

T.C
TRAKYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Dikey Arama Motorlarının İncelenmesi Ve
Bir Dikey Arama Motoru Uygulaması

Can RAZBONYALI

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Yrd. Doç. Dr. Aydın CARUS

Edirne – 2011

T.C.
TRAKYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DİKEY ARAMA MOTORLARININ İNCELENMESİ VE
BİR DİKEY ARAMA MOTORU UYGULAMASI

Can RAZBONYALI

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Ana Bilim Dalı

Bu tez .../.../2011 tarihinde, aşağıdaki jüri tarafından kabul edilmiştir.

Yrd. Doç. Dr. Aydın CARUS

Danışman

Yrd. Doç. Dr. Erdiñç UZUN

Üye

Yrd. Doç. Dr. Tarık YERLİKAYA

Üye

Yüksek Lisans Tezi

Trakya Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

ÖZET

DİKEY ARAMA MOTORLARININ İNCELENMESİ VE BİR DİKEY ARAMA MOTORU UYGULAMASI

Günümüzde, ağ bant genişliklerinin artması ve internet kullanımının yaygınlaşması ile internet neredeyse sınırsız bir veri büyüklüğüne erişmiştir. Ancak bu durum, bu büyük miktarda veri kümesi içerisinde aranan verinin bulunması problemini de beraberinde getirmiştir. Bu probleme çözüm olarak Arama Motorları kullanıma sunulmuştur. Ancak internette mevcut verileri ayırım ve sınıflama yapmadan indeksleyen Yatay Arama Motorları, kullanıcılara sadece aradığı bilgiyi içeren web sayfalarını görüntülemeye çok başarılı değildir. Bu probleme çözüm olarak ise, sadece arama motoru kapsamına giren konulara özel verileri indeksleyerek, kullanıcılara aradıkları bilgiyi içeren web sayfalarını görüntüleyen, dikey arama motorları geliştirilmiştir. Dikey arama motorları bir web sayfasının içeriğinin Dikey Arama Motoru kapsamına girip girmediğini belirlemede çeşitli yaklaşımlar kullanılmaktadır.

Bu tez çalışmasının amacı, Yatay Arama Motorları ile Dikey Arama Motorları arasındaki farkı belirterek Arama Motorlarının yapısını ve çalışma prensiplerini ayrıntılı olarak incelemektir. Tez çalışmasında ayrıca bir Dikey Arama Motoru tasarlanmış ve geliştirilmiştir. Geliştirilen Dikey Arama Motorunda bir web sayfasının içeriğinin sınıfını belirlemede Makine Öğrenmesinden yararlanılmıştır. Makine Öğrenmesi kullanan Dikey Arama Motorunun eğitim setinin değerlendirmesi yapılmış ve sonuçlar sunulmuştur.

Bu tez 2011 yılında yapılmış olup 122 sayfadan oluşmaktadır.

Anahtar Kelimeler : Arama motoru, Dikey arama motoru, Ağ örümceği, Makine öğrenmesi

Master Thesis

Trakya University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

ABSTRACT

RESEARCH ON VERTICAL SEARCH ENGINES AND DEVELOPING AN APPLICATION ON VERTICAL SEARCH ENGINE

Today, the network band widths increase with the spread of internet use and the size of the Internet has reached almost unlimited data. This, however, that the problem of finding a large amount of data sought has brought through a set of data. Search Engines are available as a solution to this problem. However, the distinction and classification of the data available on the internet without the horizontal indexing search engines, users are only viewing web pages that contain the information sought is not very successful. As a solution to this problem, the only matters which fall within the scope of the search engine indexing private data, users browsing web pages that contain information they are looking, vertical search engines were developed. Vertical Search Engines use various approaches to determine whether the scope of the contents of a web page. The aim of this thesis, stating the difference between the Traditional Search Engines and Vertical Search Engines to examine in detail the structure and operating principles. The aim of this thesis, stating the difference between the horizontal with Search Engines Vertical Search Engines, Search Engines to examine in detail the structure and operating principles. In this process also designed and developed a Vertical Search Engine. In developed Vertical Search Engine, Machine Learning System is used to determine the class of the contents of the web page. Vertical Search Engine that uses machine learning assessment was conducted in and the results are presented in training set.

This thesis has been completed in 2011 and consists of 122 pages.

Keywords : Search engine, Vertical search engine, Web crawler, Machine learning

TEŞEKKÜR

Bu çalışmanın hazırlanması sürecinde bana yol gösteren, bu alanda çalışmam için beni teşvik eden ve yapmış olduğu katkılardan dolayı danışman hocam Yrd. Doç. Dr. Aydın CARUS'a teşekkür ederim.

Tez çalışmasında uygulamanın geliştirilmesi ve sonuçların yorumlanması için yaptıkları katkılardan dolayı Yrd. Doç. Dr. Erdinç UZUN ve Arş. Gör. Edip Serdar GÜNER hocalarıma teşekkür ederim.

Benim bu alanda ilerlemem için bana ışık tutan, maddi ve manevi yardımlarını esirgemeyen babam Prof. Dr. A. Mesut RAZBONYALI' ya teşekkür ederim.

Tez çalışması esnasında bana maddi ve manevi desteklerini esirgemeyen çok değerli aileme teşekkür ederim.

Uygulama geliştirilmesi sürecinde, fikirleriyle bana yol gösteren Sabri MUMCU' ya teşekkür ederim.

Düzenli çalışmam için yapmış olduğu fedakârlıklardan ve yabancı dildeki katkılarından dolayı değerli arkadaşım Jasmin Sena BARAN'a teşekkür ederim.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
SİMGELER DİZİNİ	vi
1. GİRİŞ	1
2. ARAMA MOTORLARI VE ÇALIŞMA PRENSİBİ	3
2.1. ARAMA MOTORU TÜRLERİ	5
2.1.1. Örümcek Tabanlı Arama Motoru	5
2.1.2. Dizin Tabanlı Arama Motorları	6
2.1.3. Hibrit Arama Motoru	6
2.2. Ağ Örümceği Yapısı ve Çalışma Şekli	7
2.2.1. Ağ örümceği türleri	9
2.2.1.1. Odaklı ağ örümceği	9
2.2.1.2. Semantik ağ örümceği.....	9
2.2.1.3. Öğrenimli ağ örümceği.....	10
2.3. Html Ayırıcı ve Ayırma İşlemi	11
2.3.1. Metin.....	13
2.3.2. Açıklama Satırı.....	13
2.3.3. Basit Etiketler	13
2.3.4. Başlangıç ve Bitiş Etiketleri.....	14
2.3.5. Bağlantılar ile gezinme	15
2.4. İndeksleyici.....	16
2.4.1. İndeksleme metotları.....	16
2.4.1.1. Tam metin indeksleme.....	17
2.4.1.2. Anahtar kelime indeksleme	17
2.4.1.3. İnsan destekli indeksleme	18
3. DİKEY ARAMA MOTORU	19
3.1. Dikey Arama Motoru Kullanım Alanları	20
4. YAPAY ÖĞRENME NEDİR?	22
5. YAPAY ÖĞRENME UYGULAMALARINA ÖRNEKLER	24

5.1. İlişkilendirme Kuralları	24
5.2. Sınıflandırma	24
5.3. Bağlanım.....	26
5.4. Gözetimsiz Öğrenme.....	28
5.5. Pekiştirmeli Öğrenme	29
6. GELİŞTİRİLEN YAZILIMDA KULLANILAN YARDIMCI ARAÇLAR	30
6.1. WEKA NEDİR?.....	30
6.2. WEKA İle Sınıflamadan Önceki Hazırlıklar.....	31
6.3. WEKA İle Sınıflama	37
6.3.1. ARFF dokümanının içeriği.....	38
6.4. WEKA'nın Kullandığı Sınıflandırma Algoritmaları.....	40
6.4.1. En yakın k komşu algoritması ile karar verme	40
6.4.2. Karar ağacı algoritması ile karar verme	41
7. DİKEY ARAMA MOTORU İLE İLGİLİ MEVCUT ÇALIŞMALAR	42
8. GELİŞTİRİLEN DİKEY ARAMA MOTORU	45
8.1. Geliştirilen Ağ Robotu	45
8.2. Kullanıcı Arayüzü.....	50
9. SONUÇLAR VE ÖNERİLER	55
KAYNAKLAR	60
ÖZGEÇMİŞ	61
EK A – AĞ ÖRÜMCEĞİ	62
EK B – SPİDER.CD SINIFI	73
EK C – HTMLPARSER.CS SINIFI.....	83
EK D – EĞİTİM SETİ DOKÜMANI	84
EK E – EĞİTİM SETİ OLUŞTURMADA KULLANILAN WEB SAYFALARI	95
EK F – EN YAKIN K KOMŞU ALGORİTMASI KULLANIM KODLARI	107
EK G – KARAR AĞACI ALGORİTMASI KULLANIM KODU	108
EK H – ARAMA SAYFASI JAVASCRIPT KONTROLÜ	109
EK I – KULLANICI GİRİŞ SAYFASI JAVASCRIPT KONTROLÜ	110
EK J – KULLANICI KAYIT SAYFASI JAVA SCRIPT KONTROLÜ	111
EK K – KULLANICI KAYIT SAYFASI JAVA SCRIPT KONTROLÜ	112

SİMGELER DİZİNİ

Kısaltmalar Tablosu

@attribute.....	Öznitelik
@data.....	Veri
@relation.....	İlişki
Ağ Örümceği.....	Web Crawler
ARFF.....	Attribute-Relation File Format
Bağlantı Madenciliği.....	Link Miner
Bağlantı.....	Link
Dizin Tabanlı Arama Motoru.....	Directory Based Search Engine
Düzenleyici.....	Human Editor
Etiket.....	Tag
Hibrit Arama Motorları.....	Hybrid Search Engine
HTML.....	Hyper Text Markup Language
İnsan Destekli Arama Motorları.....	Human Powered Search Engine
İyileştirici.....	Optimizer
Nominal – Belirtme	Nominal – Specification
Örümcek Tabanlı Arama Motoru.....	Crawler Based Search Engine
PISE.....	Personal Intelligent Search Engine
WEKA.....	Waikato Environment for Knowledge Analysis
XML.....	Extensible Markup Language
Özetleyici.....	Summarizer
SSMR.....	Semantic Similarity Retrieval Model
ASCII.....	Bilgi Değişimi İçin Amerikan Standartları Kodlama Sistemi

1. GİRİŞ

Günümüzde bilgisayar donanım maliyetlerinin düşmesi ve ağ bant genişliklerinin eskiye göre önemli ölçüde artmış olmasının bir sonucu olarak internetin kullanılması da önemli oranda yaygınlaşmıştır.

İnternet kullanımının yaygınlaşması, internette bulunan bilgi miktarının da artmış olmasını beraberinde getirmiştir. İnternette mevcut bilgi miktarını tam olarak saptayabilmek mümkün değildir ancak, 2003 yılında yapılan bir araştırmaya göre, internette bulunan sayısal bilgilerin her yıl %69 oranında arttığı ortaya konmuştur[1]. 2005 yılının Ocak ayında yapılan başka bir araştırmada da, internette 11,5 milyar adet indekslenebilir web sayfasının olduğu ortaya konulmuştur[2]. Ayrıca, internetteki mevcut web sayfalarına her gün 11 milyon adet yeni web sayfasının eklendiği de belirtilmiştir[2]. Verilen bu bilgiler ışığında şu anda internette mevcut sayısal bilginin çok geniş olduğunun tahmin edilmesi zor değildir. Bu bilgi genişliğinin, verilerin saklanması ve gerektiğinde erişilmesi gibi bir takım sorunları da beraberinde getiriyor olması kaçınılmazdır.

İnternetteki veri büyüklüğü hızla artarken bilgisayar kullanıcılarının, bu artan veri yığını içinden istediği doğru bilgiye en kısa zamanı harcayarak ulaşmak isteyecekleri de kaçınılmaz bir gerçek olmuştur. Yoğun çalışma temposunu da göz önünde bulundurursak, internette veri arayacak olan kullanıcıların gereksiz verileri elemek için fazladan zaman harcamak istemeyecekleri haklı bir gerekçe olmaktadır.

İnternet ortamında birçok bilginin bulunması, bir anlamda bilgiye ulaşımın kolaylaşmış olması, büyük bir avantajdır. Ancak bu aşamada karşımıza bu kadar çok bilgi içerisinden istenen bilgiye ulaşma problemi çıkmaktadır. Buna çözüm olarak ise, aranan bilgiye ulaşılabilirliği oldukça kolaylaştıran arama motorları devreye girmektedir. Fakat arama motorları aracılığı ile ulaşılan bu bilgilerin ne kadar istenen bilgi ya da ne kadar doğru oldukları tartışmalıdır. Bu durum, zamanın çok kıymetli olduğu bu günlerde arama motoru vasıtası ile bilgi arayan kullanıcılar için problem yaratmaktadır.

Arama motoru, internet üzerindeki sayfalarda bulunan bağlantıları ziyaret eden ve bu ziyaret sonunda da sayfalardaki verileri manyetik ortamlarda indeksli olarak saklayarak sonradan arama işlemi yapan kullanıcılara, aradığı anahtar kelimeler ile ilişkili olduğunu tespit ettiği sonuçları görüntüleyen bir sistemdir. Arama motorları, ziyaret etmiş olduğu sitelerden indekslediği verileri arama yapan kullanıcılara sonuç olarak görüntülediği için, kullanıcının aradığı bilgi ile ilişkili olmayan sonuçları da kullanıcıya görüntüleyebilmektedir. Ayrıca internetin bir başka yüzü olan bilgi kirliliği de düşünülür ise, kötü amaçlı siteleri de aranan bilgiyle ilgili gibi göstermesi de bir başka eksiklik ya da dezavantaj olarak karşımıza çıkan bir başka durumdur. Bu sebepten dolayı, kullanıcı arama motorundan elde ettiği sonuçlar arasından tekrar bir eleme veya gözden geçirme yapmak durumunda kalmaktadır. Bu soruna çözüm olarak belli konuları kapsayacak verileri indeksleyen arama motorları tasarlanmıştır. Belirli bir konuya özel olan bu arama motorlarına özelleştirilmiş ya da dikey arama motoru adı verilmektedir. Ancak dikey arama motorları da sorunlarını beraberinde getirmiştir. Bu sorunların en önemlisi ise bir konuya ait verileri içeren bir dikey arama motorunun internet üzerinde mevcut birçok siteyi içerik olarak değerlendirip, sitedeki verilerin arama motoru kapsamına giren veriler olup olmadığına karar verebilmesidir. Bu aşamada farklı yaklaşımlar kullanılmaktadır.

