ARAMA MOTORU

Esra Çilek, Doğukan Berk Berber

Bilgisayar Mühendisliği Bölümü

Kocaeli Üniversitesi

[esra\_cilek@icloud.com](mailto:esra_cilek@icloud.com) , [dogukanberkberber@gmail.com](mailto:dogukanberkberber@gmail.com)

**Özet**

*Bize bu dönemde verilen üçüncü ödev arama motoru idi .Biz bu ödevi yapmak amacıyla ilk önce gereken çalışmaları yapmak için kendimize bir yol haritası çıkardık. Bu haritaya bağlı kalarak çalışmalarımızı yürüttük. Kısaca bahsetmem gerekirse;*

*Yaklaşık 5 -6 gün ödevi nasıl yapacağımızı hangi kaynaklardan yararlanabileceğimizi ve bir arama motorunu nasıl ve hangi yöntemlerle geliştirebileceğimizi araştırdık. Bu aşamayı atlatmak bizim için bir avantaj oldu, bu araştırma sonrası projemizi nasıl yapmamız gerektiği az çok kafamızda şekillenmeye başladı. Araştırma kısmını daha doğrusu arama motorunu nasıl yapılır? sorusunu kafamızda cevapladıktan sonra kodlama kısmına geçiş yaptık ve araştırmalarımızın ışığında kodlama kısmını yapmak daha kolay oldu ve böylece ödevimiz filizlenmeye başladı. Tabi bu kısımda daha önce C# dili ile görüntü işleme editörü ve sudoku çözme programı yaptığımız için bizim için kolaylık oldu ve bu dili geliştirme dili olarak kendimize seçtik.*

# Giriş

*Bu dönem bizlere verilen son ödev bir arama motoru geliştirmek. Projeye başlarken bir arama motoru nasıl yapılır, ödevimizin gerçek bir arama motoruyla arasındaki farklar nelerdir sorularından yola çıktık .Bize verilen ödev standart bir arama motoru ile farklılıklar göstermektedir. Bizlerden istenilen adımlardan kısaca bahsetmek gerekirse;*

*1-)Anahtar kelime saydırmak.*

*2-)Sayfa (URL) sıralamak.*

*3-)Site sıralamak.*

*4-)Semantik analiz gerçekleştirmek.*

*5-)Yapılan çalışmanın çevrimiçi olarak yayınlanması.*

*Bu adımlar ışığında arama motorumuzu inşaa ettik ve ortaya bazı aksilikler olsa dahi sonuçları kullanıcıya gösteren bir ödev yapmayı başardık.*

**Temel Bilgiler**

*Projemizi geliştirirken Visual Basic programını kullandık. Projeyi geliştirmede kullandığımız dil ise C# dır. Aynı zamanda geliştirdiğimiz proje Windows Form Application olarak geliştirilmiştir. Kullandığımız dilden biraz bahsetmek gerekirse;*

*Benzer şekilde C++ diline yeni eklentiler yapılarak ((C++)++) bir adım daha ileriye götürülmüş ve tamamen nesneye yönelik tasarlanmış****C#****dilinin isimlendirilmesinde, + karakterlerinin birbirlerine yakınlaşmış hali ve bir melodi anahtarı olan****C#****Major kullanılmıştır.*

# Proje İçeriği

*Projemizi incelemeyi birkaç alt başlıkta gerçekleştirebiliriz. Bunlar sırasıyla;*

* *KULLANICI TARAFINDAN GİRİLEN ANAHTAR KELİMENİN YİNE KULLANICI TARAFINDAN GİRİLEN URL DE KAÇ DEFA GEÇTİĞİNİ BELİRLEMEK*
* *KULLANICI TARAFINDAN GİRİLEN ANAHTAR KELİME KÜMESİNİN YİNE KULLANICI TARAFINDAN GİRİLEN URL KÜMESİNDE KAÇ ADET GEÇTİĞİNİ BELİRLEMEK VE VERİLEN URL LERİ SIRALAMAK*
* *KULLANICI TARAFINDAN GİRİLEN ANAHTAR KELİME KÜMESİNİN YİNE KULLANICI TARAFINDAN GİRİLEN WEB SİTE KÜMESİNDE KAÇ ADET GEÇTİĞİNİ BELİRLEMEK VE VERİLEN WEB SİTELERİNİ SIRALAMAK*
* *VERİLEN ANAHTAR KELİMELERİ VERİLEN WEB SİTELERİNDE SEMANTİK OLARAK ARATMAK*
* *YAPILAN ÇALIŞMANIN ÇEVRİMİÇİ OLARAK YAYINLANMASI*

1. **ANAHTAR KELİMENİN SAYDIRILMASI**

*Bu kısımda yaptığımız araştırmalar sonucunda ilk olarak PHP dilini kullanmaya karar verdik fakat çalışmalarımız sonucunda bu dilde bir ilerlemeye sağlayamadık ve C# diline dönüş yaptık. C# dili bize bu konuda yardımcı oldu. Anahtar kelime arama işlemini ilk olarak sitelerin kaynak kodları üzerinde gerçekleştirdik fakat piazza da sorulan soruların cevaplarını dikkate alarak yaptığımızın yanlış bizden istenilenin ise içerik de yani text dosyası içerisinde anahtar kelime araması yapılması olduğunu öğrendik. Bu kısımda ise;*

