:

بازی گذشته با دو تیم لیگ برتری بازی کرد و
============================================================
============================================================
جام حذفی توانستیم یک تیم لیگ برتری را حذف کنیم
============================================================
============================================================
تیم وجود دارد که شگفتی‌ساز لیگ باشیم. در 21 هفته
============================================================
============================================================
تلاش کرده‌ایم. هدفگذاری ما در لیگ برتر هدفگذاری دیگری بود
============================================================
============================================================
شده و 9 هفته باقیمانده لیگ برتر، سعی میکنیم از
============================================================
============================================================
تیمش و حضور آنها در لیگ برتر گفت: از شما
============================================================
============================================================
از هفته سوم ما در لیگ مشد از ناحیه کتف
============================================================
============================================================
```

```
doc_id:  4952
Score:  0.9999420218152351
Title:  احمدزاده: ملوان سرحال و شاداب است/حذف استقلال و پرسپولیس به نفع آینده فوتبال خواهد بود
URL:  https://www.farsnews.ir/news/14001021000138/احمدزاده-ملوان-سرحال-و-شاداب-است-حذف-استقلال-و-پرسپولیس-به-نفع-آینده
In doc:

تیم پس از سالها به لیگ برتر برگردد، عنوان کرد:
=============================================================
=============================================================
همه میدانند. تداوم صدرنشینی در لیگ یک کار بسیار سختی
=============================================================
=============================================================
و تلاش میکنند بهترین تیم لیگ یک باشند. احساس این
=============================================================
=============================================================
کلی فوتبال کشور و کیفیت لیگ برتر گفت: بازیهای لیگ
=============================================================
=============================================================
کیفیت لیگ برتر گفت: بازیهای لیگ برتر را خیلی دنبال
=============================================================
=============================================================
مثل 10 سال اول شروع لیگ حرفهای شاهد سلامت فوتبال
=============================================================
=============================================================
در آن سالها حداقل سلامت لیگ را میتوانستیم ببینیم. وی
```

```
doc_id:  5576
Score:  0.9999420218152351
Title:  ستاره تیم ملی گران ترین بازیکن لیگ روسیه شد+عکس
URL:  https://www.farsnews.ir/news/14001012000643/ستاره-تیم-ملی-گران-ترین-بازیکن-لیگ-روسیه-شد-عکس
In doc:

با اتمام نیم فصل اول لیگ روسیه به معرفی 10
=============================================================
=============================================================
10 بازیکن گران قیمت این لیگ پرداخت. سردار آزمون مهاجم
=============================================================
=============================================================
رده دوم گران ترین بازیکنان لیگ این کشور قرار داشت
=============================================================
=============================================================
میلیون یورو گران ترین بازیکن لیگ برتر روسیه شد. ستاره
=============================================================
=============================================================
اش به گران ترین بازیکن لیگ برتر روسیه تبدیل شد.
=============================================================
=============================================================
دوران طلایی ستاره ایرانی در لیگ روسیه با زنیت آغاز
=============================================================
=============================================================
موفق شد 3 بار قهرمان لیگ شد. همچنین آزمون یک
=============================================================
=============================================================
```

۴,۴)جیانی، اینفانتینو:

بعنوان کویری چندکلمه ای دشوار ((جیانی اینفانتینو)) در نظر گرفته شده است که اسم خاصی است. مشاهده میشود که سرعت جوابگویی آن از کویری چمد کامه ای ساده بیشتر است(به دلیل تاثیر index elimination) ولی امتیاز های آن کمتر از کویری ساده است زیرا تکرار کمتری دارد و امتیاز و شباهت کمتری دارد.

چندتا از کویری ها کاملا مرتبط هستند و در مورد فرد جستجو شده هستند ولی بیشتر آنها کاملا مرتبط نیستند.

```
------------------------------------------------------------

doc_id:  1140
Score:  0.7071067811865475
Title:  رفع تحریم میزبانی همگروه ایران/بغداد میزبان عراق-امارات شد
URL:  https://www.farsnews.ir/news/14001209000085/رفع-تحریم-میزبانی-همگروه-ایران-بغداد-میزبان-عراق-امارات-شد
In doc:


------------------------------------------------------------
------------------------------------------------------------

doc_id:  127
Score:  0.7071067811865475
Title:  رئیس فیفا تماشاگر بازی رقبای ایران می شود
URL:  https://www.farsnews.ir/news/14001222000684/رئیس-فیفا-تماشاگر-بازی-رقبای-ایران-می-شود
In doc:


------------------------------------------------------------
------------------------------------------------------------

doc_id:  2041
Score:  0.7071067811865475
Title:  75 درصد فوتبالیست‌ها مخالف جام جهانی 2 ساله
URL:  https://www.farsnews.ir/news/14001127000839/75-درصد-فوتبالیست‌ها-مخالف-جام-جهانی-2-ساله
In doc:
```

```
------------------------------------------------------------
------------------------------------------------------------

doc_id:  4580
Score:  0.7071067811865475
Title:  سخنگوی فدراسیون فوتبال: حرف بدی زده شده که اسکوچیچ ناراحت شود/اگر شرایط خطرناک نباشد اینفانتینو به ایران می‌آید
URL:  https://www.farsnews.ir/news/14001025000590/سخنگوی-فدراسیون-فوتبال-حرف-بدی-زده-شده-که-اسکوچیچ-ناراحت-شود-اگر
In doc:


------------------------------------------------------------
------------------------------------------------------------

doc_id:  4841
Score:  0.7071067811865475
Title:  واکنش فدراسیون فوتبال به انتقاد وزیر ورزش و بلیط 32 هزار یورویی یک شرکت آلمانی
URL:  https://www.farsnews.ir/news/14001022000544/واکنش-فدراسیون-فوتبال-به-انتقاد-وزیر-ورزش-و-بلیط-32-هزار-یورویی-یک-شرکت
In doc:


------------------------------------------------------------
------------------------------------------------------------

doc_id:  4910
Score:  0.7071067811865475
```

```
------------------------------------------------------------
------------------------------------------------------------

doc_id:  6370
Score:  0.7071067811865475
Title:  عزیزیخادم:اسکوچیچ با قدرت به کارش ادامه میدهد/داوران ما VAR را آموزش نبینند از چرخه حذف میشوند
URL:  https://www.farsnews.ir/news/14001001000441/عزیزیخادم-اسکوچیچ-با-قدرت-به-کارش-ادامه-میدهد-داوران-ما-VAR-را-آموزش
In doc:


------------------------------------------------------------
------------------------------------------------------------

doc_id:  6449
Score:  0.7071067811865475
Title:  ترفند مالی فیفا برای برگزاری جام جهانی هر دو سال یکبار/ 19 میلیون دلار در انتظار هر فدراسیون
URL:  https://www.farsnews.ir/news/14000930000060/ترفند-مالی-فیفا-برای-برگزاری-جام-جهانی-هر-دو-سال-یکبار-19-میلیون-دلار
In doc:


------------------------------------------------------------
------------------------------------------------------------
```