Bu çalışmada, arama motoru mimarisi ayrıntılı şekilde incelenmiş, yatay arama motorlarının sahip oldukları dezavantajlara çözüm olarak düşünülen dikey arama motorları ayrıntılı olarak açıklanmıştır. Çalışmada ayrıca, bir konuya özel olarak dikey bir arama motoru tasarlanmış ve geliştirilmiştir. Geliştirilen arama motorunda, ağ örümceği (web crawler) tarafından indirilen web sitesinin içeriğinin dikey arama motoru kapsamına girip girmediğini saptamak amacıyla makine öğrenmesi kullanılmıştır. Geliştirilen dikey arama motorunda, web sitesinin içeriğinin arama motoru kapsamına girip girmediğini tespit için WEKA (Waikato Environment for Knowledge Analysis)[3] yazılım paketinin kütüphanelerinden yararlanılmıştır. Geliştirilen dikey arama motorunda kullanılan eğitim seti üzerinden, arama motorunun başarımı değerlendirilmiş ve sonuçlar sunulmuştur.

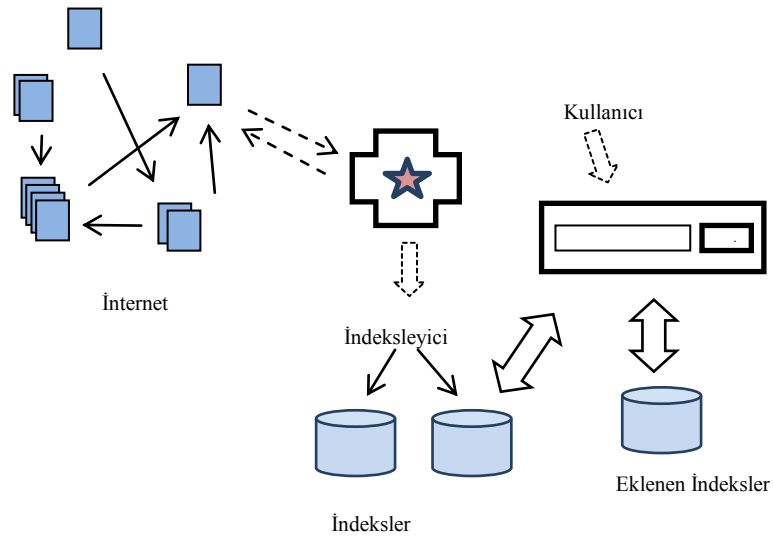
2. ARAMA MOTORLARI VE ÇALIŞMA PRENSİBİ

Arama motorları ve bilgi erişim sistemleri eş anlamlı terimler olarak sayılabilirler. Bilgi erişim sistemleri, kullanıcıların gereken bilgi ihtiyaçlarını karşılaması ve ilgili belgelerin tümüne erişirken ilgisiz olanları da ayıklamasıdır[4]. Bu bağlamda arama motorları da bilgiye erişim sağlayan sistemlerdir.

İnternetteki web sayfaları birbiri ile bağlantılar (link) aracılığı ile ilişkilidirler. Bu bağlantı, bir sayfadan diğerine geçebilme esnekliği sağlamaktadır. Genellikle, bir web sayfasında bulunan bağlantı, bulunduğu sayfanın konusu ile ilgili farklı bir web sayfasına bir bağlantıdır. Bu duruma örnek verilecek olursa, genel olarak basketbol malzemeleri ile ilgili bir web sayfası olan “www.baskethane.com” adresli web sitesi, yine basketbol malzemeleri ile ilgili diğer sayfalara bağlantılar içermektedir. Arama motoru; web sayfalarındaki verileri kendisine ait veri tabanına kaydederek, bu dokümanlardan indeks oluşturur. Web sayfaları, genellikle HTML (Zengin Metin İşaret Dili-Hyper Text Markup Language)[5] ve XML (Genişletilebilir İşaretleme Dili-Extensible Markup Language)[5] gibi işaretleme dilleri ile hazırlanırlar. Bu sebeple indeks oluşturma öncesinde, web sayfası içeriğinden, bu sayfaların diğer sayfalara olan bağlantılarını elde etmek ve sayfanın işlenebilir hale gelmesi için web sayfalarının yapısına ait olan düzenleme etiketlerinden (tag) temizlenmesi gerekmektedir. Bu amaçla sayfa ayırıcı (HTML/XML Parser) isimli arama motoru bileşeni kullanılır. Bu bileşen, ayrıntılı olarak bu bölümde açıklanmıştır.

İndeksleme işleminde, indekslenen verilerin güncelliğini korumak amacı ile sayfaların ağ örümceği tarafından belirli aralıklar ile tekrar ziyaret edilmesi gerekmektedir. İndeksler, arama motorunun veri tabanında veya sabit diskte kendisine özel yapıda saklanır. Arama motoru kullanıcıları ise kullanıcı ara birimi yardımı ile arama motorunun oluşturmuş olduğu indekslerden, anahtar kelimeler sayesinde aradıkları bilgilerin hangi sitelerde oldukları bilgisini görüntülerler. Dolayısı ile arama motorları, kullanıcıların internette aradıkları bilginin görüntülenmesinde önemli rol oynarlar. Şekil 2.1’ de görüldüğü gibi genel olarak bir arama motoru, ağ örümceği, indeksleyici ve kullanıcı ara birimi bileşenlerinden oluşmaktadır. Ayrıca ağ örümceği

tarafından indirilmiş olan sayfaların indekslenme öncesi ayrıştırılması işlemi için sayfa ayırıcı bileşeni görev yapar. Günümüzde en popüler arama motorları arasında “google.com”, “yahoo.com”, “bing.com”, “altavista.com”, “galaxy.com”, “kosmix.com”, “scirus.com”, “lycos.com”, “cuil.com” ve “msn.com” gibi arama motorları sayılabilir.



Şekil 2.1. Arama Motoru Mimarisi

Daha önce de belirttiğimiz üzere internetteki web sayfaları birbirlerine bağlantılar ile ilişkilendirilmiştir. Arama motorunun bir bileşeni olan ağ örümceği, bu sayfalar üzerindeki bağlantıları ziyaret eder. Ağ örümceğinin web sayfalarını ziyaret etmesi belli aralıklar ve belli ilkelere bağlı kalınarak yapılır. Web sayfası ve sayfanın bulundurduğu bağlantılardaki web sayfaları, ağ örümceği tarafından indirir. Ağ örümceğinin yapısı ve çalışma şekli Bölüm 2.2’ de ayrıntılı bir biçimde anlatılmıştır.

2.1. Arama Motoru Türleri

Arama motorları, bir anlamda internette bulunan web sayfalarını indeksleyen yazılımlardır. İnternetteki web sayfalarındaki verilere ulaşım, bunları indeksleyebilmesi için bir arama motorunun öncelik ile teorik olarak internetteki tüm web sayfalarına ulaşması gerekmektedir. Bu amaçla farklı yaklaşımlar kullanan arama motorları geliştirilmiştir. Arama motorlarında, genel olarak üç temel mimari kullanılmaktadır. Bunlar, Örümcek Tabanlı Arama Motorları (Crawler Based Search Engine), Dizin Tabanlı Arama Motorları (Directory Based Search Engine) veya İnsan Destekli Arama Motorları (Human Powered Search Engine) ve Hibrit Arama Motorlarıdır (Hybrid Search Engine).

2.1.1. Örümcek Tabanlı Arama Motoru

Örümcek tabanlı arama motoru denildiğinde bu konuda öncü olan Google ve Yahoo gibi arama motorlarının kullandığı bir yapıdır. Bu yapıda, daha önce de bahsedilen ağ örümceği, internetteki web sayfalarını gezer ve kullanıcılar ağ örümceğinin oluşturmuş olduğu verilerin içerisinden arama yaparlar. Bu listeler, arama motorunun indeksleyicisi tarafından yapılmış kataloglardır. Bu indeksler, ağ örümceğinin, web sayfalarını gezerek bulduğu ve bu sayfaların içeriklerinin bulunduğu küme olarak düşünülebilir. Ağ örümceği, internetteki sayfaları ve o sayfalardan bulunduğu bağlantıları sürekli ziyaret eden bir çalışma prensibine sahiptir. Bu sürekli çalışma prensibi çerçevesinde değişmeyen yöntemler kullanır. Dolayısı ile web sayfasında yapılan herhangi bir değişiklik, o web sayfasının yapılan arama sonuçlarında görüntülenme önceliğini etkileyecektir. Bunlara ek olarak, unutulmamalıdır ki gezilen web sayfasının indekslenmesi için sayfanın arama örümceği tarafından ziyaret edilmesi gerekmektedir, bu işlem belli bir periyotta gerçekleşebilmektedir. İndekslenme yapıldıkça kadar arama motorunun o web sayfasını hatırlayıp sonuçlar arasında görüntülenmesi mümkün olamamaktadır.

2.1.2. Dizin Tabanlı Arama Motorları

Birçok arama motoru sonuçları için örümcek kullanmaktadır. Bazı arama motorları ise düzenleyici (human editor) kullanmaktadır. Bu tür arama motorlarına dizin Tabanlı veya İnsan Destekli Arama Motoru denilmektedir[6]. Bu tip arama motorlarına İnsan Destekli denmesinin sebebi, indekslenecek olan web sayfalarının değerli ya da değersiz olduğu kararı da verilerek bir insan tarafından sistemin indekslerine dâhil edilmesidir. Dolayısı ile bu yapıyı kullanan bir arama motoru, sadece anahtar kelimeye ve teknolojiye güvenmeyip, insan emeğini yapının içerisine dâhil etmektedir. Bu sebeple resim, video vb gibi metin dışındaki içeriklerin olduğu web sayfalarında insan tarafından yapılan değerlendirme daha değerli olmaktadır. Web sayfalarının değerlendirilmesi esnasında hata yapılabilme durumlarının olduğu düşünülecek olursa, hata ile yanlış değerlendirilen bir web sayfası için yapılan hatanın farkedilip düzeltilmesi de büyük emek gerektirecektir. Çünkü bir web sayfası dizin tabanlı bir yapıya sahip olduktan sonra arama motoru önceliğini değiştirmek çok zordur. Birçok arama motoru gibi, LookSmart, Lycos, AltaVista, MSN, Excite ve AOL'da arama sonuçlarını daha anlamlı kılmak için dizin verisi tedarik eden sisteme dayalı yapılardan kullanmaktadır.

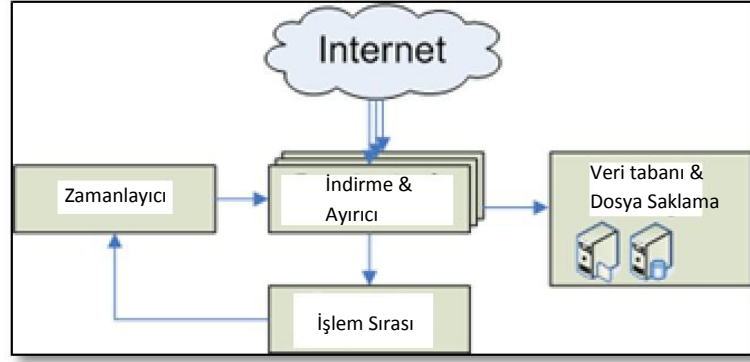
2.1.3. Hibrit Arama Motoru

İnternetin ilk zamanlarında; arama sonuçları, örümcek tabanlı veya dizin tabanlı arama motorları tarafından görüntülenmekteydi. Şu anda ise her iki çeşit arama motorunun sonuçlarını görüntülemek daha yaygındır[7]. Hibrit arama motoru, web sayfası bilgisi edinmek için, Örümcek Tabanlı ve Dizin Tabanlı Arama Motorlarının yapılarını bir arada kullanan arama motoru çeşidine verilen isimdir. Hibrit arama motorları diğer arama motorlarından farklı bir mimariye sahip olmayıp, kullanıcının aradığı anahtar kelimeye göre arama yapılır. Bu arama motoruna hibrit arama motoru denilmesinin sebebi, arama sonuçlarının hem örümcek tabanlı, hem de dizin tabanlı

arama motoru özellikleri kullanılarak elde edilen kayıtlarından görüntülenmesidir. Google gibi birçok arama motorunun da gezilecek sitelerin kullanıcılar tarafından da eklenebildiği yapıları mevcuttur. Google kullanarak bir arama yapıldığında, daha önce de belirtildiği gibi sonuçların bir arada görüntülediği görülecektir. Google örneğine göre, bir arama yapıldığında görsellerin, videoların ve web sayfalarının bir arada görüntülenmesi bir anlamda bundandır. Hibrit arama motorları, arama sonuçlarını kullanıcılarına bu şekilde görüntülemeye oldukça başarılıdır denebilir.

2.2. Ağ Örümceği Yapısı ve Çalışma Şekli

Ağ örümceği, internetteki web sayfalarını belirli bir algoritma ve yöntemle göre tarayan bir programdır. Aynı zamanda ağ örümceği, arama motorları için önemli bir bileşendir. Ağ örümcekleri; milyonlarca web sunucuları ile etkileşimde bulunduğundan, gezinme işlemi için web sayfalarının içeriklerinin değişim hızına ayak uyduramamak ve yapmış olduğu indekslerin de eski kalmasının doğurmuş olduğu sebepten dolayı, sonuçları çok çabuk değişebilen ve bundan dolayı da dezavantaj yaratabilen bir durum içerisine girmiştir[8]. Ayrıca; buna ek olarak, gezinme hızının da arama motorunun etkinliği üzerinde rolü büyüktür. Gezinme hızı ile ifade edilen, kullanıcının gezmek istediği web sayfasının tam olarak görüntülenmesi için beklediği süredir. Bu süre ile web sayfasının içeriğinde bulunan verilerin büyüklüğü doğru orantılıdır. Eğer görüntülenmek istenen web sayfasının içeriğinde bir video dosyası varsa, kullanıcı o web sayfasını görüntülemek için biraz daha fazla süre bekleyecektir. Dolayısı ile gezinme hızı, sadece internet hızına bağlı olan bir unsur olmamak ile birlikte gezilecek web sayfalarına olan hızı ile de yakından ilgili olduğu anlaşılmaktadır[8]. Bu durum, web sunucuları üzerinde de ekstra bir yoğunluk oluşturduğundan, ağ örümceği, etkin çalışan bir algoritmaya sahip olmalıdır.