*WebClient ile istenilen sayfanın html kodunu çektik. Daha sonra Türkçe karakter sorunu ortadan kaldırdık. Sonrasında ise sayfa kaynağını değilde içerikde bulunan text dosyalarında saydırma işlemini gerçekleştirdik.Böylece cevabımız CTRL+F fonksiyonuyla aynı hale geldi.*

1. **SAYFA (URL) SIRALAMA**

*Bu kısımda bizden istenilen girilen url kümesi içerisinde verilen anahtar kelimeleri aramak ve bulunma sayısına göre bir skor sistemi oluşturarak bu url kümesini sıralamaktır.*

*Biz de skor sistemini gerçeklerken kendi oluşturduğumuz algoritmadan yararlandık. Algoritmadan kısaca bahsetmek gerekirse;*

*Girilen anahtar kelimelerin girilen urllerde kaçar adet geçtiğini bularak bunların aritmetik ortalamasını almaya dayalıdır. Alınan aritmetik ortalama kelime geçme sayılarından çıkarılarak bulunan sayıları topladık. Yani her kelimenin bir url için standart sapmalarını topladık. Toplandıktan sonra elde ettiğimiz toplamı aritmetik ortalamaya bölerek bir skor tanımladık. Skoru en küçük olan url en üste yazılmaktadır.*

1. **SİTE SIRALAMA**

*Bu kısımda bizden istenilen kullanıcının girdiği web sitelerini kullanıcının girdiği kelimelere göre sıralamamızdı.*

*Öncelikle bir web sitesi içindeki alt urllerini çekmek için bir class oluşturduk. Bu class yardımıyla her web sitesinin alt urllerini aldık. Bir web sitesinin her alt urli için bir önceki aşamada oluşturduğumuz skor algoritmasını kullanarak bir alt url için skor oluşturduk. Daha sonra bir web sitesinin alt url skorlarını topladık. Topladığımız skorları web sitesinin alt url sayısına bölerek bir orantı yakaladık. Bu orantı üçüncü aşamamız için yeni bir skor sistemi oluşturmamızı sağladı. Daha doğrusu ikinci aşamada oluşturduğumuz skor formülünü geliştirerek web sitelerini sıralamak için daha gelişmiş bir skor sistemi oluşturduk. Elde ettiğimiz orantılar sonucunda skoru en büyük olan girilen kelimelerle en alakalı site olmuş oldu.*

**D.SEMANTİK ANALİZ**

*Bu kısımda bizden istenilen görev;*

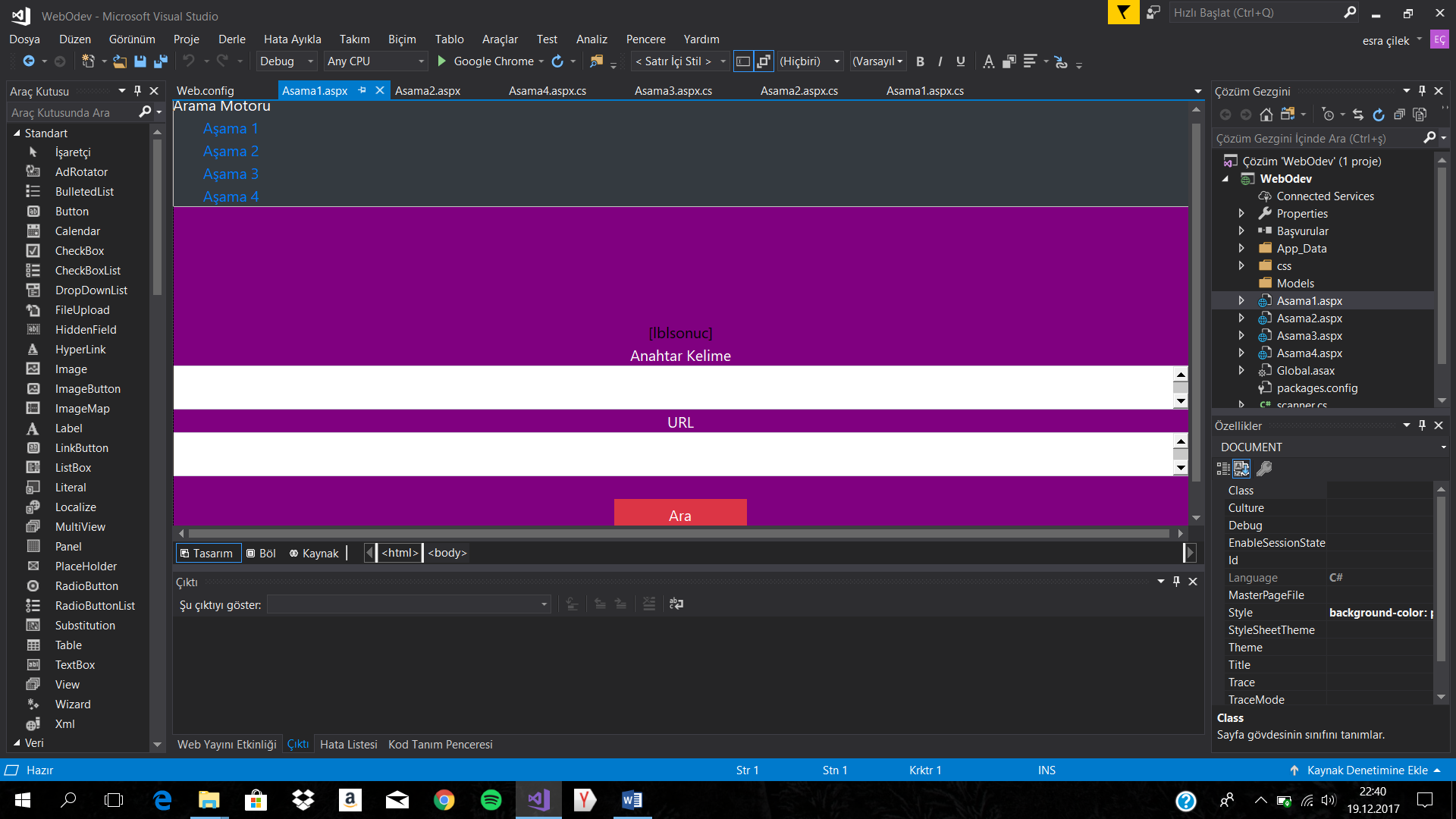
*Girilen anahtar kelimeyi verilen url lerde aramanın yanı sıra girilen anahtar kelimenin eş anlamlısını da bulup onu da verilen url de aramaktır. Bu aşamayı gerçekleştirirken eş anlamlı kelimelerin bulunduğu bir txt dosyası oluşturduk ve girilen anahtar kelimeleri bu dosyadan eşleme yaparak aratmaya başladık. Girilen anahtar kelimenin eş anlamlısını ilk önce dosyada bulma işlemini gerçekleştirdik ve daha sonra bu iki kelimeyi üçüncü aşamada olduğu gibi anahtar kelime aratma metodhu ile aratmaya başladık. Sonuç olarak girilen kelimenin eş anlamlısını da bulduğumuz için daha keskin ve daha doğru bir sonuç elde ederek sıralama işlemini gerçekleştirdik.*

