

YAZILIM MİMARİSİ

Seyyid Ahmed Doğan - 14011038

Mehmet Seda Uçucu – 13011701

Faruk Aşçı – 15011043

Muhammed Melih Durmaz – 14011071

Mustafa Bodur – 15011026

Ahmet Onur Akman – 16011059

Mehmet Emin Hatipoğlu – 14011027

Proje Tanımı

Aranan kelimeyi daha önce indekslenmiş URL'lerde arayarak alaka sırasına göre kullanıcıya döndüren bir sistem tasarlanacaktır. Sistem, her URL'nin içeriğini dairesel şekilde çevirerek her kelime için bir skor belirleyecektir. Bu skorlar, aranan kelimeye göre her URL'nin skorunu belirleyecektir. Her URL'nin skoru belirlendikten sonra bu skorlara göre sıralı bir halde URL'leri döndürecektir.

İşlevsel Gereksinimler

- Dışarıdan kullanıcı tarafından aranacak sayfalar belirlenip URL bilgileri verilmelidir.
- Aranacak anahtar kelimeler indeksleriyle belirlenmelidir.
- Sayfalarda geçen cümlelerin kelimeleri sırasıyla indekslenmelidir. İlk kelime 0. indis, ikinci kelime 2. indis, 3. kelime 3. indis, şeklinde indekslenmelidir.
- Aramalar sadece bir kelime ile yapılmalıdır.
- Girilen kelimeye olan yakınlığına göre kullanıcıya sıralı URL, içerik ve skor bilgisi döndürülecektir.

İşlevsel Olmayan Gereksinimler

- Python 3.x yazılım dilini destekleyen tüm işletim sistemlerinde çalışabilmelidir.
- Dosya boyutu 20 MB'yi geçmemelidir.
- Girilen kelimeye göre tüm sonuçları döndürmelidir.
- Yeni özelliklerin eklenebilmesi için esnek ve modüler bir tasarım yapılması gerekmektedir.
- Kullanıcının isteğini tatmin edilecek sürede cevap vermelidir.

Gereksinim Tanımları

Program Girdileri

Sistem, indekslenecek URL listesini bir “txt” dosyasından alacaktır. Dosyadaki her satırdaki veri formatı şu şekilde olmalıdır:

içerik; URL;

Ara Çıktılar

Her satırdaki içerik puanlaması yapıldıktan sonra bir veri tabanına her kelime, içinde bulunduğu içerik, bağlı olduğu URL ve kelimenin içerik içerisinde ne kadar başta olduğu skoru ile birlikte bir veri tabanına aşağıdaki şekilde kaydedilmelidir:

[{indeks, {içerik, URL}}]

Program Çıktıları

Aranan kelimeye göre veri tabanından URL’ler çekildikten sonra aşağıdaki formatta bir liste döndürmelidir:

[{indeks,içerik, URL, skor}]

Burada skor, aranan kelimeye göre her URL’ye verilen puanı temsil etmektedir.

Kısıtlar

context Indexer::add_index (context: String, url: String)