Şekil 2.2.Ağ örümceği yapısı

Ağ örümcekleri için çeşitli uygulamalar olmasına rağmen, temelde hepsi aynı işlemi yaparlar. Bütün ağ örümcekleri, web sayfasını indirir. İndirilen sayfayı ayrıştırır ve bağlantılarını çıkarır ve her ulaşılan bağlantı için bu işlemi tekrarlar. Bu işlemler, şekil 2.2’ de gösterilmiştir. Ağ örümceği internet üzerinde ziyaret ettiği web sayfalarını indirir. İndirilen bu web sayfalarının anlaşılabilmesi ve işlenebilir hale gelmesi için ayrıştırılması gerektiğinden daha önce de söz edilmişti. Ayrıştırılan sayfalar içinde bulunan diğer sayfalar ile bağlantılar tekrar ziyaret edilir. İndirilen sayfalar işlem sırasına alınır ve sıra ile sayfaların hatırlanmasını kolaylaştırmak ve sayfanın bütün içeriği yerine onu ihtiyaç duyulduğunda hedef olarak gösterecek indeksleri verilir. Daha sonra bu indeksler veri tabanına veya Arama Motoru için özel tasarlanmış başka bir yapıda sabit diske kayıt edilir. Sayfanın mevcut içeriğinin tamamını kaydetmektense, indekslerin kaydedilmesi, sabit diskte kaplayacak alanın azaltılması açısından avantaj sağlayacağından indekslerin saklanması tercih edilmektedir. Ayrıca indeks, mevcut saklanmış veriye en kısa sürede erişmek için önem teşkil etmektedir

2.2.1. Ağ örümceği türleri

2.2.1.1. Odaklı ağ örümceği

Ağ örümceklerinin genel amaçları, belirli web sayfası adresleri kümesinden ulaşabildikleri kadar çok web sayfası toplamaktır. Odaklı ağ örümceklerinin amacı ise, önceden tanımlanan konu kümeleri ile ilişkili sayfaları seçerek taramaktır. Konular, anahtar kelime kullanılarak değil örnek doküman olarak belirlenir. Yani, odaklı ağ örümceği, belirli bir konudaki dokümanları toplamak için tasarlanmıştır. Ayrıca odaklı ağ örümceği; gezinti sınırlarını, gezilmesi daha uygun en ilişkili web sayfasını bulmak için analiz eder. Bu da donanım ve ağ kaynakları için büyük bir kazanç yaratmaktadır. Bununla birlikte, ağ örümceklerinin web sayfaları üzerinde, ziyaret edilme bakımından ekstra yoğunluk oluşturduğu bilinmektedir. Odaklı ağ örümceklerinin geliştirilmesi ile ağ trafiği, ağ yoğunluğu ve sayfa indirme miktarları azaltılmaktadır.

Odaklı ağ örümceklerinin avantajlarının olduğu kesindir. Bu avantajlardan biri ağ trafiğini azaltmalarıdır. Diğer bir avantajı da konu ile ilgisiz sayfaların ilgili sayfalara göre ayıklanma oranındaki başarısıdır. Bu oranın yüksek olması sayesinde kullanıcı tarafından, konu ile ilgisi bulunmayan sayfaların ayıklanması için çok zaman harcanmaz. Bu oranın yüksek olması odaklı ağ örümceği ile normal bir ağ örümceğinin arasında bir fark yaratılır. Ayrıca bu oran yüksekliği, odaklı ağ örümceği kullanmak için önemli bir neden olarak görülebilir.

2.2.1.2. Semantik ağ örümceği

Klasik ağ örümcekleri, sayfa – konu ilişkisini, doküman benzerliği ile sağlamaktadır. Klasik bilgi erişim metotlarına dayalı olan bu benzerlik ölçümleri, aslında sözcüksel benzerliğe dayalıdır. Bu ölçme yapısında, eğer iki doküman aynı

terimleri içeriyorsa bu iki dokümanın ilişkili olduğu kabul edilir. Ancak iki dokümanda da bulunan az sayıda ortak terimler, bu iki dokümanın ilişkisiz olduğu anlamına gelmemektedir. Örneğin; iki terim farklı yazıldıkları halde anlamsal olarak aynı olabilmeleri mümkündür. Klasik ağ örümcekleri; anlamları aynı fakat yazılışları farklı terimler içeren dokümanları ilişkilendirememektedir. Semantik ağ örümcekleri, terim taksonomi ve ontolojisini kullanarak bu probleme çözüm getirmiştir. Doküman benzerliği de Semantic Similarity Retrieval Model (SSRM)[9] modeli ile hesaplanmaktadır.

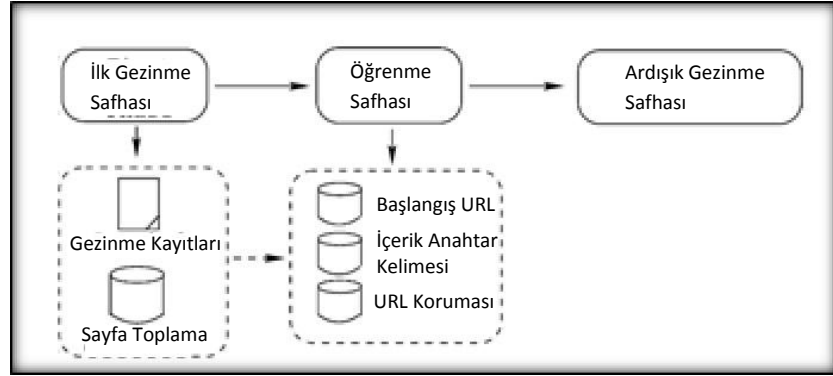
2.2.1.3. Öğrenimli ağ örümceği

Öğrenimli ağ örümceği, örnek web sayfaları kümesi üzerinden kullanıcının önceliğini öğrenir. Spesifik olarak; kullanıcı, web sayfası kümesi oluşturur ve ilgiye göre hangisinin konu ile ilişkili olduğu belirlenir. Böylece öğrenimli ağ örümcekleri, sayfaların ziyaret edilme önceliklerini belirlemede ve gezinme sürecine yol göstermede öğrenim aşaması uygular. Öğrenimli ağ örümcekleri tarafından ilgili sayfalara ulaşmak için ilişkili bağlantılar, bu sürece yol gösterici niteliktedirler. Öğrenimli ağ örümceklerinin öğrendikleri ile web sayfalarını bilinçli şekilde dolaşmaları için üç farklı evreden geçmeleri gerekmektedir. Bu evreler, Şekil 2.3' te gösterilmektedir.

Öğrenimli ağ örümceğinin birinci evresi, ilk gezinme evresidir. Bu evre öncelikli bilgi olmadan web sayfası toplama evresidir. Bu evre süresince, ağ örümceği sadece anahtar kelimeye sahiptir ve bu anahtar kelime ile kullanıcı tarafından tedarik edilmiş ilgili konuyu tanımlar. Daha sonra da ağ örümceği bu anahtar kelimeleri arama motoruna aday başlangıç web sayfası kümesi oluşturması için gönderir. Buradan gelen sonuçlar, ikinci evre olan öğrenim evresinde öğrenim temelini inşa etmek için kullanılır.

İkinci evre olan öğrenme evresinde; ağ örümceği, birinci evrede topladığı web sayfalarının ışığında nasıl daha etkin şekilde, konu ile daha ilgili web sayfalarını toplayabileceğini öğrenmeye çalıştığı evredir.

Üçüncü ve son evre olan ardışık gezinme evresi, ağ örümceğinin kazandığı bilgi ile web sayfalarını indirdiği evredir. Bu evre süresince, başlangıç web sayfası adresleri ve konuya ilişkin anahtar kelimeler öğrenilmiş, web sayfası tahmin bilgi tabanı kullanılarak web sayfalarının sıralama işlemi yapılmaktadır.



Şekil 2.3. Öğrenimli Ağ Örümceği Şeması

2.3. Html Ayırıcı ve Ayırma İşlemi

Şu ana kadar ağ örümceğinin web sayfalarını indirip, sayfa – konu ilişkisi kurması için ayrıca bu sayfaların arama motoru tarafından unutulmaması ve tekrar o sayfayı bularak görüntülemek için indirilen web sayfalarını indekslediğinden söz edilmiştir. Ağ örümceklerinin web sayfalarını indirdikleri kadar onları nasıl ve neye göre anlamlandırdıkları da önemlidir. Bu noktada HTML/XML gibi etiket ayırma işlemi indirilen web sayfasının anlaşılabilir ve işlenebilir hale getirilmesinde devreye girmektedir.

Metin halindeki veriyi ayıran, anlamlı hale getiren programlar ayırıcı olarak tanımlanmaktadır. HTML/XML ayırıcı yazmak zorlu ve birçok kontrolün yapılmasını gerektiren bir işlemdir. Bunu sebebi web sayfasının yapısal içeriğinin yazım kurallarının uyum içerisinde olmamasıdır. Örneğin, web sayfasında bulunan veriler genellikle

HTML olarak bulunmakta olup, bir HTML dokümanında karmaşık başlangıç etiketi, bitiş etiketi, metin ve basit etiketler bulunabilmektedir. Ağ örümceğinin bir web sayfasının işlenebilir hale getirebilmesi için topladığı web sayfalarının HTML etiketlerinden ayrıştırılmaları gerekmektedir. Bu ayrıştırma işlemi de ayrıştırıcı tarafından yapılmaktadır.

HTML sayfalarının yapısal özelliklerinden dolayı ayrıştırılmaları kolay bir işlem olmadığından biraz önce bahsedilmiştir. Ayrıca ayrıştırma işlemi, arama motorlarının görüntüleyecekleri sonuçları da etkileyebileceği için çok dikkatli ve doğru yapılması gerekmektedir. Ayrıştırıcı, aşağıda anlatılan HTML sayfası bileşenlerini tanıyıp, anlayıp bunlar arasından işe yaracak olanları işleme alan bir programdır. Örneğin, geliştirmek istediğimiz ayrıştırıcı bir web sayfasını başka bir dile çevirme işlemi yapacaksa, web sayfasının HTML bileşenlerinden ayrıştırılmış ve sadece metin kısımlarının işleme hazır hale getirilmiş olması gerekir. Arama motorları için kullanılacak bir ayrıştırıcı ise indirilen web sayfasının metin içeriğini, bağlantılarını çok iyi ayrıştırması gerekmektedir. Bu ayrıştırmanın sonunda bağlantılı sayfalar gezilecek, metin içeriğinde anahtar kelimeler aranacaktır.

Bir HTML dokümanı, beş temel bileşenden oluşmakta olup, bu bileşenler Şekil 2.4' te örnek bir HTML dokümanı içeriği ile gösterilmektedir. Bunlar;

- Metin
- Açıklama Satırı
- Basit Etiketler
- Başlangıç Etiketleri
- Bitiş Etiketleri

2.3.1. Metin

Bu bölüm HTML yapısındaki sayfanın, metin olarak görüntülenen kısmıdır. Bir HTML dokümanında, Script' ler haricinde bir etiketin parçası olmayan veri, metin olarak gözlenir. Aşağıda örnek olarak metin, formatlanmakta ve metni kapsayan etiketler ile kontrol edilmektedir.

```
<h1>Merhaba</h1>
```

2.3.2. Açıklama Satırı

Açıklama satırı, kullanıcıya görüntülenmeyen HTML dokümanının tasvir edildiği bölümdür. Bu kısım genellikle, HTML programcısının ondan sonra yazmış olduğu kodları inceleyecek kişi için yardım amaçlı bıraktığı açıklama veya dikkat edilmesi gereken bilgilendirme notlardır. Çünkü açıklama satırları kullanıcı tarafından görüntülenmezler. HTML dokümanında açıklama satırları, aşağıda da görüldüğü gibi <!-- etiketi ile başlar ve --> etiketi ile sonlanır.

```
<!-- Bu bir açıklama satırıdır. -- >
```

2.3.3. Basit Etiketler

Basit etiketler de HTML etiketleridir. Fakat onları diğer etiketlerden farklı kılan özellikleri, yalnız olarak temsil edilmeleridir. Daha açıklayıcı olmak gerekirse, basit etiketler normal HTML etiketlerindeki bitiş etiketine sahip olmayan etiketlerdir. En çok kullanılan basit etiketler arasında alt satıra geçmek için kullanılan (
) etiketi ve resim eklemek için kullanılan () etiketidir. Daha öncede belirtildiği üzere, bu

etiketlerin basit etiketler olarak nitelendirilmelerinin sebebi (</br> ve) etiketlerinin kullanımına ihtiyaç olmamasıdır. HTML dokümanında basit etiketler aşağıdaki gibi gözlenir:

Metnin ilk satışı

Metnin ikinci satırı

Metnin üçüncü satırı

2.3.4. Başlangıç ve Bitiş Etiketleri

Çoğu HTML dokümanı başlangıç ve bitiş etiketlerinden oluşmaktadır. Başlangıç etiketleri basit etiketlere çok benzer şekilde yazılarak başlar. İkisi arasındaki en belirgin fark ise başlangıç etiketini daha sonradan takip eden bitiş etiketidir. Aşağıda da görüldüğü üzere form etiketi başlangıç etiketi, “/form” etiketi de bitiş etiketidir.

```
<form method="post">
```

```
<input type=submit>
```

```
</form>
```

Başlangıç ve bitiş etiketleri arasında belirtilen kod, HTML dokümanın işlevselliğini kontrol etmektedir.

2.3.5. Baęlantılar ile gezinme

Baęlantılar, derinlemesine incelenmesi gereken HTML etiketleridir. Çünkü baęlantıların genel amacı, kullanıcılara dięer sayfaya kolayca geęme ve dokümanlar arasında kolayca gezinme olanaęı saęlamasıdır.

Ařaęıda verilmiř bir baęlantı örneğinde etiketler arasındaki kullanıcı tarafından görüntölenen kısımdır. řüphesizdir ki “href” kısmı, baęlantının kullanıcısının görüntüleyebildięi bilginin az bir kısmıdır. Bilginin ne olduęuna görüntüledięi baęlantıya tıkladıęı zaman ulařacaktır.

```
<a href = “www.trkya.edu.tr”> Trakya Üniversitesi </a>
```

```
<!DOCTYPE <html>
<html dir="ltr" lang="tr-TR">
<head>
<meta charset="UTF-8" />
<title>řİRA SANAT MERKEZİ | SANATIN TEK ADRESİ</title>
<link rel="profile" href="http://gmpg.org/xfn/11" />
<link rel="stylesheet" type="text/css" media="all"
href="http://www.sirasanatmerkezi.com/wp-content/themes/twentyten/style.css" />
<link rel="alternate" type="application/rss+xml" title="řİRA SANAT MERKEZİ &raquo; Beslemesi"
href="http://www.sirasanatmerkezi.com/feed/" />
<link rel="pingback" href="http://www.sirasanatmerkezi.com/xmlrpc.php" />
<link rel="alternate" type="application/rss+xml" title="řİRA SANAT MERKEZİ &raquo; Yorum
Beslemesi" href="http://www.sirasanatmerkezi.com/comments/feed/" />
<link rel="EditURI" type="application/rsd+xml" title="RSD"
href="http://www.sirasanatmerkezi.com/xmlrpc.php?rsd" />
<link rel="wlwmanifest" type="application/wlwmanifest+xml"
href="http://www.sirasanatmerkezi.com/wp-includes/wlwmanifest.xml" />
<link rel="index" title='řİRA SANAT MERKEZİ' href='http://www.sirasanatmerkezi.com/' />
```

řekil 2.4. Örneđ HTML Belgesi

2.4. İndeksleyici

Ağ örümceklerinin, internetteki web sayfalarını gezerek ve bu sayfaları daha sonra kullanıcı ara yüzü ile yapılan aramalarda hatırlayabilmesi için yaptığı işlem indekslemedir. Çünkü arama motorlarının hatırlayabildiği, saklayıp ulaşabildiği web sayfası kadar sonuç görüntüleyebileceği daha önce belirtilmiştir.