**SONUÇ**

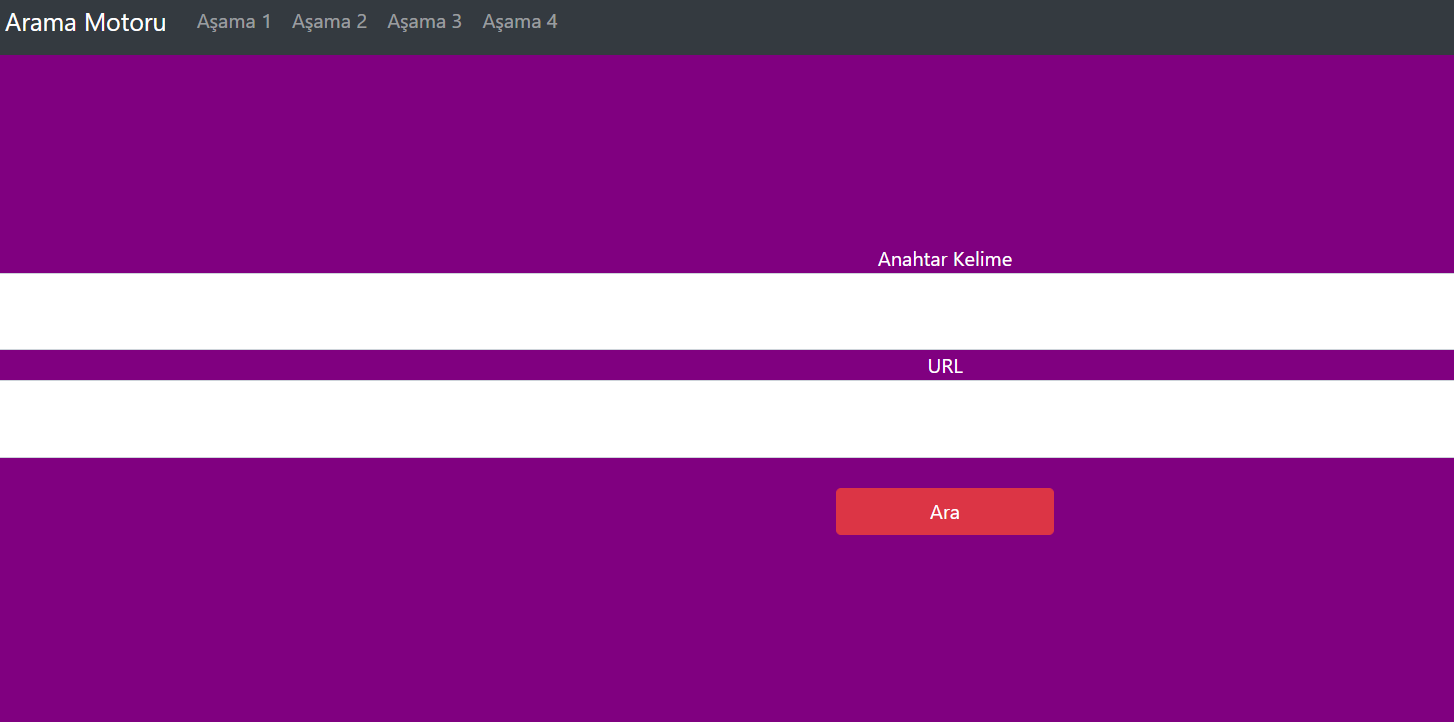
*Sonuç olarak biz projede C# kullanma bilgimizi geliştirdik. Bunun yanında ilk defa bu platformda ödev hazırladığımız için bu konuyla ilgili bilgi ve tecrübe sahibi olduk.Yeni kavramlarla karşılaştığımız için bunları kavramamız ve kullanmamız bizim için bir hayli zorlu olsa da bunun üstesinden araştırma ve ders notlarından yararlanma yollarıyla geldik. Sonuç olarak ortaya girilen anahtar kelimenin verilen url de kaç defa geçtiği bulan, anahtar kelime sıklığı ve alt url sıklığı arasında ilişki kurup verilen siteleri sıralayan bir program geliştirmeyi başardık.Tabi ki orijinal bir arama motorunun yanında bir çok eksiği olsa da bize verilen ödevden istenilenleri karşılamaya yeten bir program ortaya çıkardık.*

