**YERLİ ARAMA MOTORU “ABULLABUT”**

Damla MUTLU – Hasibe Nur ÇAPRAZ

Bilgisayar Mühendisliği Bölümü

Kocaeli Üniversitesi

damlamutlu123@gmail.com – [hnurcapraz@gmail.com](mailto:hnurcapraz@gmail.com)

1. **Problem tanımı:**

Projede istenen basit bir arama motoru yapımıdır. Arama motorunun yapacağı işlemler bellidir ve bunların aşama aşama gerçekleşmesi beklenmektedir. Birinci aşama “anahtar kelime saydırma” işlemidir. Bu aşamada verilen anahtar kelimenin, verilen URLde kaç kere geçtiğinin hesaplanması gerekmektedir. İkinci aşama ise “sayfa (url) saydırma” işlemidir. Bu aşamada verilen anahtar kelime kümesi ve URL kümesi için işlem gerçekleşmelidir. Anahtar kelimelerin içerikte yer alma sayısına dayalı bir skor formülü oluşturmamız ve skora göre URLleri yazdırmamız gerekmektedir. Aynı zamanda Her bir URL için (bir web sayfası), sıralamasının puanını ve her bir anahtar kelimenin kaç kez yer aldığının bilgisini ekrana yazdırmamız gerekmektedir. Üçüncü aşama ise “site sıralama” işlemidir. Bu aşamada verilen anahtar kelime ve URL kümesi için URLde ve tüm alt URL’lerinde anahtar kelime yer alma sayılarına dayalı bir skor formülü oluşturmamız istenmiştir. Bu işlemin üçüncü derinliğe kadar yapılması yeterlidir. Daha sonra URLleri anahtar kelimelerin yer alma sayılarına göre sıralamamız istenmiştir. Her URL için sırasını, skorunu, alt URL’lerin ağaç yapısını ve her düğümdeki her bir anahtar kelimenin yer alma sayısı ile birlikte yazdırmamız istenmektedir. Dördüncü aşama ise “semantik analiz” işlemidir. Bu aşamada URLdeki anahtar kelimemizle semantik olarak alakalı kelimeleri bulmamız ve tekrarlı olarak üçüncü aşamayı yapmamız istenmektedir. Beşinci ve son aşamada ise çalışmamızın çevrim içi yayınlanması istenmektedir. Proje Notepad++’ta PHP dilinde yapılmıştır.

1. **Yapılan Araştırmalar:**

Projeye başlarken çok sayıda sorunla karşılaştık. Hangi dilden başlayacağımızı, hangi ortamda geliştireceğimizi bilmiyorduk. Çeşitli araştırmalar yaptık, bu konuda videolar izledik. PHP dilinde yapmanın mantıklı olduğuna karar verdik. Notepad++’ı seçme sebebimiz ise PHP dahil olmak üzere çok sayıda dil desteğinin bulunması oldu.

Karşılaştığımız asıl sorunlara gelecek olursak, başlangıçta istenen her aşama, basamak bizim için bir sorundu. Örneğin ekrana nasıl buton koyacağımızı, textbox koyacağımızı bilmiyorduk. Bu konuda bir video serisi izledik. Böylece buton, textbox nasıl eklenir, yeni form nasıl açılır, başlık nasıl eklenir bunları öğrendik. Bu sorunlara çözüm bulunca büyük aşama kaydetmiş olduk. İlerleyen zamanlarda ise Türkçe karakter sorunu çıktı. Birinci aşama gerçekleşiyordu ancak Türkçe karakter içeren bir kelime girince içermeyen hali varsa URLde onu da saymalıydı. Ya da tam tersi işlem gerçekleşmeliydi. Bunun için anahtar kelimedeki ve içerikteki tüm Türkçe karakterleri İngilizce karşılıkları ile değiştirdik. Sorun böylece çözülmüş oldu. Diğer bir sorun ise büyük-küçük harf durumuydu. Örneğin aradığımız kelime “Bilgisayar” ama program yalnızca ilk harfi büyük olanları sayıyor “bilgisayar”, BİLGİSAYAR” vb şekilde olanları saymıyordu. Bunu çözmek için de anahtar kelimenin ve içeriğin tüm harflerini küçük harf yaptık. Sorun da böylece ortadan kalkmış oldu. Bunların yanında PHP diline özel “count, explode, file\_get\_contents, file\_get\_contents” vb kavramları araştırdık ve sık kullandığımız çok yardımı olan kavramları öğrendik. Ancak karşılaştığımız en büyük sorun skor hesaplayıp URLleri sıralamak oldu. Skor algoritmasını oturtmak günlerimizi aldı. Bu süreçte işimizi kolaylaştıracak kaynakları inceledik, video izledik ancak sonunda kendi mantığımız ile bir algoritma oluşturduk. Sıralama olayına geldiğimizde ise skorları diziye atıp sıralamayı denedik. Klasik C’deki sıralama algoritmalarını uyarlamaya çalıştık ancak pek başarılı olamadık. Sonunda kendimiz bir algoritma oluşturup sıralama işlemini ona göre yaptık.