İndeksleyici, ağ örümceği tarafından indirilen sayfalarını indeksleyen bir program olmak ile birlikte, aynı zamanda da arama motorunun bir bileşenidir. Ağ örümceğinin en çok yaptığı ve önem arz eden görevi, web sitesinin hangi konu ile ilgili olduğu kararını vermektir. Web indeksleme işlemi, ağ örümceklerinin, arananı kolayca bulmaları için kitapların sonunda bulunan indeks bölümüne benzer bir yapı oluşturmayı amaçlamaktadır. Ayrıca indeksleyici HTML kodlarını önemli kılabilecek özelliklerine göre gözden geçirir. Bu bağlamda koyu, italik kelimelere ve başlık etiketlerine önem verilir.

2.4.1. İndeksleme metotları

Yukarıda söz edildiği üzere ağ örümceklerinin indirdiği web sayfalarının arama motoru tarafından unutulmaması ve aranan konu ile ilgili web sayfasına kolayca ulaşılabilmesi için indekslenmektedirler. İndeksleme işlemi, bu web sayfalarının veri tabanına bütün sayfa olarak değil de onları çağrıştıracak kısa yollar atayarak, bu kısa yolların veri tabanına yazılması olarak da nitelendirilebilir.

Ağ örümceği tarafından indirilen sayfaların indekslenmesi için çeşitli yöntemler kullanılmaktadır. Bahsedilen veri tabanının oluşturulmasında iki önemli metot kullanılmaktadır. Bunlar; tam metin anahtar kelime indeksleme ve insan destekli indekslemedir.

2.4.1.1. Tam metin indeksleme

İsminden de anlaşılacağı gibi web sayfasındaki bütün kelimelerin, arama için veri tabanına kaydedildiği bir indeksleme yöntemidir. Alta Vista, Google, Infoseek, Excite bu tür indeksleme veri tabanına sahip arama motorları arasında verilebilecek örneklerdir[8]. Tam metin indeksleme, belirli bir isme veya terminolojiye verilmiş olan bir örneğin, kolaylıkla bulunmasında yardımcı olmaktadır. Fakat genel bir konu hakkında yapılacak arama, bu veri tabanı için kullanışlı değildir ve birçok aranan ile ilgili ilgisiz arama sonucu görüntülenmektedir. Bunlara ek olarak, günümüzde paylaşılmakta olan çok büyük boyutlardaki elektronik verinin tam metin indekslenmesi de her geçen gün daha da zorlaşmakta olup maliyetleri de katlanarak arttırmaktadır.

2.4.1.2. Anahtar kelime indeksleme

İnternette bulunan sayısal verilerin paylaşımı ve büyüklüğünün artıyor olması ile birlikte, yapılan aramaları kolaylaştırmak amacı ile sayfa içeriklerinin birtakım anahtar kelimeler ile tanımlanması bulunan avantajlı bir yöntemdir. Anahtar kelimeler otomatik olarak üretilebilmekte ya da kullanıcı tarafından belirlenebilmektedir.

Anahtar kelime indeksleme, sadece önemli kelimelerin veya cümlelerin veri tabanına kaydedildiği bir indeksleme türü olup, bu indeksleme işlemi de ağ örümceği tarafından yapılmaktadır. Sayfaların içerik – konu eşleşmesi, diğer ilgili bilgiler beraberinde, veri tabanına ya da sabit diske kaydedilir. Bu indeksleme türü, geniş çaplı arama motorları için en iyi seçim olmasını rağmen, anlamlı sorguları tamamlayabilme özellikleri bakımından kısıtlı durumdadır.

2.4.1.3. İnsan destekli indeksleme

Arama motorlarında genel olarak, indeksleme işlemi ağ örümceğinin bir parçası olarak indeksleyici tarafından yapılır. İnsan destekli indeksleme yönteminde ise web sayfası bir kişi tarafından gözden geçirilir ve o web sayfasını tarif etmek için çok az anahtar cümle veya kelime kullanılır. İnsan destekli indeksleme metodunu kullanan bir indeksleyici, teknoloji ve onun kullanılabilirliğini ikinci plana atmış ve indeksleme gibi konu – sonuç ilgililiğini önemli şekilde etkileyen bir bileşeni, insan gücü ile yapmaya çalışıyor demektir. Bu, kullanıcılara kişi tarafından tarif edildiği için büyük bir avantaj sağlar. Buna ek olarak; bu indeksleme türü, dizin temelli veri tabanlarının gelişmesinde öncü olmuştur. Fakat insan emeği ile yapılan işlerde, her zaman hata payının daha yüksek olduğu da unutulmaması gereken önemli bir ayrıntıdır. İnsanın her zaman dikkat problemi olup, kullanılacak anahtar kelimenin yanlış kullanılması durumunda, o web sayfası için sonuçlar arasında görüntülenmede sorunlar doğuracaktır. Bu aşamada oluşacak böyle bir hatanın veya gözden kaçırmanın düzeltilmesi çok emek gerektirecektir.

3. DİKEY ARAMA MOTORU

Sayısal bilgilerin günden güne hızla artması, bunun yanına teknolojinin de hızlı bir artış içerisinde olması, sayısal bilgilerin daha kolay paylaşılır hale gelmesi sonucunu beraberinde getirmiştir. Buna bağlı olarak, paylaşılan bilgiler arasından aranana ve doğru olanına en kısa zamanda ulaşmak, ciddi zorluklar yaratmaktadır. Bu zorlukların aşılmasına, şu ana kadar bahsedilen indeksleme yöntemleri de yetersiz kalmaktadır. Bahsedilen bu zorluklara çözüm bulmak amaçlı var olan yapılardan daha farklı bir yapı oluşturmak kaçınılmaz olmaktadır. Dikey arama motorları, bu zorluklara çözüm bulmak aşamasında yeni bir çıkış yolu olmuştur. Artık en popüler arama motoru olan “google.com”, sistemine eklemiş olduğu “Diğer” seçeneği ile dikey arama motorunu sistemine entegre etmiştir.

Dikey arama motoru, belirli kullanıcılar ile arama sonuçlarının ilişkilendirilmesi için, belirli konuların web yelpazesinden toplanması olarak nitelendirilebilir. Yani, dikey arama motoru ile belirli konularda arama yaparak, aranan konu ile tamamen ilgili arama sonuçlarının görüntülenmesi olarak açıklanabilir. Böylelikle arama sonuçlarının arasından, aranan konu ile ilgili olmayanları ayıklamak için zaman kaybı önlenmiş olur.

Dikey arama motorları, yeni bir sistem olmak ile birlikte tercih edilmelerinin en etkin sebebi, arama motorlarının kullandığı karmaşık anahtar kelime kombinasyonlarının tersine, bunu kullanmadan web sayfasını konuya özel indeksledikten sonra kaydetmesidir. Bunun sonucunda da aranan konu ile görüntülenen arama sonuçlarının ilgililik derecelerinin yüksek olduğu anlamına gelir. Dikey arama motorlarının, genel arama motorlarında yaşanan zorluklarına çözüm olarak getirilmesinin sebebi de budur. Bilindik dikey arama motorları arasında Kosmix, Scirus, Medstroy, MALDO, Trovit, Libra sayılabilir.

3.1. Dikey Arama Motoru Kullanım Alanları

Daha önce de belirtildiği üzere, arama motorlarının kullanımında karşılaşılan problem olarak da sayılabilecek zorlukları, arama sonuçlarının büyük miktarlarda paylaşılan bilgi ile doğru orantılı olarak çok fazla olması ve görüntülenen sonuçların çok fazla sayıda olup, aranan konu ile ilgili ilgisiz bütün web sayfalarını içermesiydi. Dikey arama motorları, bu zorluklara çözüm olarak sunulan yeni bir yapı olup, belirli konularda arama yaptıklarından dolayı ilgili ilgisiz sonuç kalabalığında kaybolmayı önlemektedir. Sonuç kalabalığının önlenmesi ile birlikte vakit kaybı da önlenerek daha kısa zamanda doğru bilgiye erişim imkânı yaratılmıştır.

Pazarlama şirketleri açısından dikey arama motorları eşi bulunmaz bir fayda sağlayan araç niteliği taşımaktadır. Yapılan araştırmalara göre, dikey arama motoru kullanan ve bu yolla reklam veren bir pazarlama şirketinin maliyeti ile aynı reklamı Google AdWords ile vermesi arasında çok büyük farkın olduğu ortaya konmuştur[10]. Çıkan sonuca göre Google AdWords gibi Google arama motorunun reklam vermek için kullanılan ek özelliği ile reklam vermenin maliyetinin daha yüksek olduğu ortaya çıkartılmıştır[10].

Dikey arama motorları; yapılan belirli konudaki arama sonuçlarına göre kullanıcıların dikkatinin, aradığı konudan başka bir konuya kaymasını önlemesinden dolayı işletmelerin çok büyük şekilde dikkatlerini çekmektedir. Genel arama motorlarından reklam vermenin maliyetinin yüksek olmasından dolayı; işletmeler, dikey arama motorları yardımı ile reklam yapmaya yönelmektedirler. Çünkü bu sayede işletmeler, daha iyi oranlarda anlaşma yapmakta ve bunun sonucunda da doğal olarak daha iyi gelir elde etmektedirler. Yapılan araştırmalara göre 2005 yılında 7,4 milyar dolar arama motorundan pazarlamaya harcanmıştır[10]. Bunun %16 sı şirketler arası e-ticarettir (b2b)[10]. % 40 orta ölçekli pazarlamacılar, neredeyse bütçelerini arama motorlarına vakfetmiştir[10]. Yahoo' nun yaklaşık % 38' i ve Google'ın yaklaşık % 50' si şirketler arası e – ticaret olarak bilinmektedir[10]. Yaklaşık arama motoru kullanıcılarının % 64' ü ilk olarak şirket bilgisi aramaktadırlar[10]. Durum böyle olunca, elektronik reklam vermek giderek popüler bir hal almaktadır. Reklam giderlerinin

şirketlerin giderleri içerisinde önemli yer tuttuğu düşünülür ise, Amazon.com gibi alışveriş sitelerinin her hafta Google gibi arama motoru şirketlerine milyonlarca dolar kendi ürünleri ile ilgili anahtar kelimeler için ödeme yapması da düşünülür ise bu güce sahip olmayan, özellikle daha küçük ölçekli pazarlama şirketleri için, dikey arama motorları katlanarak artan bir gider azaltma kapısı olması bakımından önem teşkil edecektir.

Özetle; en az gider ile en çok kazancı sağlamak olan pazarlama ilkesi, zorlaşan şartlar ve artan giderler ile kolay bir ilke olmamaktadır. Maddi olarak güçlü pazarlama şirketleri, çok büyük meblağları gözden çıkarıp arama motorlarına yatırım yaparak güçlerine güç katmaktadırlar. Fakat orta ve küçük ölçekli pazarlama şirketleri için bu maliyetler karşılanması son derece güç hatta imkânsız miktarlardadır. Bu imkânsızlık; orta ve küçük şirketleri, popüler arama motorları aracılığı ile yapılan reklamlar ile ulaşılacak kitleler kadar çok olmasa bile, onları doyuracak kadar kazanç sağlayan diğer bir çözüm yolu olan dikey arama motorları vasıtası ile reklam verme yoluna itmektedir. Böylece dikey arama motorları, özellikle elektronik ortamda pazarlama imkânı olan şirketler için giderek yayılan bir çözüm yolu olmaktadır.

Buraya kadar anlatılanların dışında, dikey arama motorlarının aranılan bilgiye kısa zamanda ve ilgi oranının arama motorlarına kıyasla çok yüksek olduğu unutulmamalıdır. Ayrıca, dikey arama motorları kullanılarak yapılan aramaların sonucunda; kullanıcıların dikkati, aradıkları bilgi ile ilgili görüntülenen arama sonuçlarından dolayı başka konulara kayması büyük oranda önlenmektedir.

4. YAPAY ÖĞRENME NEDİR?

Bilgisayarların düşünebilen, akıllı makineler olmadıkları gerçektir. Onlara bir işi yaptırabilmek için izlemeleri gereken yolu veya yolları tanımlamak gerekir. Hatta bu yolların herhangi biri bazen tek başına yetersiz kalır. Bu yolların en etkin olanını tanımlamak günümüzde daha önem kazanmaktadır. Bahsedilen bu yolların her birine algoritma denir. Başka bir deyişle algoritma problemi çözenin ya da sonuca ulaşmanın herhangi bir yoludur. Örnek olarak problem veya ulaşılması istenen sonuç olarak, bir dizi sayının sıralanması düşünülebilir. Bu sıralamayı yapabilecek çok sayıda algoritma içerisinde, en az komut ile en az bellek kullanarak en etkin sıralama yaparı tercih edilir.

Bazı problemlerin çözümü için bir algoritma bulunmayabilir veya geliştirmek mümkün olmayabilir. Örnek olarak istenmeyen elektronik postaların temizlendiği bir algoritma verilebilir. Burada amaç gelen elektronik postaların içeriklerine bağlı olarak istenen veya istenmeyen olduğunu belirlemektir. Bu problemi tam olarak çözebilen bir algoritma bulunmamaktadır. Buna ek olarak hangi elektronik postanın istenen, hangisinin istenmeyen olduğu kişiden kişiye değişir. Burada yapılacak iş, derleyeceğimiz istenen ve istenmeyen elektronik postalardan oluşan veri kümesinden bilgisayarın, elektronik postada bulunan hangi özelliğin istenen, hangi özelliğin istenmeyen yaptığını öğrenmesini sağlamaktır. Diğer bir deyişle bilgisayardan bu iş için duruma göre bir yol yani algoritma oluşturmasını istemektir.

Günümüz internet teknolojileri sayesinde bir veriye kolaylıkla ulaşabiliyor ve bu veriyi saklayabiliyoruz. Veri toplamak için kullanılan yolların geneli, bu verileri sayısallaştırmaya dayalı olduğundan, bu özellik bulunan verilerin güvenli şekilde saklanabilmesini ve bir ağ üzerinden rahatça ulaşılabilmesini sağlamaktadır. Birçok mağazası olan marketler zincirinin hangi gün, hangi müşterisinin, hangi ürünü aldığı bilgilerini kayıt altına aldığı bilinen bir gerçektir. Bu, her gün gigabaytlarca veri anlamına gelmektedir. Bu marketler zinciri, müşterilerinin yapmış olduğu alışverişlerini inceleyerek gelecekte aynı müşterinin olası alışverişine ilişkin bir öngörüye sahip olmak isteyebilir. Bu öngörüye sahip olabilmek için o müşterinin yapmış olduğu alış

veriřlerinin incelenip, çok kesin olmasa da, yaklaşık bir öngörüye ulaşabileceğine inanılmaktadır. Çünkü bir müşteri, gömlek alıyorsa yanına kravat alır, kışın kalın kazak alır, yazın ince t-shirt alır. Yani, yaptığı alış veriřler rastgele değildir. Hangi müşterinin ne zaman hangi gömleđi alacağını, hangi tiřörtü veya kazađı alacağını, hangi marketten alış veriř yapacağını, hangi web sayfasını gezip oradan sipariř vereceđini bilemeyiz. Eđer kesin řekilde bunları bilecek olsaydık verileri incelemek gibi bir ihtiyacımız olmazdı. Yaptıđı alış veriřlerdeki yani verideki önemli örüntülerin ve düzenliliklerin sezinlenebileceđi düşünölmektedir[11]. Yapay öğrenmenin bir alanı da bu olmuřtur[11].