**KAYNAKÇA**

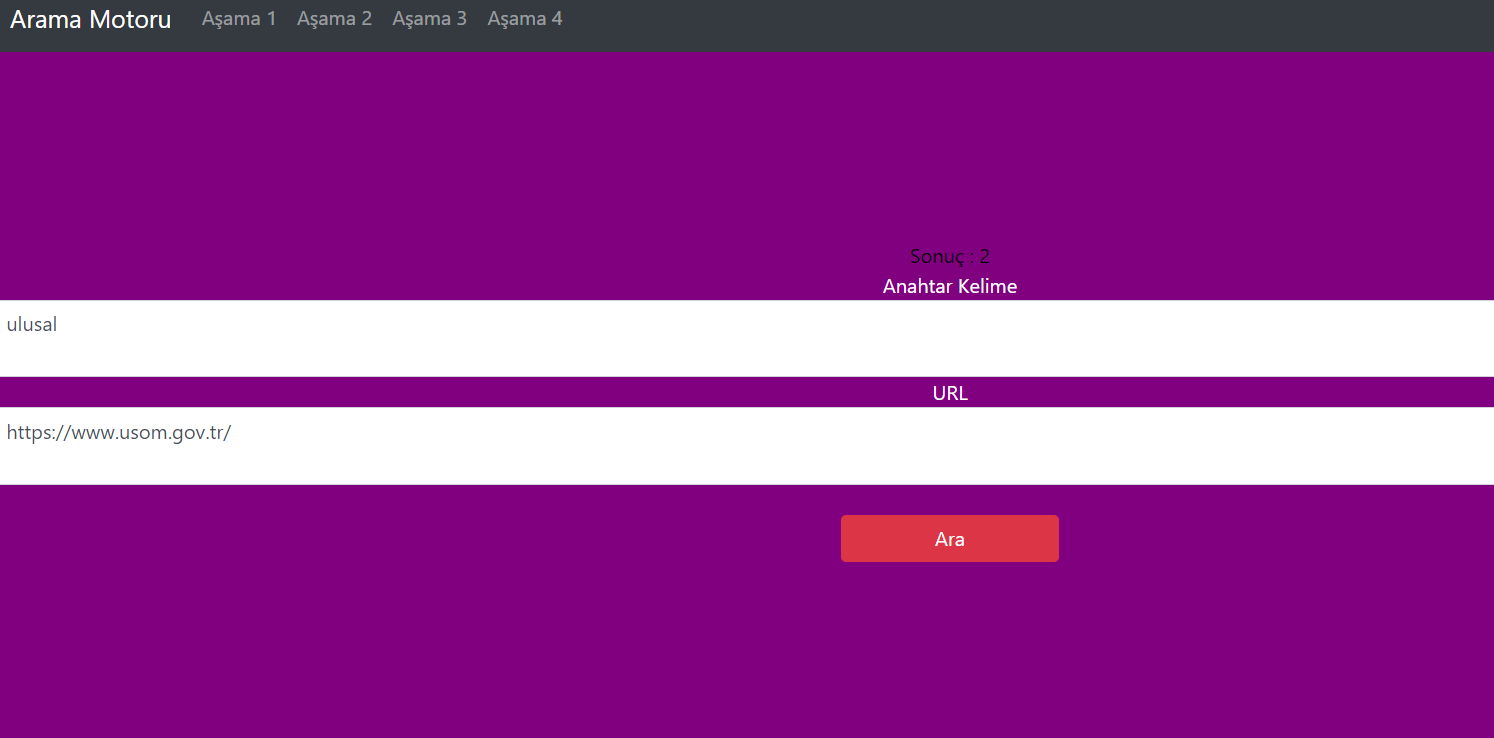
* [*https://stackoverflow.com/questions/23783679/how-to-convert-string-to-html-safe-string*](https://stackoverflow.com/questions/23783679/how-to-convert-string-to-html-safe-string)
* [*https://stackoverflow.com/questions/4935446/string-to-htmldocument*](https://stackoverflow.com/questions/4935446/string-to-htmldocument)
* [*https://msdn.microsoft.com/tr-tr/library/system.windows.forms.htmldocument(v=vs.110).aspx*](https://msdn.microsoft.com/tr-tr/library/system.windows.forms.htmldocument(v=vs.110).aspx)
* [*https://stackoverflow.com/questions/19604577/how-to-get-all-display-text-from-a-webpage-in-c-sharp*](https://stackoverflow.com/questions/19604577/how-to-get-all-display-text-from-a-webpage-in-c-sharp)
* [*https://stackoverflow.com/questions/2038104/parsing-html-to-get-content-using-c-sharp*](https://stackoverflow.com/questions/2038104/parsing-html-to-get-content-using-c-sharp)
* [*https://stackoverflow.com/questions/2248411/get-all-links-on-html-page*](https://stackoverflow.com/questions/2248411/get-all-links-on-html-page)
* [*https://stackoverflow.com/questions/18909452/getting-list-of-values-out-of-list-of-objects*](https://stackoverflow.com/questions/18909452/getting-list-of-values-out-of-list-of-objects)
* [*https://stackoverflow.com/questions/3622751/c-sharp-getting-a-list-from-a-field-out-of-a-list*](https://stackoverflow.com/questions/3622751/c-sharp-getting-a-list-from-a-field-out-of-a-list)
* [*https://stackoverflow.com/questions/4572562/find-an-element-in-list-of-list-c-sharp*](https://stackoverflow.com/questions/4572562/find-an-element-in-list-of-list-c-sharp)
* [*http://www.ahmetcansever.com/programlama/c-class-ve-list-kullanarak-veri-tabanindan-veri-cekme/*](http://www.ahmetcansever.com/programlama/c-class-ve-list-kullanarak-veri-tabanindan-veri-cekme/)
* [*https://stackoverflow.com/questions/26654369/why-im-getting-exception-too-many-automatic-redirections-were-attempted-on-web*](https://stackoverflow.com/questions/26654369/why-im-getting-exception-too-many-automatic-redirections-were-attempted-on-web)
* [*https://stackoverflow.com/questions/2933464/how-to-build-a-search-engine-in-c-sharp*](https://stackoverflow.com/questions/2933464/how-to-build-a-search-engine-in-c-sharp)
* [*https://csharpiledundenbugune.blogspot.com.tr/2017/09/c-streamreader-turkce-karakter-sorunu\_15.html*](https://csharpiledundenbugune.blogspot.com.tr/2017/09/c-streamreader-turkce-karakter-sorunu_15.html)
* [*https://stackoverflow.com/questions/5766573/what-is-the-best-way-to-compare-2-integer-lists-array-in-c-sharp*](https://stackoverflow.com/questions/5766573/what-is-the-best-way-to-compare-2-integer-lists-array-in-c-sharp)
* [*https://stackoverflow.com/questions/852439/how-to-sort-elements-of-array-list-in-c-sharp*](https://stackoverflow.com/questions/852439/how-to-sort-elements-of-array-list-in-c-sharp)