1. **Yazılım mimarisi:**
   1. *PHP*

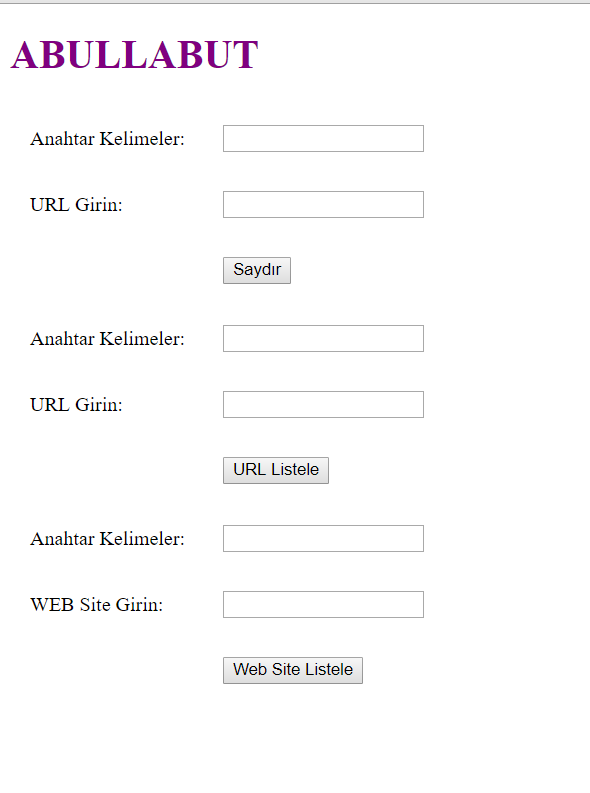
Hypertext Preprocessor (Türkçe: Üstünyazı Önişlemcisi) (Aslen: Personal Home Page - Kişisel Ana Sayfa) , internet için üretilmiş, sunucu taraflı, çok geniş kullanımlı, genel amaçlı, HTML içerisine gömülebilen betik ve programlama dilidir.[1] <https://tr.wikipedia.org/wiki/PHP>

* 1. *Notepad++*

[GNU](https://wmaraci.com/nedir/gnu) lisansı ile ücretsiz olarak dağıtı[lan](https://wmaraci.com/nedir/lan) gelişmiş bir [kod](https://wmaraci.com/nedir/kod) düzenleyici yazılımıdır. Programlama işlemleri için [yazılım](https://wmaraci.com/nedir/yazilim) dünyasının sunduğu en ideal düzenleme aracı olan [Notepad](https://wmaraci.com/nedir/notepadplusplus)++, birçok programlama dilini desteklemektedir. Bunlar arasında; [C](https://wmaraci.com/nedir/cplusplus), C++, [Java](https://wmaraci.com/nedir/java), C#, [XML](https://wmaraci.com/nedir/xml), [HTML](https://wmaraci.com/nedir/html), [PHP](https://wmaraci.com/nedir/php), [CSS](https://wmaraci.com/nedir/css) ve [ASP](https://wmaraci.com/nedir/asp) gibi sıklıkla kullanılanlar [da](https://wmaraci.com/nedir/da) bulunmaktadır.[2] <https://wmaraci.com/nedir/notepadplusplus>

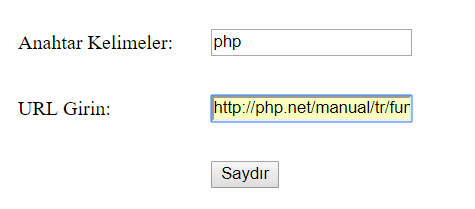
1. **Genel yapı**

Projemizi yaparken localhost’ta çalıştık daha sonra onu internet sitesi haline getirdik. Sitemizin adı aabullabut.tk. sitemize ilk girdiğimizde üç kısım görüyoruz.

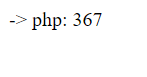


Şekil 1:Ana ekran

İlk kısımda anahtar kelimemizi ve URLi girip saydır butonuna tıklıyoruz. Burada anahtar kelimenin URLde kaç defa geçtiğini buluyoruz.