Yapay öğrenme, bilgisayarların örnek veri ya da geçmiş deneyimleri kullanarak başarımlarını arttıracak řekilde programlanmasıdır. Elimizde bazı parametrelere bađlı olarak tanımlanmış bir model ve veri ya da geçmiş deneyim üzerine modelin başarımlarını ölçmek için bir ölçüt tanımlıdır[11]. Amaç, modelin parametrelerini bu başarımlar ölçütüne göre en iyi yapan parametre deđerini bulmaktır[11]. Bu zaman kadar bahsedilen model, gelecek ile ilgili bir konu hakkında tahminler yapacak öngörücü ya da mevcut verilerden bilgiye ulaşmak için açıklayıcı bir model olabilir[11].

Yapay öğrenmenin temelinde, gözlenmiş örneklerden çıkarım yapmak yatar. Bundan dolayı da model oluřurmada istatistik kuramı kullanılır. Bilgisayar bilimlerinin ise ilk görevi, büyük miktarlarda olan veriler üzerinde uygulanan ve öğrenme ile ilgili iyileřtirme probleminin etkin řekilde çözölmesidir. Diđer görevi de öğrenilmiş modelin en az bellek kullanmasını sađlayarak etkin bir çıkarım yapmasını sađlamaktır.

5. YAPAY ÖĞRENME UYGULAMALARINA ÖRNEKLER

5.1. İlişkilendirme Kuralları

Yapay öğrenme uygulamasına en basit örnek, bir mağaza için sepet çözümlemesi olarak verilebilir. Bu örneği açıklamak gerekirse, bir gömlek alan müşteri yanında da kravat almışsa ve diğer bir müşteri de gömlek almışsa, bu müşteri kravat almaya aday bir müşteridir.

İlişkilendirme kuralı bulunurken, müşterinin geçmişte yapmış olduğu alışveriş bilgilerini içeren “gömlek” kümesine koşullu olarak “kravat” ürünü için $P(\text{gömlek} | \text{kravat})$ olasılığı öğrenmek amaçlanmaktadır. Mesela mevcut verilerden $P(\text{gömlek} | \text{kravat}) = 0.6$ bulunduğunu düşünelim. Bulunan bu değer, gömlek alan müşterilerin % 60' ı kravat da alır anlamına gelen bir kural çıkartılır. Bu kural ile birlikte müşterilere kişisel bilgiler ekleyerek ve G de müşterilerin gelir düzeyini gösterirken, $P(\text{gömlek} | \text{kravat}, G)$ olasılığını tahmin etmek isteyebiliriz. Verilen mağaza örneğini, bir web sayfasının konferans, kongre, seminer, vb. sınıflarından hangisine ait olduğunu belirlemede, ev eşyası alımında hangi ürünün yanında hangi ek ürünün alınacağına ve daha çeşitli şekillere uyarlanabilmesi mümkündür.

5.2. Sınıflandırma

Bir brokerin, çalışacağı yatırımcılarının hangi tipte olduklarını onları tanımadan tahmin etmekte ve ona göre bir fon yönetim stratejisi belirlemek için kullanacağı bir model düşünelim. Bu modelin içeriği hakkında bilgi vermeden önce, brokerlerin çalıştığı yatırımcı tipleri hakkında bilgi vermek daha doğru olacaktır.

Brokerler, bilindiği üzere borsada yatırımcılarının parası ile fon alımı – satımı yapan ve bu alım – satımdan da yatırımcılarının ona vermiş olduğu parayı

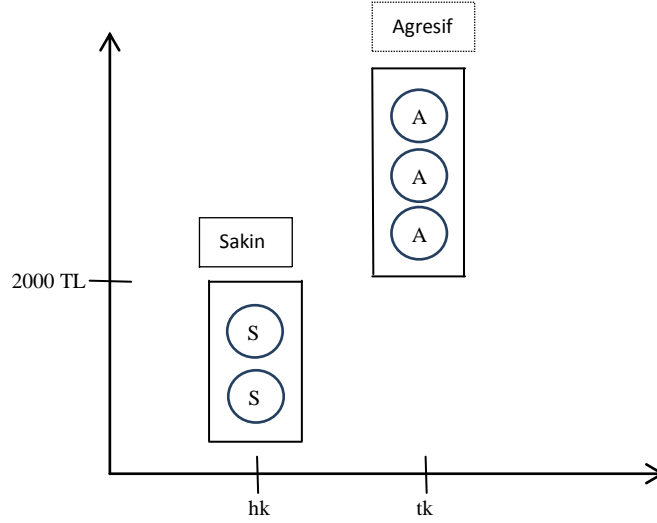
değerlendirerek, yatırımcılarına gelir sağlayan kişilerdir. Diğer ifade ile brokerler, yatırımcılarına borsa aracılığı ile para kazandıran insanlardır. Brokerlerin, para kazandırmayı amaçladıkları yatırımcıları hakkında edindikleri bilgi, onların nasıl bir davranışsal tutum içerisinde alım satım yapmaları gerektiği belirler. Bu davranışsal tutum agresif ya da sakin olarak yatırımcıdan yatırımcıya farklılıklar gösterir. Brokerler, aldığı fondan çok para kazanmak isteyen yatırımcılarının fon yönetimi stratejilerinde agresif tutuma sahiptirler. Bu tür müşteriler çok kazanmadıkları sürece kaybetmeyi göze alan yatırımcı tipidir. Bu yatırımcı tipi ya hep ya hiç kazanma mantığına sahiptir. Broker, bu tip yatırımcısı için risk oranı daha yüksek, sürpriz olarak nitelendirilen fonlara yatırım yapar. Eğer brokerin sahip olduğu yatırımcı, bir anda çok kazanmak yerine, az az ama sürekli kazanmayı hedefleyen bir mantığa sahipse, broker için bu yatırımcı fon yönetimi stratejisinde sakin bir tutuma sahiptir. Broker, bu tip yatırımcısı için risk oranı daha düşük, o günkü şartlar altında yükselebileme oranı fazla olan fonlara yatırım yapar.

Daha önce bahsedilen modelin, brokerlerin sahip oldukları yatırımcıları için izleyecekleri fon yönetimi strateji tipini, agresif ya da sakin olarak tahmin edebilecek bir model olduğunu düşünelim. Bu iki sınıflı bir sınıflandırma problemidir. Sınıflar, agresif fon yönetimi stratejisi uygulanacak yatırımcılar ve sakin fon yönetimi stratejisi uygulanacak yatırımcılardır. Bu uygulamada, brokerin daha önce sahip olduğu yatırımcılarının ona teslim ettikleri para miktarı, kazanma mantıkları kısacası yatırımcı bilgileri yani daha önceki tecrübeleri modelin girdisidir. Bu girdiye bağlı olarak iki sınıftan biri seçilir.

Şekil 5.1’ de[11] sınıfların nasıl ayrıldığı ve örnek veri kümesini görmekteyiz. “m” değeri, yatırımcının brokere teslim edeceği para miktarını gösterirken, “hk” değeri, yatırımcının yatırımı yaparken ki kazanma mantığının az az ama hep kazanmak olduğunu, “tk” değeri de yatırımcının yatırımı yaparken ki kazanma mantığının bir kerede çok fazla kazanmak olduğunu göstermektedir. Brokerin edindiği tecrübeler ile öğrenilmiş bir sınıflandırma kuralı,

EĞER Miktar > m **VE** Mantalite = tk **İSE** agresif strateji **DEĞİLSE** sakin strateji

şeklinde yazılabilir.



Şekil 5.1. Fon Yönetim Stratejisi Belirleme Modeli

5.3. Bağlanım

İkinci el buzdolabı fiyatlarını öngören bir uygulama tasarlamaya çalıştığımızı düşünelim. Buzdolabının fiyatını etkileyebilecek özellikler, ne kadar elektrik harcadığı, markası, iç hacmi, özel bölme sayısı, dolabın içerisinde bulunan yiyeceklerin kokularının birbirine karışmasını engelleme özelliğinin bulunması, vs. uygulamanın girdileridir. Aynı uygulamanın çıktısı da buzdolabının fiyatıdır. Bu uygulamada da olduğu gibi, çıktının bir sayı olduğu uygulamalara bağlanım uygulaması denir.

Buzdolabının özelliklerini X, fiyatını da Y ile gösterelim. Geçmişte alınıp satılmış buzdolaplarından bir öğrenme veri kümesi oluşturup bu tecrübe edilmiş verileri

kullanarak Y değerini X değerlerinin bir işlevi olarak öğrenmek istemekteyiz. Şekil 5.2' de[11], örnek bir işlev verilmektedir. Doğrusal olan bu model μ ve μ_0 değerleri için,

$$y = \mu x + \mu_0$$

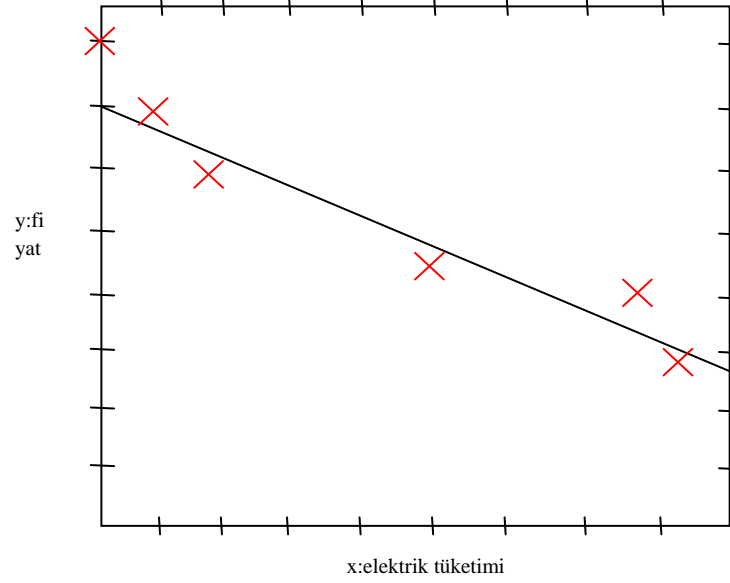
şeklinde yazılır. Verilen bu örnekte tek girdi olarak buzdolabının ne kadar elektrik harcadığıdır.

Gözetimli öğrenme uygulamaları, X değişkeninin girdi, Y değişkeninin çıktı olduğu ve amacının da girdiden çıktıya bir eşleme öğrenmek olduğu uygulamalardır. Yapay öğrenme yöntemleri kullanılarak, çıktıyı, girdiyi kullanan ve bazı parametrelere göre tanımlı bir model ile yaklaşık olarak öngörebilmek amaçlanır:

$$y = g(x|\theta)$$

Yukarıdaki formülde görülen model $g(\cdot)$ sembolü, parametreleri de θ sembolü ile gösterilmektedir. Y ile gösterilen ve uygulamanın çıktısı olmak ile birlikte bağlanım uygulamasında bir sayı, sınıflandırma uygulamasında ise 0 ya da 1 değerini alan sınıf etiketidir. $g(\cdot)$ ile gösterilen model, bağlanım uygulamalarında kullanılan bağlanım işlevi ya da sınıflandırma uygulamalarında kullanılan farklı sınıfları ayıran ayıraç işlevine sahiptir. Yapay öğrenme uygulaması, belirlemiş olduğu öngörü ile gerçek çıktılar arasındaki farkı, yani hatasını en aza indirecek parametreyi bulmaya çalışır. Şekil 5.2' de[11] model doğrusal olarak verilmektedir. Buna ek olarak μ ve μ_0 değerleri öğrenim verisine çok yakın geçecek şekilde bulunmaktadır. Böyle doğrusal modellerin çok kısıtlayıcı kaldığı durumlarda ikinci ya da daha da yüksek dereceden denklemler kullanılarak çıktı ön görülmeye çalışılır.

$$y = \mu x^2 + \mu_1 x + \mu_2$$



Şekil 5.2. İkinci El Buz Dolabı Fiyatları İçin Örnek Veri ve Yaklaşık Bir Model

5.4. Gözetimsiz Öğrenme

Gözetimli öğrenmede amaç, girdi ile doğru değeri bir gözetmen tarafından verilen çıktı arasında bir eşleme öğrenmektir[11]. Gözetimsiz öğrenmenin gözetimli öğrenmeden en büyük farkının gözetimli öğrenmede bulunan gözetmenin olmamasıdır. Gözetimsiz öğrenmede elimizde sadece girdi verileri bulunmaktadır. Gözetimsiz öğrenmede amaç, girdideki düzenlilikleri bulmaktır[11].

5.5. Pekiřtirmeli Öğrenme

Bazı uygulamaların çıktıları bir sayı veya sayılar dizisi olabileceđi gibi bazı uygulamaların çıktıları eylem dizisi olabilir. Bu tip uygulamalarda önemli olan tek bir eylem deđil, sonuca ulaşmak için kullanılacak eylemler dizisinin iyi bir şekilde belirlenmiş olamamasıdır. Böyle bir durum karşısında yapay öğrenme uygulaması, sonuca ulaşmak için kullanılacak eylem dizisinin yani politikanın geçmiş eylemlerden tecrübe ettiđi politikalardan iyi bir politika üretmesi beklenmektedir. Anlatılan bu türdeki öğrenmeye, pekiřtirmeli öğrenme denir.

Bu öğrenme tipine verilebilecek en uygun örnek satranç oyunudur. Satranç oyununun kuralları azdır. Kurallarının az olmasına karşılık herhangi bir durumda birçok hamle yapabilme seçeneğinin bulunmasından dolayı karmaşıktır. Bu yüzden, satranç oyununda tek bir hamle deđil, doğru hamleler dizisi yapmak, oyunu kazanma veya kaybetme sonucunda belirleyici özelliđe sahiptir. Eğer bir hamle, doğru ve iyi bir politikanın bir parçası ise, o hamle sonuca ulařtıran, iyi bir hamledir. Dünyaca ünlü, iyi satranç oyuncularının oyun politikalarını öğrenebilen algoritmalar geliştirildiđi takdirde, bu algoritmaları gerçek hayatta, gerçek uygulamalarda da kullanılabiliriz.