* [*https://stackoverflow.com/questions/599275/how-can-i-download-html-source-in-c-sharp*](https://stackoverflow.com/questions/599275/how-can-i-download-html-source-in-c-sharp)
* [*https://stackoverflow.com/questions/5909864/multiline-textbox-multiple-newline*](https://stackoverflow.com/questions/5909864/multiline-textbox-multiple-newline)
* [*https://stackoverflow.com/questions/5183385/htmlagilitypack-get-whole-html-document-as-string*](https://stackoverflow.com/questions/5183385/htmlagilitypack-get-whole-html-document-as-string)
* [*https://stackoverflow.com/questions/23783679/how-to-convert-string-to-html-safe-string*](https://stackoverflow.com/questions/23783679/how-to-convert-string-to-html-safe-string)
* [*https://stackoverflow.com/questions/9279174/how-to-convert-string-in-html-asp-net*](https://stackoverflow.com/questions/9279174/how-to-convert-string-in-html-asp-net)
* [*https://stackoverflow.com/questions/19980657/display-string-as-html-in-asp-net-mvc-view*](https://stackoverflow.com/questions/19980657/display-string-as-html-in-asp-net-mvc-view)
* [*https://stackoverflow.com/questions/2038104/parsing-html-to-get-content-using-c-sharp*](https://stackoverflow.com/questions/2038104/parsing-html-to-get-content-using-c-sharp)
* [*https://stackoverflow.com/questions/3622751/c-sharp-getting-a-list-from-a-field-out-of-a-list*](https://stackoverflow.com/questions/3622751/c-sharp-getting-a-list-from-a-field-out-of-a-list)
* [*https://www.codeproject.com/Questions/204778/Get-HTML-code-from-a-website-C*](https://www.codeproject.com/Questions/204778/Get-HTML-code-from-a-website-C)
* [*https://stackoverflow.com/questions/286813/how-do-you-convert-html-to-plain-text*](https://stackoverflow.com/questions/286813/how-do-you-convert-html-to-plain-text)
* [*https://stackoverflow.com/questions/3231969/download-file-from-url-to-a-string*](https://stackoverflow.com/questions/3231969/download-file-from-url-to-a-string)
* [*https://stackoverflow.com/questions/4758283/reading-data-from-a-website-using-c-sharp*](https://stackoverflow.com/questions/4758283/reading-data-from-a-website-using-c-sharp)
* [*https://stackoverflow.com/questions/8129154/how-to-get-text-off-a-webpage*](https://stackoverflow.com/questions/8129154/how-to-get-text-off-a-webpage)
* [*https://stackoverflow.com/questions/3991840/simple-text-to-html-conversion*](https://stackoverflow.com/questions/3991840/simple-text-to-html-conversion)
* [*https://github.com/xoofx/NUglify*](https://github.com/xoofx/NUglify)
* [*https://social.msdn.microsoft.com/Forums/en-US/365798c6-fb1a-4fee-9e73-0cb5f4dd00de/parse-and-extract-text-from-html-webpage?forum=csharpgeneral*](https://social.msdn.microsoft.com/Forums/en-US/365798c6-fb1a-4fee-9e73-0cb5f4dd00de/parse-and-extract-text-from-html-webpage?forum=csharpgeneral)