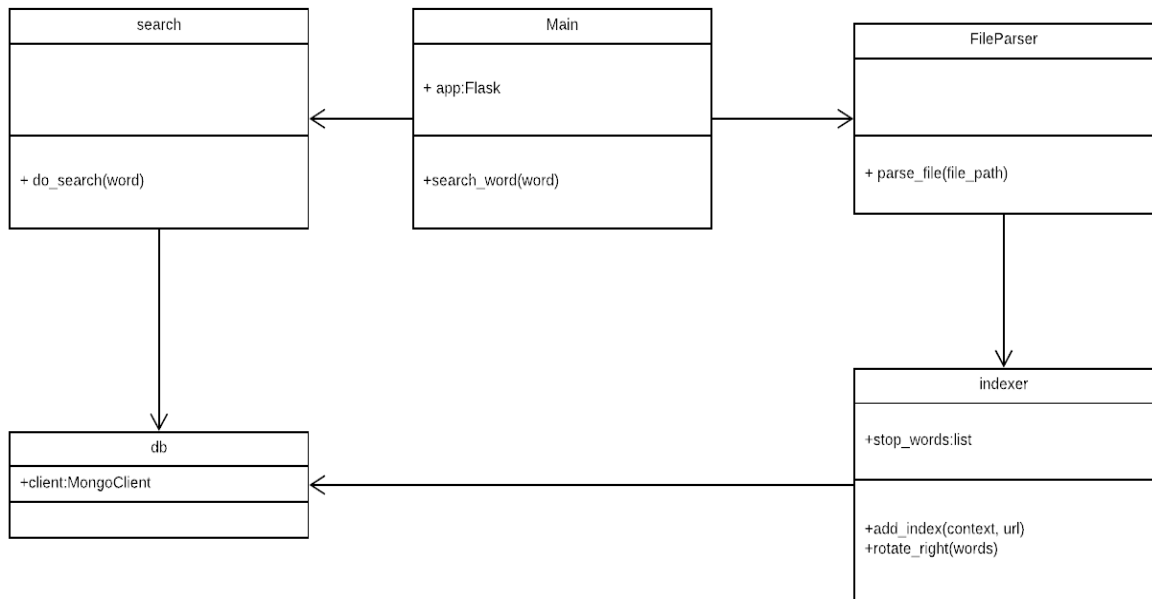
pre: stop_words -> size() > 0

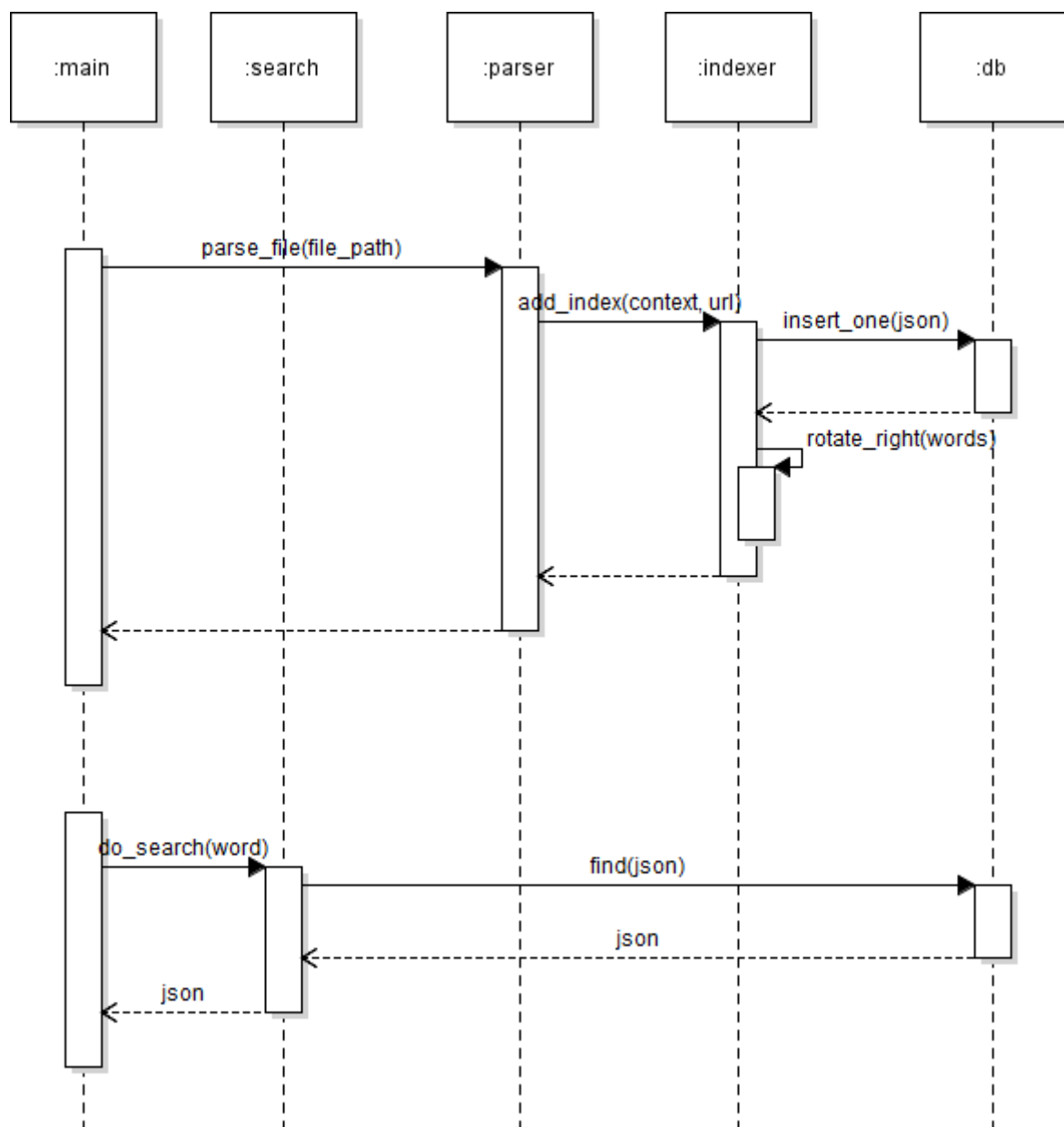
post: words = words@pre

context: Parser::parse_file(file_path: String)