6. GELİŞTİRİLEN YAZILIMDA KULLANILAN YARDIMCI ARAÇLAR

6.1. WEKA Nedir?

WEKA, Yeni Zelanda' nın Waikato Üniversitesi' nde geliştirilmiş, Java programlama dili kullanılarak yazılmış, popüler bir makine öğrenme yazılım paketidir. WEKA, görselleştirme araçları, veri analizi için algoritmalar ve tahmin modelleme içeriklerine sahip olup, bütün bu içeriklere kolayca ulaşabilmek amaçlı bir kullanıcı arayüzüne de sahiptir. WEKA'nın Java olmayan orijinal sürümünü, diğer programlama dillerinde uygulanan TCL / TK ön uç modelleme algoritması, C programlama dilinde veri ön işlem yardımcı uygulaması ve çalışan makine öğrenme tecrübeleri için dosya üretme tabanlı bir sistem idi. Bu orijinal sürüm, tarımsal alanda alınan verileri analiz etmek amaçlı üretilmiş bir araç olarak tasarlanmış idi. Fakat geliştirilmeye 1997 yılında başlanmış olan daha yeni Java tabanlı tam sürüm (WEKA 3), birçok eğitimsel amaçlı ve akademik araştırmalar gibi farklı uygulama alanlarında kullanılmaktadır.

WEKA kullanımının birtakım avantajlarının bulunduğu kuşkusuzdur. Bu avantajlar; GNU lisansı altında serbestçe kullanılabilirliği, tamamen Java programlama dili ile uygulanabilir hale geldiğinden beri hemen hemen her modern programlama platformu üzerinde çalışması, kapsamlı veri ön işleme ve modelleme teknikleri içermesi ve kullanıcı ara yüzleri nedeni ile kolay kullanımı olarak sayılabilirler.

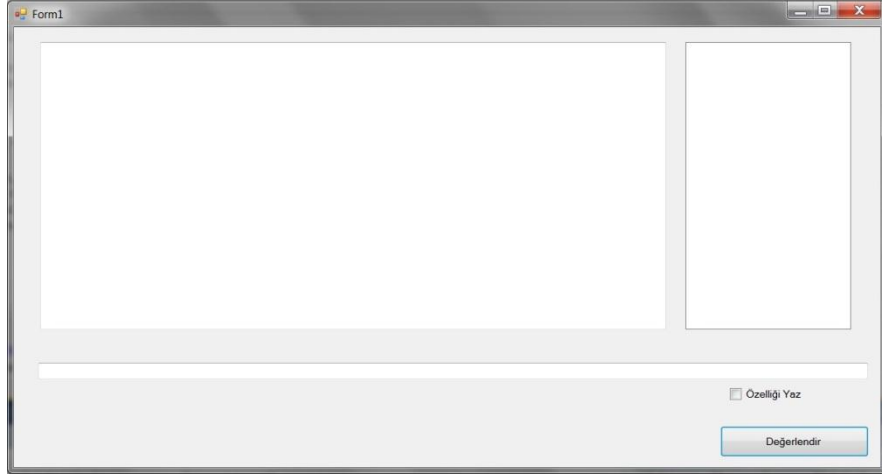
WEKA, birtakım veri madenciliği işlemlerini içerir. Daha spesifik olarak bunlar, veri ön işleme, kümeleme, sınıflandırma, regresyon, görüntüleme ve özellik seçimidir. WEKA'nın sahip olduğu bütün teknikler, verilerin tek bir düz dosya ya da ilişkili olarak kullanılabilir olduğu varsayımına dayanmaktadır. Dolayısı ile WEKA, Java veri tabanı bağlantısını kullanarak SQL veri tabanına erişimi ve veri tabanı sorgusu tarafından döndürülen sonucun işlenmesini desteklemektedir. Bu çok ilişkili veri madenciliği için uygun değildir fakat birbiri ile ilişkili veri tabanı tablolarını tek bir tabloya dönüştüren ayrı bir yazılım bulunmaktadır. WEKA bu ayrı yazılımın yaptığı işlemi yapabilmektedir.

6.2. WEKA İle Sınıflamadan Önceki Hazırlıklar

Daha önce de bahsedildiği üzere WEKA'nın içerdiği veri madenciliği işlemlerinden bir tanesi de sınıflandırmadır. Geliştirilmiş olunan tez çalışmamın uygulama bölümünde, WEKA'nın sahip olduğu bu özelliklerin kütüphaneleri uygulamaya eklenerek yararlanılmıştır. Bu özellikten yararlanmak için arff (Attribute-Relation File Format) uzantılı dokümana eğitim seti oluşturuldu ve incelenen metnin eğitim seti dokümanı yardımı ile kendisine ait sınıflandırma algoritmasına göre sınıflandırma işlemini tamamlaması sağlanmıştır. Bu eğitim seti ile bilgisayara verilen kriterler doğrultusunda incelenen metnin hangi durumlarda hangi sınıfa ait olacağını öğretilmesi amaçlanmıştır.

Bu eğitim setinin oluşturulması aşamasında, arff uzantılı eğitim dokümanı oluşturması için Visual Basic.NET programlama dilinde yardımcı uygulama geliştirilmiştir. Geliştirilen dikey arama motorunda aranacak olan konular ile ilgili haberlerin sınıflandırılmasında kullanılacak eğitim seti dokümanının oluşturulması, incelenen metnin içerisinde bahsedilen sınıflara ait kelimenin en çok sıklıkla geçme oranına bakılarak düzenlemek amacı ile bu yardımcı uygulama kullanılmıştır.

Yardımcı uygulamanın adı "MetinSınıflandırma" dır. Şekil 6.1' de bu uygulamanın arayüzü gösterilmiştir. İnternette bulunan ve arama motoru tarafından incelenecek web sayfalarının sınıflandırmasında kullanılacak yapay öğrenme mekanizması için gereken eğitim setinin oluşturulmasında, bu uygulama kullanılmıştır. Şekilde 6.1' de de görüldüğü üzere arayüz, taranan metnin girilip, incelenmesi için bir text alanına sahiptir. Eğer sıfırdan bir eğitim dokümanı hazırlanıyor ise "Özelliği Yaz" check box alanı işaretlenmeden eğitim dokümanının başlangıç için gereken alanları ve metin kutusunun içinde bulunan metin için bir eğitim satırı "Değerlendir" butonuna basıldığında "Eğitim.arff" dosyasına yazılır. Daha sonraki metinler için eğitim satırı eklemek için, önceki eğitim satırlarının kaybolmasını önlemek amacı ile, "Özelliği Yaz" kontrol kutusu işaretlenerek "Değerlendir" butonuna basılarak eğitim setinin genişletilmesi sağlanmış olur.



Şekil 6.1. Metin Sınıflandırma Uygulaması Arayüzü

“Değerlendir” butonuna basıldığında, bu butonun altında çalışan kodlar ve çağırılan fonksiyonlar Şekil 6.2’ de gösterilmektedir. Örnek kaynak kodda da görüldüğü üzere “VarmiKontrol”, “EgitmenSatiriOzellikYazdir”, “EgitmenSatiriYazdirSonuc” fonksiyonları çağırılmaktadır. Bu fonksiyonların detaylı açıklamalarına ilerleyen bölümlerde değinilecektir. Daha önce de değinildiği üzere “Özelliği Yaz” kontrol kutusu işaretlendikten sonra “Değerlendir” butonuna basıldığı takdirde “EgitmenSatiriOzellikYazdir” fonksiyonunun çağırıldığını, bu kontrol kutusu işaretlenmeden butona basıldığında “EgitmenSatiriYazdirSonuc” fonksiyonunun çağırıldığını Şekil 6.2’ de görmekteyiz. Bu kontrol kutusunun işaretlenme durumuna göre neden ayrı fonksiyonların çağırıldıkları açıklanmıştı.


```

PrivateSub DeğerlendirCmd_Click(sender As System.Object, e As System.EventArgs) Handles DeğerlendirCmd.Click
Dim kelime AsString
Dim Metin AsString = MetinTxt.Text
Dim i, uzunluk, sonraki, KelimeSayac AsInteger
Dim onceki AsInteger = 1
    uzunluk = Len(Metin)
For i = 1 To uzunluk
If Mid(Metin, i, 1) = " "Then
    sonraki = i
    kelime = Mid(Metin, onceki, (sonraki - onceki))
For KelimeSayac = 0 To OgreticiKelimeListesiAdet - 1
If VarmiKontrol(KelimeListesi(KelimeSayac).Kelime, kelime) Then
    KelimeListesi(KelimeSayac).Adet += 1
EndIf
Next KelimeSayac

    onceki = sonraki
EndIf
Next i
If OzellikYazChk.Checked Then
    EgitmenSatiriOzellikYazdir()
Else
    EgitmenSatiriYazdirSonuc()
EndIf
EndSub

```

Şekil 6.2. “Değerlendir” Butonu Altında Çalışan Kaynak Kodlar

Geliştirilen dikey arama motorunda, görüntülenecek sonuçların birbirinden ayırmak ve onları doğru sınıflamak için eğitim seti için anahtar kelimeler belirlenmiştir. Bu belirlenen anahtar kelimeler için structure yapısı oluşturulmuştur. Bu yapıda, kelimenin metin değerinin bulunduğu kısım, değerlendirme aşamasında bir sınıf olup olmadığını belirtmek amaçlı kullanılacak tür kısmı ve incelenen web sayfasının içerisinde kaç adet geçtiği değerini tutmak amaçlı adet kısmı bulunmaktadır. Anahtar kelimenin, değerlendirme aşamasında sınıf olduğunu belirtmek için, o anahtar kelimenin tür değerine “true” atanmıştır. Anahtar kelimenin, değerlendirme aşamasında sınıf olmadığını belirtmek için ise, o kelimenin tür değerine “false” atanmıştır. Tür değerleri, kullanılacak yapay öğrenme mekanizmasının yapacağı sınıflandırmada belirleyici olduğuna karar verilen kelimeler olup, çeşitli seminer, kongre, konferans, vb. haberleri incelenerek seçilmişlerdir. Şu ana kadar anahtar kelimelerin, structure yapısı ve aldıkları değerler ile ilgili bahsedilen kısım Şekil 6.3’ te gösterilmektedir.

```

StructureKelimeListesiYapi
Public Tur AsBoolean
Public Kelime AsString
Public Adet AsInteger
EndStructure

KelimeListesi(0).Kelime= "seminer" : KelimeListesi(0).Tur = True
KelimeListesi(1).Kelime = "sempozyum" : KelimeListesi(1).Tur = True
KelimeListesi(2).Kelime = "konferans" : KelimeListesi(2).Tur = True
KelimeListesi(3).Kelime = "kongre" : KelimeListesi(3).Tur = True
KelimeListesi(4).Kelime = "etkinlik" : KelimeListesi(4).Tur = False
KelimeListesi(5).Kelime = "panel" : KelimeListesi(5).Tur = True
KelimeListesi(6).Kelime = "çalıřtay" : KelimeListesi(6).Tur = True
KelimeListesi(7).Kelime = "diđer" : KelimeListesi(7).Tur = True
KelimeListesi(8).Kelime = "katılım" : KelimeListesi(8).Tur = False
KelimeListesi(9).Kelime = "toplantı" : KelimeListesi(9).Tur = False
KelimeListesi(10).Kelime = "oturum" : KelimeListesi(10).Tur = False
KelimeListesi(11).Kelime = "tarih" : KelimeListesi(11).Tur = False
KelimeListesi(12).Kelime = "bilimsel" : KelimeListesi(12).Tur = False
KelimeListesi(13).Kelime = "sunum" : KelimeListesi(13).Tur = False
KelimeListesi(14).Kelime = "bildiri" : KelimeListesi(14).Tur = False
KelimeListesi(15).Kelime = "ulusal" : KelimeListesi(15).Tur = False
KelimeListesi(16).Kelime = "uluslararası" : KelimeListesi(16).Tur = False

```

Şekil 6.3. Kelime Listesi Kaynak Kodları

Yardımcı uygulama olarak geliştirilmiş “MetinSiniflandirma” uygulamasının anahtar kelime seçimi ve anahtar kelimelerin alacağı değerler ile ilgili bilgilerden daha önce bahsedilmiştir. Bir kelimenin, incelenmekte olan ve daha önce Şekil 6.1 ile gösterilmiş “MetinSiniflandirma” uygulamasının ara yüzündeki metin kutusu alanında bulunan metnin içerisinde bulunup bulunmadığını geri döndüren fonksiyon Şekil 6.4’ te gösterilmektedir. Bu fonksiyon, daha sonra uygulamanın mevcut diğer fonksiyonları tarafından çağırılarak metnin içerisinde bulunan ve benim tarafımdan belirlenmiş olan anahtar kelimelerin o metin içerisinde bulunup bulunmadıklarını geri döndürmektedir. Eğer kelime, metnin içerisinde bulunuyorsa o kelimeye ait adet değeri “Değerlendir” butonu altında çalışan kod parçası ile belirlenmektedir..

```

PublicFunction VarmiKontrol(Kelime1 AsString, Kelime2 AsString) AsBoolean

Kelime2 = LCase(Trim(Kelime2))
Dim TutanKarakterAdet = 0, uzunluk AsInteger

If Len(Kelime1) >= Len(Kelime2) Then
    uzunluk = Len(Kelime2)
Else
    uzunluk = Len(Kelime1)
EndIf
If uzunluk > 0 Then
    For i = 1 To uzunluk
        If Mid(Kelime1, i, 1) = Mid(Kelime2, i, 1) Then
            TutanKarakterAdet += 1
        EndIf
    Next

    If uzunluk = TutanKarakterAdet ThenReturnTrue
    ReturnFalse
EndIf

EndFunction