* [*https://stackoverflow.com/questions/286813/how-do-you-convert-html-to-plain-text*](https://stackoverflow.com/questions/286813/how-do-you-convert-html-to-plain-text)
* [*http://www.gorselprogramlama.com/dizi-array-sinifinin-metotlari-c-net-ders-44/*](http://www.gorselprogramlama.com/dizi-array-sinifinin-metotlari-c-net-ders-44/)
* [*https://getbootstrap.com/*](https://getbootstrap.com/)
* [*https://v4-alpha.getbootstrap.com/components/navbar/*](https://v4-alpha.getbootstrap.com/components/navbar/)
* [*https://getbootstrap.com/docs/3.3/components/#navbar*](https://getbootstrap.com/docs/3.3/components/#navbar)
* [*http://www.dilbilgisi.net/sozlukler/es-anlamli-kelimeler-sozlugu/*](http://www.dilbilgisi.net/sozlukler/es-anlamli-kelimeler-sozlugu/)
* [*https://stackoverflow.com/questions/19640590/reading-from-a-local-text-file-into-c*](https://stackoverflow.com/questions/19640590/reading-from-a-local-text-file-into-c)
* [*https://stackoverflow.com/questions/15232813/add-strings-to-string-array*](https://stackoverflow.com/questions/15232813/add-strings-to-string-array)
* [*https://stackoverflow.com/questions/15926142/regular-expression-for-finding-href-value-of-a-a-link*](https://stackoverflow.com/questions/15926142/regular-expression-for-finding-href-value-of-a-a-link)
* [*http://www.yazilimdilleri.net/YazilimMakale-3138-C--Web-Sitesinden-Veri-Cekmek--HmtlAgilityPack-Kullanimi--ile-HTML-Parse-Etme.aspx*](http://www.yazilimdilleri.net/YazilimMakale-3138-C--Web-Sitesinden-Veri-Cekmek--HmtlAgilityPack-Kullanimi--ile-HTML-Parse-Etme.aspx)
* [*http://www.osmankurt.net/post/2017/03/07/c-sharp-ile-html-kodlarini-temizleme.aspx*](http://www.osmankurt.net/post/2017/03/07/c-sharp-ile-html-kodlarini-temizleme.aspx)
* [*https://social.msdn.microsoft.com/Forums/tr-TR/48c27fb3-03c5-45ac-a860-4e9165643cc3/c-stenilen-kelimeleri-string-fadeden-silmek-?forum=csharptr*](https://social.msdn.microsoft.com/Forums/tr-TR/48c27fb3-03c5-45ac-a860-4e9165643cc3/c-stenilen-kelimeleri-string-fadeden-silmek-?forum=csharptr)
* [*http://kod.gen.tr/php-html-tag-temizleme*](http://kod.gen.tr/php-html-tag-temizleme)
* [*https://stackoverflow.com/questions/18153998/how-do-i-remove-all-html-tags-from-a-string-without-knowing-which-tags-are-in-it*](https://stackoverflow.com/questions/18153998/how-do-i-remove-all-html-tags-from-a-string-without-knowing-which-tags-are-in-it)
* [*http://www.c-sharpcorner.com/code/927/c-sharp-regular-expression-code-to-remove-html-tags-in-c-sharp.aspx*](http://www.c-sharpcorner.com/code/927/c-sharp-regular-expression-code-to-remove-html-tags-in-c-sharp.aspx)
* [*https://www.google.com.tr/search?q=substring+c%23&oq=substri&aqs=chrome.2.69i57j0l2j69i61j0l2.8012j0j7&sourceid=chrome&ie=UTF-8*](https://www.google.com.tr/search?q=substring+c%23&oq=substri&aqs=chrome.2.69i57j0l2j69i61j0l2.8012j0j7&sourceid=chrome&ie=UTF-8)
* [*https://github.com/Rebelmail/html-uglify*](https://github.com/Rebelmail/html-uglify)