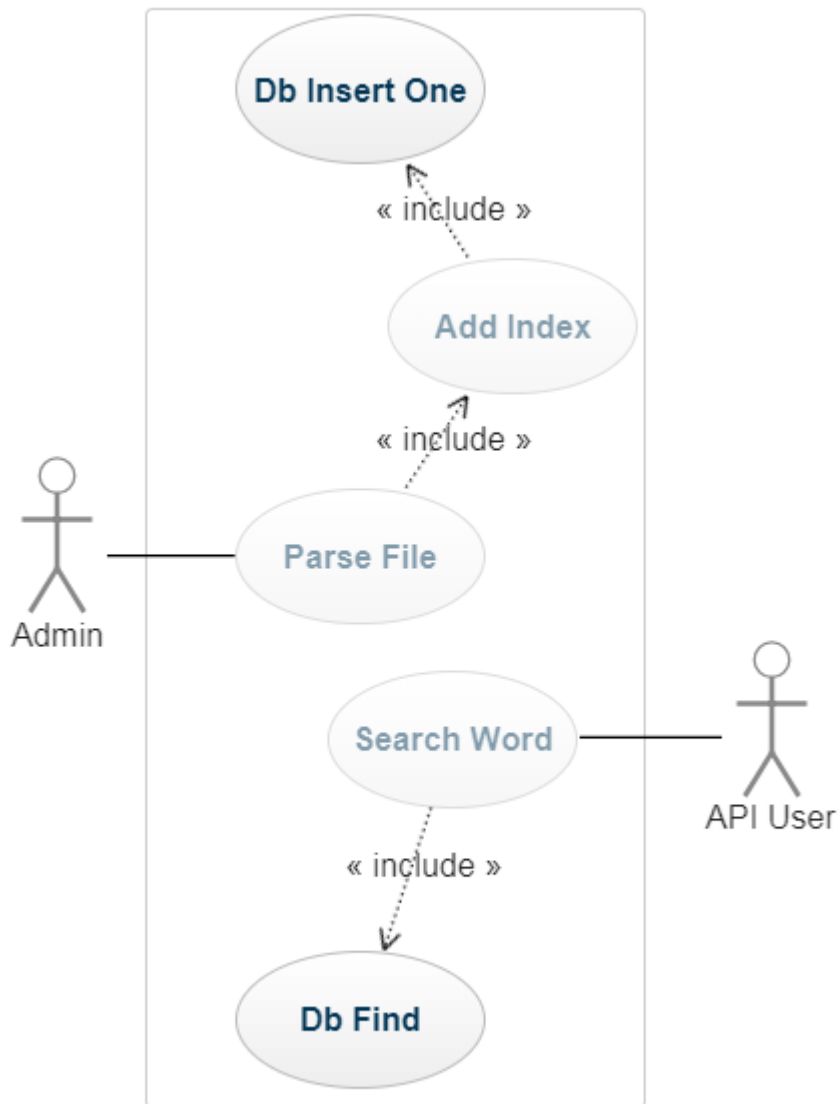
pre: file_path -> size() > 0

UML DİYAGRAMLARI





USE CASE DİYAGRAMI



Program Tanıtımı

Programda belirli bir kurala göre verilmiş kelime toplulukları ve bu kelime topluluklarına karşılık gelen URL'ler indekslenerek bir veri tabanına yerleştirilmiştir.

İndeksleme işlemi için ilk önce verilen kelime toplulukları sık kullanılan ve önemli olmayan kelimelerden arındırılmıştır (örneğin “the”, “and”, “a”...). Daha sonra kalan kelimelere indeks verilmesi işlemi yapılmıştır. İndeks verilirken ilk kelimeye 0, ikinci kelimeye 1... şeklinde indeks verilerek her kelimedenden sonra indeks verilen kelime sona atılmış ve sonraki kelime indekslenmiştir. İndekslenen kelimeler hangi kelime topluluğuna ait olduğu, kaç puan değerinde olduğu değerleri ile birlikte daha sonra kullanılmak üzere veri tabanında saklanmıştır.

Örnek Çıktılar

“hello”, “mad”, “lord” ve “man” kelimelerinin aramalarından dönüş örnekleri gösterilmiştir.

POST: ~/search/hello

```
[  
]
```

POST: ~/search/mad

```
[  
  {  
    "context": "Mad Max: Fury Road ",  
    "url": " https://www.imdb.com/title/tt1392190/ ",  
    "point": 0  
  }  
]
```

POST: ~/search/lord

```
[  
  {  
    "context": "The Lord of the Rings: The Return of the King ",  
    "url": " https://www.imdb.com/title/tt0167260/ ",  
    "point": 0  
  },  
  {  
    "context": "The Lord of the Rings: The Fellowship of the Ring ",  
    "url": " https://www.imdb.com/title/tt0120737/ ",  
    "point": 0  
  },  
  {  
    "context": "The Lord of the Rings: The Two Towers ",  
    "url": " https://www.imdb.com/title/tt0167261/ ",  
    "point": 0  
  }  
]
```

POST: ~/search/man

```
[  
  {  
    "context": "The Man in the High Castle ",  
    "url": " https://www.imdb.com/title/tt1740299/ ",  
    "point": 0  
  },  
  {  
    "context": "Iron Man ",  
    "url": " https://www.imdb.com/title/tt0371746/ ",  
    "point": 1  
  },  
  {  
    "context": "One Punch Man: Wanpanman ",  
    "url": " https://www.imdb.com/title/tt4508902/ ",  
    "point": 2  
  }  
]
```