```

Şekil 6.4. Kelime Kontrolü Örnek Kaynak Kodu

Daha önce “EgitmenSatiriOzellikYazdir” ve “EgitmenSatiriYazdirSonuc” fonksiyonlarının ayrı ayrı çağırılma sebebine değinilmiştir. Şimdi aralarındaki farka daha ayrıntılı olarak değinilecektir. Bu incelemeye “EgitmenSatiriOzellikYazdir” fonksiyonu ile başlamak gerekirse, uygulamanın “Egitim.arff” adında bir dosya oluşturduğunu görürüz. Bu dosya, arama motorunda kullanacağımız ve WEKA kütüphanesi fonksiyonlarının vereceği karar için gerekli olan ve hangi durumlarda ne karar vermesi gerektiğinin öğretildiği eğitim setidir. Ayrıntılı şekilde anlatılmaya çalışılan “EgitmenSatiriOzellikYazdir” fonksiyonunun görevi, bu eğitim seti dokümanının WEKA standartları ile oluşturulması ve “MetinSiniflandirma” uygulamasının metin kutusu alanında incelenen metnin eğitim satırının oluşturulmasıdır. Bu noktada fonksiyonun kodları incelendikten sonra WEKA standartında “arff” uzantılı dokümanın içeriği anlatılacaktır. “EgitmenSatiriOzellikYazdir” fonksiyonuna geri dönecek olursak bu fonksiyon ile ilgili kodlar Şekil 6.5’ te gösterilmektedir. Bu fonksiyonda, “Egitim.arff” dokümanı için gerekli ve “@relation” kısmına isim verildikten sonra, belirlenmiş kelimeler ve bu kelimelerin eğitim satırında alacağı olası değerler yazdırılmaktadır. Geliştirilen uygulamada, kelimelerin olası değerleri “true” ya

da “false” olarak belirlenmiştir. Bu bölümün dokümana yazımı sonlandıktan sonra, “@attribute” kısmının yazım aşamasına geçilir. “@attribute” kısmında, belirlediğimiz ve uygulamada oluşturulan kelime listesinin tür değerine “true” verilen kelimeler yazılır. Bilindiği üzere tür değeri “true” olan kelimeler karar verme aşamasında sınıf olarak nitelendirileceklerdir. Geliştirilmiş uygulama için gerekli sınıflar “seminer”, “sempozyum”, “konferans”, “kongre”, “panel”, “çalistay” ve “diger” olarak belirlenmiştir. Türler ile ilgili dokümana yazım kısmı da tamamlandıktan sonra WEKA’ nın sınıflandırmasında karar vermesi için önemli olan “@data” kısmının yazılmasına geçilmektedir. “@data” kısmı sınıflandırmada önemli bir kısım olup tecrübe edilmiş geçmiş verilerin incelenip, öğrenilerek karar vermenin etkin şekilde gerçekleşebilmesi için gereken bölümdür. Karar verme aşamasında “@data” kısmına yazılacakların ve yazılan eğitim satırına ait sınıfın doğru belirlenmesi daha sonraki aşamalardaki sınıfın ne kadar doğru belirleneceğini etkileyeceğinden, önem derecesi büyüktür. Bu eğitim satırı, “MetinSınıflandırma” uygulaması ile incelenen metnin kelimeleri içerisinde, eğitim seti için belirlenmiş olan anahtar kelimelerin geçip geçmediğinin yazılı olduğu ve bu kelimelerin en fazla sıklıkla yer alana ait sınıfının belirlenip yazıldığı içeriğe sahiptir. WEKA’ nın standartlarına uygun “arff” uzantılı dokümanların içeriği ile bilgi Bölüm 6.3.1’ de verilmektedir.

“EgitmenSatiriOzellikYazdir” fonksiyonundan bahsettikten sonra, “EgitmenSatiriYazdirSonuc” fonksiyonuna ait kodlar Şekil 6.6’ da gösterilmiştir. “Özelliği Yaz” kontrol kutusu işaretlenmediğinde, bu fonksiyonun çağırıldığından bahsedilmiştir. “EgitmenSatiriYazdirSonuc” fonksiyonu ile “EgitmenSatiriOzellikYazdir” fonksiyonu arasındaki tek fark, “EgitmenSatiriOzellikYazdir” fonksiyonunda oluşturulan “@relation” ve “@attribute” bölümlerinin oluşturulmayıp eğitim satırının oluşturulmasıdır. Bu farkın dışında “EgitmenSatiriOzellikYazdir” fonksiyonunu açıklarken anlatılanların hepsi “EgitmenSatiriYazdirSonuc” fonksiyonu için de geçerlidir.

```

PublicFunction EgitmenSatiriYazdirSonuc()

Dim KelimeSayac, EnYuksekFrekansliKelimeIndisi AsInteger
FileOpen(1, "D:\CR_Doküman\Yüksek Lisans\Trakya Üniversitesi\Tez ve Seminer\Tez\CANDATA\Egitim.arff",
OpenMode.Append)
Print(1, vbNewLine)
For KelimeSayac = 0 To OgreticiKelimeListesiAdet - 1
If KelimeListesi(KelimeSayac).Adet > 0 Then
If KelimeListesi(KelimeSayac).Adet > KelimeListesi(EnYuksekFrekansliKelimeIndisi).Adet And
KelimeListesi(KelimeSayac).Tur Then
EnYuksekFrekansliKelimeIndisi = KelimeSayac
EndIf
Print(1, "TRUE,")
Else
Print(1, "FALSE,")
EndIf
Next
If KelimeListesi(EnYuksekFrekansliKelimeIndisi).Kelime <>"diger"Then
Print(1, KelimeListesi(OgreticiKelimeListesiAdet - 10).Kelime)
Else
Print(1, KelimeListesi(EnYuksekFrekansliKelimeIndisi).Kelime)
EndIf

FileClose(1)
EndFunction

```

Şekil 6.5. “EgitmenSatiriOzellikYazdir” Fonksiyonu Kaynak Kodu

6.3. WEKA İle Sınıflama

Şu ana kadar WEKA kütüphanesi içinden kullanılacak sınıflama algoritmaları için gereken “arff” uzantılı dokümanın oluşturulması için gereken yardımcı uygulamanın bölümlerinden bahsedildi. WEKA kütüphanesi içinden kullanılacak sınıflama algoritmalarının anlatımına geçmeden önce, bu algoritmalar için gereken “arff” uzantılı dokümanın içeriğinden bahsetmek daha faydalı olacaktır.

6.3.1. ARFF dokümanının içeriği

ARFF dosyası nitelik – ilişki dosya formatı anlamına gelip, öznitelik kümesinden elde edilen örnekler listesinden oluşan Bilgi Değişimi İçin Amerikan Standartları Kodlama Sistemi (ASCII) metin dosyasıdır. Şekil 6.6’ da örnek bir “arff” uzantılı dosya örneği görülmektedir. Bu dosyanın bölümlerini detaylıca inceleyelim.

Bir konu üzerine kompozisyon yazarken veya Microsoft Word kullanılarak bir çalışma yaparken konuları birbirinden ayırmak için başlık kullanılır. “arff” uzantılı dosya içinde bu geçerlidir. “arff” uzantılı dosyalarda başlık, ilişki (@relation) ve öznitelik (@attribute) ifade etmek için kullanılır. Şekil 6.6’ da da görüldüğü üzere, “arff” uzantılı dosyalarda “@relation”, “@attribue” ve “@data” bölümleri için başlık kullanılmaktadır.

İlişki adı, “arff” uzantılı dosyaların ilk satırında ifade edilir. İlişki adı, string veri tipinde olmak ile birlikte, boşluk içeriyor ise tırnak içerisine alınmalıdır. İlişki adının ifade edilme şekli Şekil 6.6’ nın ilk satırında görülmektedir.

Şekil 6.6’ ya bakıldığında “@relation” ifadesinden sonra gelen ikinci kısım, öznitelik kısmıdır. Öznitelik kısmı, öznitelik ifadelerinin sıralı biçimde ifade edildiği kısımdır. Her özniteliğin kendine ait öznitelik ifadesi bulunmalıdır. Şekil 6.6’ nın öznitelik kısmı incelendiğinde, sadece bir tane “seminer” adında öznitelik adının bulunduğu gibi her özniteliğin tek kendisine ait öznitelik adı ve veri tipi bulunmalıdır. Eğer öznitelik adında boşluk kullanılacak ise o öznitelik adı tırnak içerisinde ifade edilmelidir. Özniteliklerin sırası, “arff” uzantılı dosyanın veri bölümünün kolon pozisyonunu işaret etmektedir. Örneğin Şekil 6.6’ da bulunan ilk veri satırının ilk virgüle kadar olan değeri, öznitelikler kısmında bulunan “seminer” özniteliğinin değerini göstermektedir. Bu, “arff” uzantılı dosyada bulunan bütün veri satırları için geçerlidir. Şu ana kadar özniteliklerin alabilecek adın nasıl olması gerektiği ve özniteliklerin veri satırında nerede işaret edildiği ile ilgili bilgi verilmiştir. Özniteliklerin ifade edilirken alacakları değerlerinde belirtilmesi gerektiğinden bahsedilmiştir. Fakat hangi veri türleri alabileceklerinden bahsedilmedi.

Öznitelikler, dört farklı veri tipinde ifade edilebilirler. Bunlar, sayısal (numeric), nominal – belirtme (nominal-specification), metin (string) ve tarih (date) veri tiplerinden biri olabilir. Bu veri tipleri WEKA yazılımı tarafından desteklenen veri tipleridir. Sayısal, metin ve tarih veri tipleri büyük küçük harfe duyarlıdır. Şekil 6.6’ da görüldüğü üzere, veri tipi, özniteliğin adının yanında yer alan kısmıdır. Sayısal veri tipi değeri, gerçek veya tam sayı olabilir. Nominal – belirtme değeri, olası değerlerin sıralanması ile sağlanır. Değerler boşluk içeriyor ise tırnak içerisinde yazılması gerekmektedir. Metin veri tipi değeri, isteğe bağlı metinsel değere sahip öznitelikler oluşturmamızı sağlar. Bu veri tipi özellikler metin madenciliği uygulamalarında büyük kolaylık sağlamaktadır. Metin veri tipinde özniteliklerden oluşan veri seti oluşturup, metni filtreleyecek uygulamalar WEKA yardımı ile yazılabilmektedir. Son olarak özniteliklerin sahip olabileceği veri tipi tarih veri tipidir. Tarih veri tipi, tarih değerinin nasıl ayrıştırılacağını ve yazdırılacağını seçimli şekilde belirtildiği veri tipidir. Varsayılan tarih formatı, ISO-8601 formatı olarak kabul edilir. Bu format ise birleştirilmiş tarih ve saat formatıdır (yyyy-MM-dd'T'HH:mm:ss).

Özniteliklerin sahip olabilecekleri veri tiplerinden bahsettikten sonra, “arff” dosyasının son bölümü olan veri bölümünden bahsedelim. Bu bölüm, “@data” ifadesinin bulunduğu tek satırdan ve asıl örnek satırlarından meydana gelmektedir. Bu bölümde her örnek, tek satır halinde sunulur. Her örneğin sonunda da satır başı yapılarak bir sonraki örnek gösterimi başlatılır. Her örnek için öznitelik değerleri virgüller ile ayrılmış olup, öznitelik kısmında yazılmış olan özniteliklerin sırası ile aynı şekilde yazılırlar.

```

@relation Bilimsel

@attribute seminer {TRUE, FALSE}
@attribute sempozyum {TRUE, FALSE}
@attribute konferans {TRUE, FALSE}
@attribute kongre {TRUE, FALSE}
@attribute etkinlik {TRUE, FALSE}
@attribute panel {TRUE, FALSE}
@attribute çalistay {TRUE, FALSE}
@attribute diger {TRUE, FALSE}
@attribute tur {seminer,sempozyum,konferans,kongre,panel,çalistay,diger}

@data
FALSE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,sempozyum
FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,konferans
FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,konferans
FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,panel
TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,diger
FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,kongre

```

Şekil 6.6. ARFF Uzantılı Dosya Örneği

6.4. WEKA'nın Kullandığı Sınıflandırma Algoritmaları

Geliştirilen uygulamada, WEKA'nın kullandığı iki sınıflandırma algoritması kullanılmıştır. Bunlar, karar ağacı ile sınıflandırma ve en yakın k komşu kestirimi ile sınıflandırma algoritmalarıdır.

6.4.1. En yakın k komşu algoritması ile karar verme

Bu algoritma, veri madenciliği alanına ait olup bellek tabanlı bir algoritmadır. Bu yöntem, örnek kümesindeki gözlem değerlerinden yararlanır. Bu örnek kümesinin de sınıfları bellidir. Bu yöntemin amacı ise örneğe katılacak yeni gözlemin hangi sınıfa ait olduğunu belirlemektir. Bu yöntemde, örnekler "n" boyutlu uzayda bir nokta olarak alınır ve verilen noktaya en yakın komşuların sayısı olan "k" parametresi belirlenir. Bu

yöntem, uzaklık hesaplama üzerine kurulu olduğu için verilen noktaya diğer tüm noktaların uzaklıkları tek tek hesaplanır. Bu hesaplama işlemi öklid bağıntısı ile yapılır. Hesaplanan uzaklık değerlerine göre satırlar sıralanır ve en küçük “k” tanesi seçilir. Seçilenlerin hangi kategoriye ait oldukları belirlenir ve en çok tekrarlanan kategori değeri seçilir. Son olarak da seçilen bu kategori, tahmin edilmesi beklenen gözlem değeri kategorisi olarak kabul edilir.

6.4.2. Karar ağacı algoritması ile karar verme

Karar ağaçları eğitici öğrenme için en yaygın olarak kullanılan bir yöntemdir. Bu yöntemde, genellikle bir eğitim seti üzerinden karar tabanlı sınıflandırma ve adaptif öğrenme gerekmektedir. Karar ağacı, bu tür taktiklerin uygulanması için popüler bir yardımcı yöntemdir. Karar ağacı, karar modelleme aracı verilen çıkış için belirli bir giriş sınıf etiketinin sınıflandırma sürecini grafiksel olarak gösterir. Bundan dolayı karar ağaçları, ilk bakışta akış şemalarına benzemektedirler. Bu yapının elemanları, adından da tahmin edilebileceği gibi dal ve yapraktır. Her bir elemanın özelliği bir düğüm ile gösterilir. En son yapı “yaprak”, ilki “kök” olarak ve bu ikisi arasında kalanlar da “dal” olarak nitelendirilir. Karar ağaçları sınıflamanın yapılabilmesi için uygun bir alt yapı sağlamaktadır. Karar ağacı ile verilecek bir kararın algoritması incelenecek olursa, ilk olarak öğrenme kümesi oluşturularak başlanır. Algoritmaya örnek vererek devam etmek adına bu öğrenme kümesine “Q” adını verelim. Bu oluşturulan “Q” kümesini oluşturan örnekleri birbirinden en iyi şekilde ayıran nitelik belirlenir. Sonraki adımda seçilen bu nitelik ile ağacın bir düğümü oluşturulur ve düğümden ağacın çocuk düğümleri veya ağacın yaprakları oluşturulur ve çocuk düğümlere ait alt veri kümesinin örneklerini belirlenir. Son adımda da yaratılan her alt veri kümesi için örneklerin hepsi aynı sınıfa aitse, örnekleri böylecek nitelik kalmamış ise veya kalan niteliklerin değerini taşıyan örnek yok ise işlem sonlandırılır. Bahsedilen bu durumlar dışında oluşan durum için “Q” kümesini oluşturan örnekleri birbirinden en iyi şekilde ayıran niteliğin belirlendiği adımdan devam edilir.