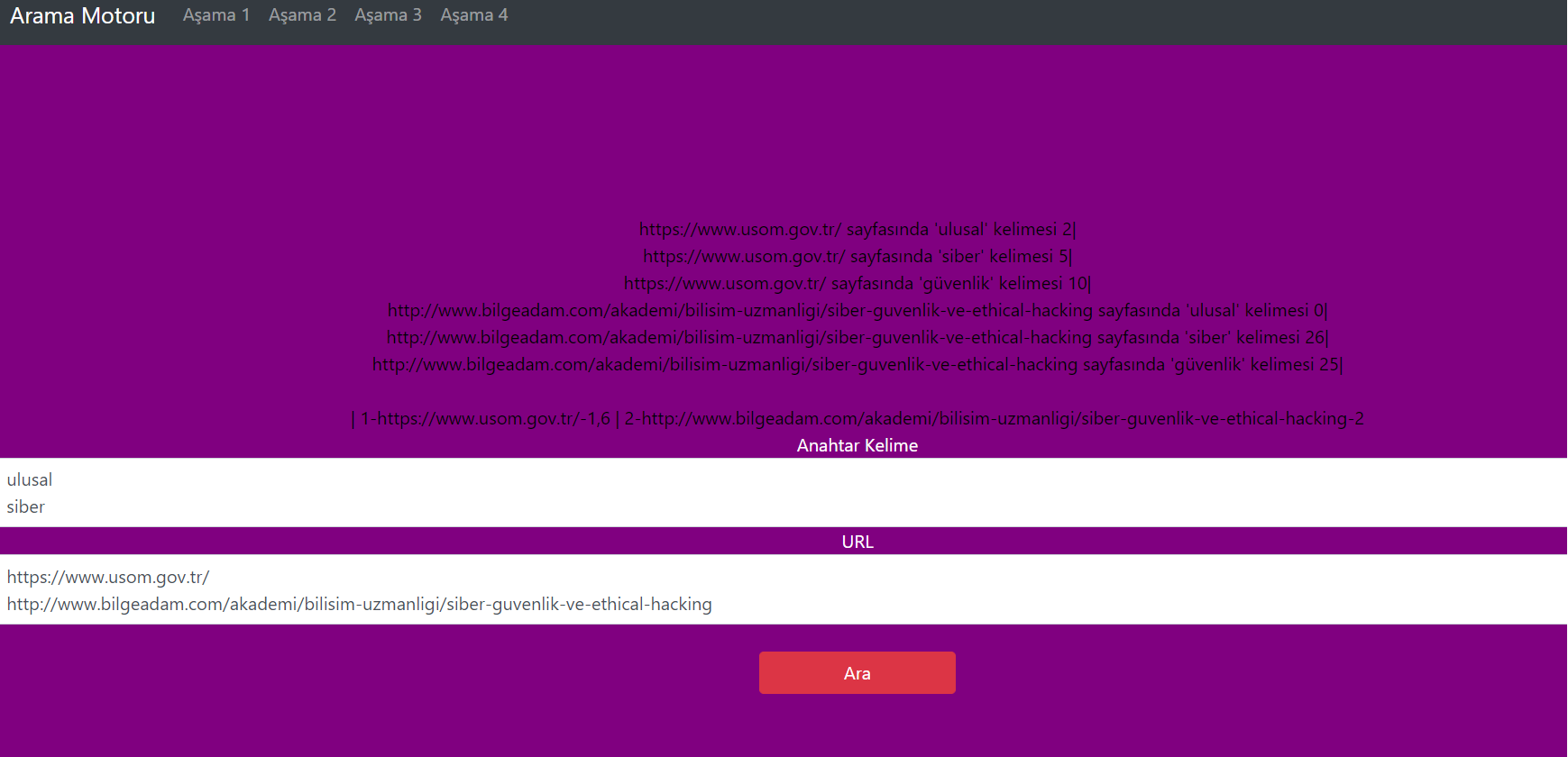
***Tasarım***



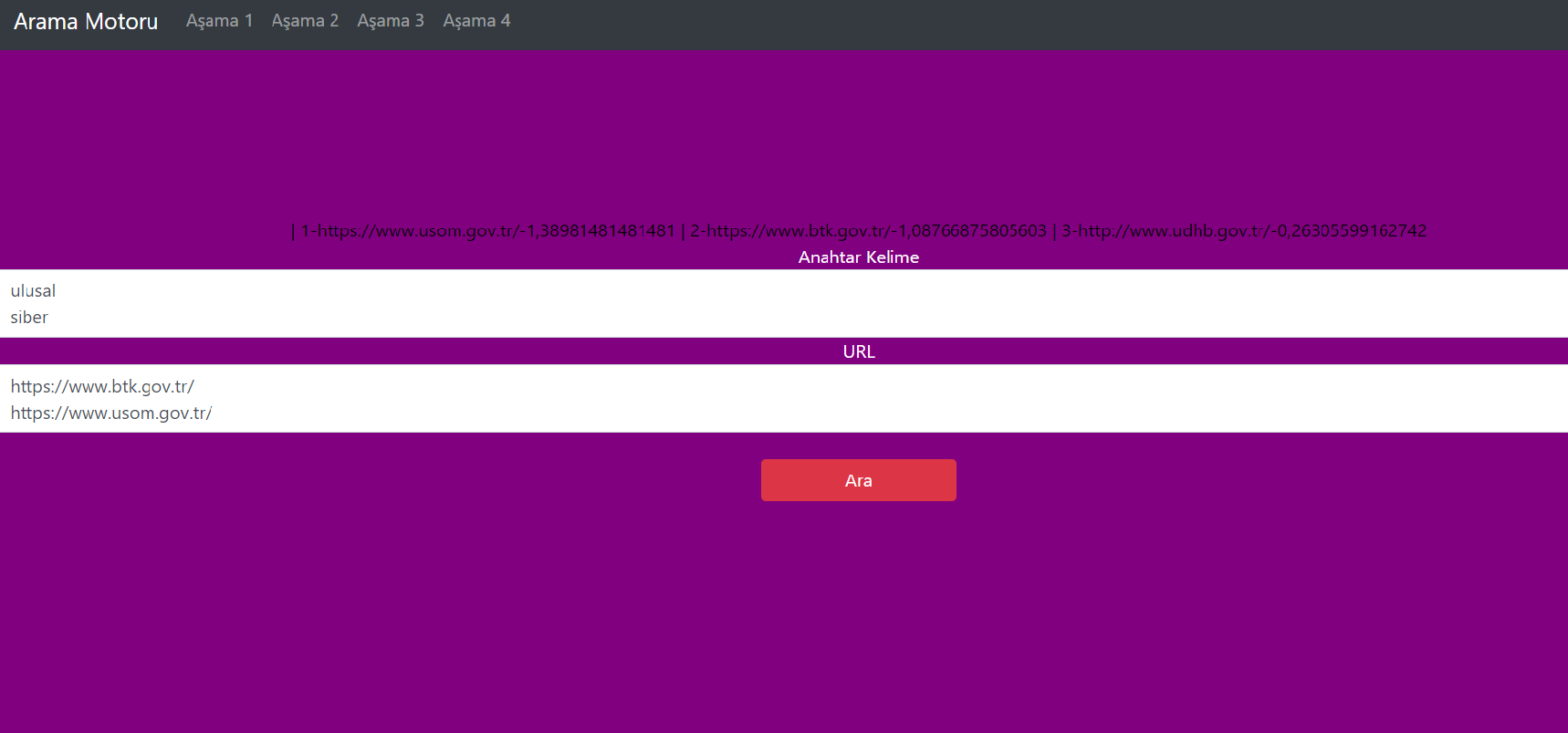
