

Kocaeli Üniversitesi
Bilgisayar Mühendisliği Bölümü
BLM 307: Yazılım Lab. I, 2017-2018 Güz
Proje III

YAM: Yerli Arama Motoru

Proje Başlangıç Tarihi 20 Kasım 2017
Proje Teslim Tarihi: 15 Aralık 2017

Amaç: Google'dan daha iyi sonuç veren bir arama motoru yapmak.

Dil: Serbest

1 Anahtar kelime saydırma (%10+)

Projenin içeriği ve adımları aşağıda gösterilmiştir:

Verilen bir anahtar kelimenin ve bir URL için:

- Anahtar kelimenin URL içeriğinde (URL'in gösterdiği sayfa içeriğinde) kaç defa yer aldığını bulunuz.

Çıktı: Bu aşamada projenin web sitesindeki bir sayfada (örn: asama1.php, asama1.asp, ...) bir anahtar kelime girilecek bir alan ve bir URL girilecek bir alan verilecektir. Sonrasında verilen URL'de verilen kelimenin kaç defa yer aldığı bilgisi hesaplanarak yazdırılacaktır.

(Örnek anahtar kelime: “güvenlik” , örnek URL: <https://www.btk.gov.tr/tr-TR/Sayfalar/SG-SIBER-GUVENLIK-KURULU>)

2 Sayfa (URL) Sıralama (%30+)

Verilen bir anahtar kelime kümesi ve bir URL kümesi için,

- Anahtar kelimelerin içeriklerde yer alma sayısına dayalı bir skor formülü tanımlayınız.
 - Örneğin, direkt kelimelerin URL’de geçme sayısını toplayarak bir skarlama yapalım. Bu durumda 1. kelimenin 100 defa, 2. ve 3. Kelimenin 3’er defa geçtiği bir URL, her bir kelimenin 10’ar defa geçtiği bir URL’den daha fazla skor alacaktır. Yani, bu skarlama formülü iyi sonuçlar vermeyecektir.
- URL’leri skoruna göre sıralayınız.
- Her bir URL için (bir web sayfası), sıralamasını, puanını ve her bir anahtar kelimenin kaç kez yer aldığı sayısını yazdırınız.

Örnek anahtar kelime kümesi: “ulusal, siber, güvenlik”, örnek URL kümesi:

<https://www.btk.gov.tr/tr-TR/Sayfalar/SG-SIBER-GUVENLIK-KURULU>,

<http://www.udhb.gov.tr/h-12-siber-guvenlik.html>,

<http://sibertehtit.com/ulusal-siber-guvenlik-stratejisi-ve-eylem-plan/>

3 Site Sıralama (%50+)

Verilen bir web sitesi kümesi ve bir anahtar kelime kümesi için,

- URL’de ve tüm alt URL’lerinde anahtar kelime yer alma sayılarına dayalı bir skor formülü tanımlayın (derinlik 3, ana sayfa=1, ana sayfadan linklenmiş bir sayfa:2, ana sayfadan linklenmiş bir sayfadan linklenmiş bir sayfa:3).
- Web sitelerini anahtar kelimelerin yer alma sayılarına göre sıralayınız (tüm alt URL’leri dahil)
- Her URL için (bir web sitesi), sırasını, skorunu, alt URL’lerin ağaç yapısını ve her düğümdeki her bir anahtar kelimenin yer alma sayısı ile birlikte yazdırın.

(Örnek anahtar kelime kümesi: “ulusal, siber, güvenlik” , örnek web sitesi kümesi: “<https://www.btk.gov.tr/>, <http://www.udhb.gov.tr/>, <http://sibertehtit.com/>)

4 Semantik Analiz (%80+)

- Verilen web siteleri içerisinde anahtar kelimelerle semantik olarak alakalı kelimeler olabilir. Örneğin, “ulusal” yerine “milli” kelimesi yer alabilir. Bu kelimelere “alakalı anahtar kelimeler” diyelim.
- Alakalı anahtar kelimeleri bulun ve yazdırın.
- Yinelemeli olarak 3. aşamadaki analizi yapın.

5 Çalışmanın Çevrimiçi Yayınlanması (%100+)

- Arama motorunuz için bir web sitesi geliştirin.

6 Ödev Teslimi

- Proje teslim gününde rapor (hard copy) teslim etmeniz gerekmektedir.
- Teslim tarihinden sonra getirilen projeler kabul edilmeyecektir.
- Proje ile ilgili sorularınız için Yrd. Doç. Dr. Süleyman Özarslan, Arş. Gör. Süleyman Eken veya Arş. Gör. Furkan Göz'e ulaşabilirsiniz.
- Demo tarihleri daha sonra duyurulacaktır.
- Demo sırasında algoritma, geliştirdiğiniz kodun çeşitli kısımlarının ne amaçla yazıldığı ve geliştirme ortamı hakkında sorular sorulabilir.
- Kullandığınız herhangi bir satır kodu açıklamamız istenebilir. Demo sırasında geliştirdiğiniz kod farklı girdiler için çalıştırılacaktır.
- Yazılım Laboratuvarı I Proje III için https://piazza.com/kocaeli_university/fall2017/blm307 adresinde sorularınızı iletebileceğiniz bir ders açılmıştır (erişim kodu: blm307). Aynı zamanda proje ile ilgili belgeler de buradan paylaşılacaktır.
- Yeniden proje grubu oluşturulmayacaktır. Eski grup arkadaşımızla projeyi yapmanız gerekmektedir.