Şekil 2:Birinci Aşama



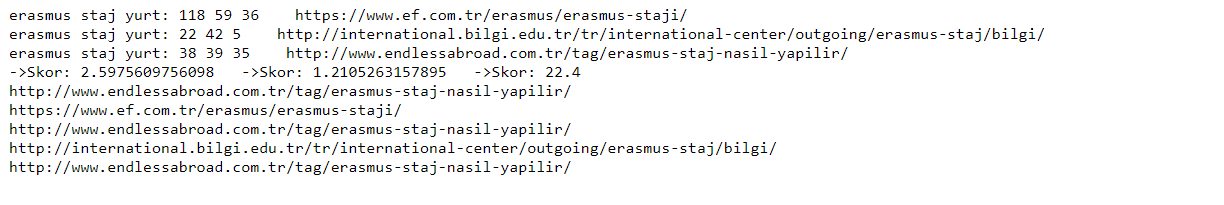
Şekil 3:Ekran çıktısı

Bu işlemi yaparken ilk önce kelimemizin uzunluğun alıyoruz for döngüsü ve substr yardımıyla URL içeriğinde dolaşıyoruz harf harf geziyoruz ve anahtar kelimemiz içerikteki parçaya eşit olunca sayı değişkenimiz artıyor. Bunu da ekrana yazdırıyoruz. Aynı zamanda büyük-küçük harf ve Türkçe karakter sorunu çıkmaması adına anahtar kelimenin ve içeriğin tüm harflerini küçük harfe çeviriyoruz. Bunun için strtolower kullanıyoruz. Türkçe karakter için de içeriğin ve anahtar kelimenin eğer varsa Türkçe karakterlerini İngilizce karaktere çeviriyoruz arka planda kod Türkçe karakterle uğraşmıyor böylece ikisi de sorun çıkartmamış oluyor.

İkinci kısımda ise URL sıralama işlemini yapıyoruz. Anahtar kelime kümemizi ve URL kümemizi giriyoruz.



Şekil 4:İkinci Aşama



Şekil 5: URL Sıralama

Burada URL’lerimizi araya “\*” koyarak yazıyoruz. Bu yüzden ilk olarak sıralama işlemi için explode yardımı ile URL’lerimizi “\*” gördükçe bölüyoruz ve bunların sayısını tutuyoruz. Daha sonra aynı bölme işlemini anahtar kelimede boşluk gördükçe yapıyoruz ve bu sayıyı da tutuyoruz. Birinci aşamada yaptığımız kelime sayma işlemini her kelime ve URL için ayrı ayrı yapıyoruz daha sonra bunlara bağlı olarak bir skor elde ediyoruz ve skoru büyük çıkan üstte olacak şekilde URL’lerimizi sıralıyoruz. Büyük-küçük harf ve Türkçe karakter sorunu çıkmaması adına anahtar kelimenin ve içeriğin tüm harflerini küçük harfe çeviriyoruz. Bunun için strtolower kullanıyoruz. Türkçe karakter için de içeriğin ve anahtar kelimenin eğer varsa Türkçe karakterlerini İngilizce karaktere çeviriyoruz arka planda kod Türkçe karakterle uğraşmıyor böylece ikisi de sorun çıkartmamış oluyor. Bu kısma semantik analizi de dahil ettik. Anahtar kelime kümesinde bulunan kelimelerin eş anlamları da txt dosyamızda aranıyor ve sayma ilmeine dahil oluyor.

Üçüncü kısımda ise derinliği baz alarak URL sıralamamız isteniyor ancak bizim arama motorumuz bu işlemi tam olarak gerçekleştiremiyor yalnızca birinci derinlikteki linkleri çekebiliyoruz. Büyük-küçük harf uyumu ve Türkçe karakter uyumu işlemleri bu basamakta da gerçekleşiyor.



Şekil 6:ÜçüncüAşama



Şekil 7:Derinlik

Son olarak da arama motorumuzu web sitesi haline getirip yayınladık. Sitemizin adı aabullabut.tk. Bu işlemi freehosting ile yaptık ilk önce bir domain alıp sonra hosting servisinden yararlandık. [3] <https://www.youtube.com/watch?v=rpwL69Piop0>

1. **Referaslar**

\*<https://tr.wikipedia.org/wiki/PHP>

\*<https://wmaraci.com/nedir/notepadplusplus>

\*<https://www.youtube.com/watch?v=rpwL69Piop0>

\*https://www.youtube.com/channel/UCO5Dm8YsOb9bOtBdIxmP7Rw