***Arayüzümüz***



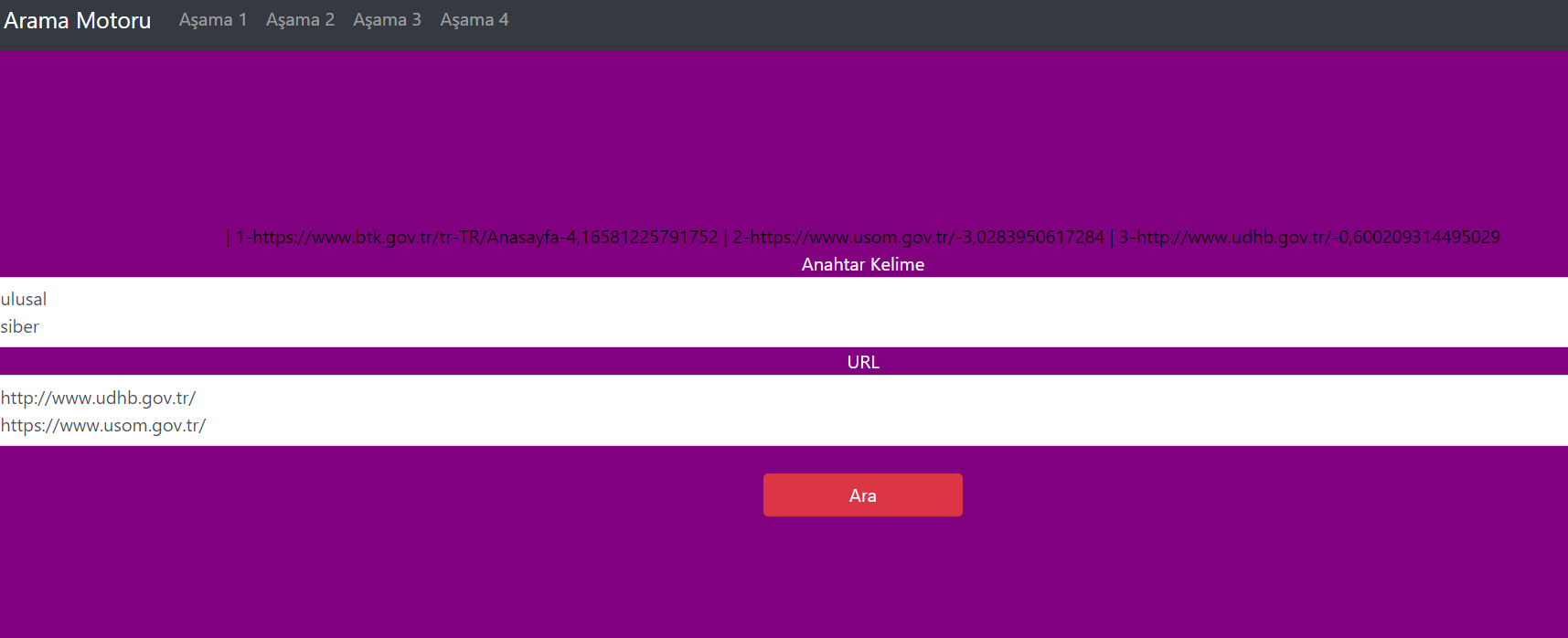
***Aşama-1***



***Aşama-2***



***Aşama-3***



***Aşama-4***