محمد مهدی نظری – ۹۹۳۱۰۶۱

در ساخت شاخص مکانی از لینک زیر کمک گرفته شده :

https://www.geeksforgeeks.org/python-positional-index/

برای نگه داری توکن ها از یک دیکشنری استفاده میکنیم که key آن همان توکن و value آن لیستی از تعداد تکرارهای آن در کل داکیومنت ها و دیکشنری از موقعیت تکرار آن در هر داکیومنت است : دیکشنری دوم کلید برابر DocID و مقدار برابر لیستی از پوزیشن هایی است که آن توکن در داکیومنت آمده است(برای هر داکیومنت یک دیکشنری موجود است).

مراحل ساخت Positional Index در کد بصورت خط به خط با کامنت توضیح داده شده است. تابع فعلی فقط برای یک فایل تکست نوشته شده پس فقط یک داکیومنت داریم که ID آن برابر ۰ در نظر گرفته شده اما در گام بعدی پروژه، تابع برای کار با فایل json بروزرسانی خواهد شد و داکیومنت های مختلف بررسی میشوند.

برای استفاده ازتابع ابتدا باید متن را پیش پردازش کنیم یعنی نرمال سازی و حذف کلمات پرتکرار و ریشه یابی انجتم شده و توکن ها استخراج میشوند و متن نهایی به عنوان آرگومان به تابع تحویل داده میشود. خروجی برنامه را با متن ورودی

```
"بارها و بارها از نخبگان و دلسوزان نظام این جمله بیان شده که قبل از وقوع
انقلاب بلکه از هزاران سال پیش مردم ایران به پوشش و حجاب ارج می‌نهادند و در حال
حاضر این مسئله همچون رفتار رضا شاه میشه .خوبه که داریم می رویم به سمت حجاب
اختیاری و نظام مورد علاقه ما مثل رضا شاه نمی شه".
```

ملاحظه میکنیم :

```
"C:\Users\MMNazari1380\PycharmProjects\Information Retrival\venv\Scripts\python.exe" "C:/Users/MMNazari1380/PycharmProjects/Information Retrival/Phase1
.py"
{'بار': [2, {0: [0, 2]}], 'و': [5, {0: [1, 4, 19, 23, 38]}], 'نخبگ': [1, {0: [3]}], 'دلسوز': [1, {0: [5]}], 'نظا': [2, {0: [6, 39]}], 'جمله': [1, {0:
[7]}], 'بیان': [1, {0: [8]}], 'شده': [1, {0: [9]}], 'قبل': [1, {0: [10]}], 'وقوع': [1, {0: [11]}], 'انقلاب': [1, {0: [12]}], 'بلکه': [1, {0: [13]}], 'هزار':
[1, {0: [14]}], 'سال': [1, {0: [15]}], 'پ': [1, {0: [16]}], 'ایر': [1, {0: [17]}], 'مرد': [1, {0: [18]}], 'پوش': [1, {0: [20]}], 'حجاب': [2, {0:
36]}], 'ارج': [1, {0: [21]}], 'می': [2, {0: [22]}], 'نهادند': [1, {0: [24]}], 'حال': [1, {0: [25]}], 'حاضر': [1, {0: [26]}], 'مسئله': [1, {0: [27]}],
'همچون': [1, {0: [28]}], 'رفتار': [1, {0: [29, 44]}], 'رضا': [2, {0: [30, 45]}], 'شاه': [2, {0: [31]}], 'میشه': [1, {0: [32]}], 'خوبه': [1, {0: [33]}],
'دار': [1, {0: [34]}], 'رو': [1, {0: [35]}], 'سم': [1, {0: [37]}], 'اختیار': [1, {0: [40]}], 'مورد': [1, {0: [41]}], 'علاقه': [1, {0: [42]}], 'ما': [1,
{0: [43]}], 'مثل': [1, {0: [46]}], 'نمی': [1, {0: [48]}]}

Process finished with exit code 0
```