7. DİKEY ARAMA MOTORU İLE İLGİLİ MEVCUT ÇALIŞMALAR

Arama motorları ve dikey arama motorları arasındaki performans, sayfa – konu ilişkisi, web motorunu daha etkin hale getirebilmek için yeni algoritma geliştirme, bu arama motorları kullanarak reklam vermenin maliyetleri konularında çeşitli karşılaştırmalar, ölçümler ve çalışmalar yapılmıştır. Bu çalışmalardaki genel amaç, mevcut olan odaklı ağ örümceklerinin daha etkin çalışmaları için yeni yaklaşımlar geliştirilmiş, dikey arama motorları ile reklam yapmanın arama motorları ile reklam yapmanın maliyetleri yönünden karşılaştırma yapılmıştır.

Yukarıda belirtilen çalışmalardan bir tanesi, dikey arama motoru ile reklam verme alanında yapılmış bir çalışmadır[10]. Yapılan çalışmalar pazarlama şirketlerinin arama motoru ile ve dikey arama motoru ile reklam yaptıklarındaki maliyetleri karşılaştırılmaktadır. Ayrıca bu çalışmada maliyetler ile birlikte dikey arama motoru ile bir dişçinin kullanmakta olduğu seramik maddesini Google arama motoru kullanılarak ve bir dikey arama motoru kullanarak arama yaptığındaki görüntülediği sonuçlara da değinilmiştir. Araştırma ve çalışmaların sonucunda, şirketlerin 7,4 milyar dolar arama motorundan pazarlama yapmaya harcadıklarına ulaşılmıştır. Durum böyle olunca çevrimiçi reklamcılığın artışı kaçınılmaz olmaktadır. 2006 yılında çevrimiçi reklamcılığa 1.3 milyar dolar harcandığı yapılan araştırmalarda ortaya çıkmıştır[10]. Bu harcamaların da büyük bir kısmının Google AdWords e harcandığı belirlenmiştir. Yapılan karşılaştırmalarda en bilinen dikey arama motoru olan LookSmart ile yapılan çevrimiçi reklamın Google AdWords ile yapılandan çok daha az maliyete sahip olduğu ortaya konmuştur[10]. Böylelikle de pazarlama şirketleri yapmış oldukları yatırımlarında daha fazla kazanç elde etmektedirler[10].

Arama motorlarının genel sorunu kullanıcı ara yüzü ile yapılan aramanın sonuçları arasında aranan ile ilgisiz çok fazla sonuç görüntülemesiydi. Buna ek olarak, arama motorları anahtar kelimeye göre arama sonuçlarını görüntülemektedirler. Bu büyük ölçekli arama motorları için en iyi seçenek olmasına rağmen daha anlamlı sorgu

tanımlayabilme kapasiteleri kısıtlıdır. İnternette bulunan verilerin çok büyük bir kısmının da HTML formatında olması, araştırmacıları XML sorgu motoru tasarlamaya itmiştir[12]. Yapılan çalışmada XML sorgu motoru tasarlanmıştır. XML sorgu motoru, ağ örümceklerinin internetten topladığı web sayfalarını XML formatına dönüştürüp, indeksleme yapan bir mimariye sahiptir[12]. Böylelikle bu çalışmada, maliyeti yüksek olan ama ilişki derecesi arama motorlarına göre daha yüksek sonuçlar görüntüleyen kavram tabanlı ve bağlamsal olarak farkında olan bir sorgulama motoru geliştirilmiştir[12].

Diğer bir çalışma ise bağlam odaklı ağ örümceği olup, bu çalışmada aramayı mükemmelere erişirmek planlanmıştır[13]. Bu çalışmada internetten hedef dokümana nasıl ulaşıldığını tasvir eden bağlam grafiği ve sınıflandırıcılar kullanılmıştır. Bağlam odaklı ağ örümceklerinin kullandığı algoritmanın ilk aşamada, her bir kök doküman için bağlam grafiği kümesi ve ortak sınıflandırıcılar oluşturulur[13]. İkinci aşama olan gezinme aşaması, sınıflandırıcıları kullanarak aramaya yol gösterir ve bağlam grafiğini çevrimiçi güncelleme işlemi yapar[13].

Bir başka çalışmada adı PISE (Personal Intelligent Search Engine) olan sistem tasarlanmıştır. Bu sistemde kullanıcının arama uzayında gezinmek için, kullanıcının ilgisi ve anahtar kelimesi ile gerçek zamanlı arama algoritması geliştirilmiştir[14]. Bu sistemde kullanıcının geçmişi yapılarak farklı kullanıcıların aynı kelimeyi girseler de PISE farklı sonuçlar getirmesi sağlandı[14]. Aynı kullanıcı aynı kelimeyi farklı zamanda girse bile kullanıcının daha önce sorguladığı kelimelerin tutulduğu iki boyutlu web sayfası domain adı verilen tablodan yararlanarak ilgisine en yakın web sayfalarını az zaman harcayarak görüntülenmektedir[14].

Başka bir çalışmada, arama sonuçlarını mükemmelere yaklaştırmak ve içeriği özetlemek amaçlı tasarlanmış bir sistemdir[15]. Bu çalışmada tasarlanmış sistemin kullandığı alt fonksiyonlar bağlantı madenciliği (linkminer), iyileştirici (optimizer) ve özetleyicidir (summarizer) [15]. Bağlantı madenciliğinde bağlantıların yapısı incelenir[15]. Bu aşamada; HTML dosyası ayrıştırılır, gereken bağlantılar çıkartılır, bağlantının yapısı incelenir, bağlantı optimizasyonu ile fazlalık bağlantılar atılır, istenmeyen bağlantılardan kurtulduktan URL ayrıştırılır ve son olarak da doküman veri

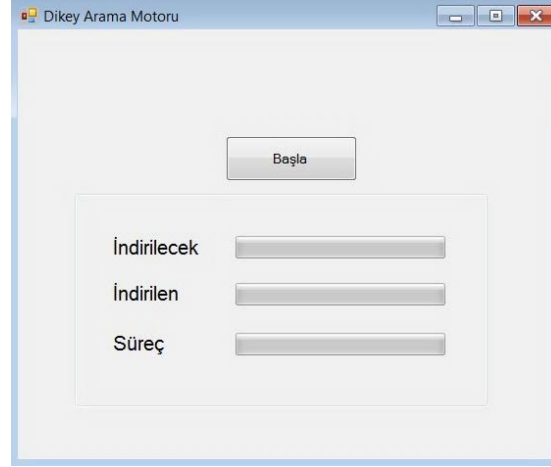
tabanına içerikler indirilir[15]. Bağlantı madenciliğinden sonra doküman veri tabanına aktarılan bağlantılar kullanışlı oldukları kesindir[15]. Bu aşamada dokümanlar arasındaki benzerliği belirlemek için içerdikleri kelimeleri karşılaştırılır[15]. Son aşamada ise içerikler optimize edildiği zaman kümeler şeklinde özetlenebilir hale gelmiş olur[15]. Böylelikle son kullanıcı da mükemmel yakın bir özet almış olur. Bu çalışmada derinlemesine özet geliştirme konusuna girilmemiştir[15]. Böylece bu çalışmada geliştirilen yeni sistem ile arama sonuçlarının etkinliğinin arttığı ortaya konmuştur. Ayrıca; etkin şekilde içerikleri özetlemek için bit metot sunulmuş, etkin sonuç elde etmek için sorguyu düzenlemeye yoğunlaşmayıp dokümanları kümeleme yoluna gidilmiştir. Bu kümeleme işlemi de sonuçların organize edilmiş küme içerisinde kategorize edilmesini sağlamıştır.

8. GELİŞTİRİLEN DİKEY ARAMA MOTORU

Yapılan çalışmada bölüm 6.3.1’ de açıklanmış olan eğitim setini kullanarak indirilen sayfanın konusuna en yakın k komşu algoritması ve karar ağacı sınıflama algoritmaları kullanarak karar veren dikey arama motoru geliştirilmiştir. Arama motorları ve dikey arama motorlarının, ağ örümceği ve kullanıcı arayüzü bileşenlerine sahip olduğu daha önceki bölümlerde bahsedilmiştir. Dikey arama motorunun geliştirilmesinde, C#.NET programlama dili kullanılmıştır. Geliştirilen Ağ Örümceğinin indirdiği sayfalarındaki, web sayfası adreslerinin indekslenip kaydedilmesi ve kullanıcı arayüzü ile bu kayıtlar arasından görüntüleme yapılması için Microsoft SQL Server veri tabanı kullanılmıştır. Görüntüleme ve kaydetme işlemi için, uygulamanın gereken yerlerinde, SQL sorgu dili kullanılmıştır. Aşağıdaki bölümden başlayarak geliştirilen uygulamadan detaylıca bahsedilmektedir.

8.1. Geliştirilen Ağ Robotu

Ağ robotu; geliştirilen uygulama için, Yüksek Öğretim Kurumu’ nun web sayfasından, Türkiye’ de bulunan devlet üniversitelerinin, vakıf üniversitelerinin ve diğer yüksek öğretim kurumlarının web adreslerini girdi olarak alarak ilk işlemlerine başlamaktadır. Bu işleme başlamadan önce, geliştirilen görsel bir arayüz ile Ağ örümceğine başlama komutu verilir. Ağ örümceğinin görsel arayüzü Şekil 8.1’ de gösterilmektedir. Gösterilen forma ait kaynak kodlar Ek – A’ da verilmektedir.



Şekil 8.1. Ağ Robotu Arayüzü

Uygulamanın ağ robotu bileşeninde, başlangıç olarak bahsedilen Türkiye’ de bulunan eğitim kurumları Excel dokümanında girdi olarak verilmiştir. Bu dokümandan satır satır okuma ve ağ robotunun yapacağı HTML sayfasını indirip işleyeceği okulun web sayfası adresini, ilgili C# sınıfına gönderme işlemi “Başla” butonuna basılması ile tetiklenmiş olur. Uygulama çalışmaya başladıktan sonra kullanılan iki ayrı sınıf bulunmaktadır. Bu sınıflardan biri “HTMLparser.cs” sınıfı, diğeri de “spider.cs” sınıfıdır. “spider.cs” sınıfından yaratılan nesne ile, Excel dokümanından okunmuş bir satır olan okul web sayfası adresinin HTML kodlarını indirmek için ilgili fonksiyonu çağırabilme olanağı sağlanmış olur. Web adresinin HTML kodlarını indirmek, yapılacak olan işlemlerin ilk aşaması ve temelidir. Bu sınıf ile indirilen web sayfasının metin kısmında Türkçe karakterler ile ilgili sorunlar çözülür. Bu sorunlar çözüldükten sonra indirilen web sayfasının içeriği, gizli durumda olan “rtxtHTML” isimli bir metin kutusu alanının değerine atanır. Bu atama işlemi, indirilen sayfanın içeriğini kaybetmeden diğer işlemler için kullanılabilmesi içindir. Bu aşamaya kadar gelirken, diğer yandan da indirilen web sayfasının diğer sayfalara olan bağlantıları bulunur ve bu bağlantılar, veri tabanına ilgili alanlara kaydedilir. Eğer bulunan bağlantı veya bağlantılarda eksiklik söz konusu ise, o bağlantı veya bağlantılar olması gerektiği haline dönüştürülür. Böylece bulunan eksik, çalışmayacak olan bir bağlantı düzeltilmiş olur. Bulunan bu bağlantılar,

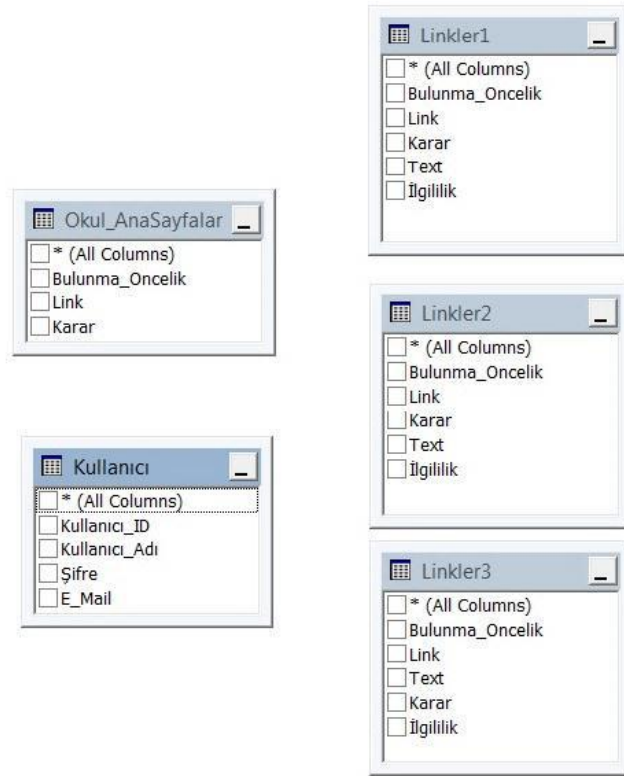
Excel dokümanında bulunan eğitim kurumlarının web adreslerinin incelenmesi bittikten sonra, işleme alınacak oldukları için veri tabanındaki ilgili tabloya kaydedilmektedirler. Eğitim kurumlarının web sayfa adresleri yardımı ile indirilen web sayfasının içeriği, arama motorunun diğer bileşeni olan HTML ayırıcı tarafından temizlendiği bilinmektedir. Gizli olan metin kutusu alanında bulunan, indirilen web sayfasının içeriğine özel kodları, “HTMLparser.cs” sınıfından bir nesne oluşturularak ve bu sınıfa ait fonksiyonlar çağırılarak temizlenmektedir. Bu sınıfa ait fonksiyonda, temizleme işlemi, düzenli ifadeler (Regex.replace) fonksiyonu yardımı yapılmıştır. Bu temizlemeden kasıt, indirilen web sayfasının içeriğine özel ve geneli HTML etiketleri olan ifadelerden, java script’lerden ayırıp, sade metin haline, yani işlenebilir hale getirmektir. Web sayfası içeriklerinin metin kısmı ile çalışılmasının sebebi, metin içerisinde geçen kelime sıklıklarına göre hangi sınıfa dâhil edileceği kararının verilecek olmasıdır. Bu kararın verilmesi aşaması, bu bölümün ilerleyen kısmında daha detaylı şekilde açıklanacaktır. Web sayfası içeriği etiketlerinden ayrıştırılan sayfanın metin değeri, yine görsel arayüzde bulunan “MetinTxt” isimli metin kutusu alanına atanır ve “MetinTxt” metin kutusu alanının mevcut değeri sonraki adımlarda kullanılacak fonksiyonlara parametre olarak gönderilir. Ek – C’de HTML kodlarını etiketlerden ayıran “HTMLparser.cs” sınıfına ait kodlar bulunmaktadır.

Şu ana kadar içeriği indirilen web sayfalarının etiketlerinden ayrılmış metne çevirme işlemi ve bu işleme kadar aynı anda yapılan diğer işlemler anlatılmıştır. Bu aşamadan sonra ağ robotu; metni sınıflandırıp, sınıf adı ile indirilen web sayfasının adresini veri tabanına yazma işlemlerinden bahsedilecektir. Ağ robotunun sınıflandırma bölümü için WEKA’ nın dâhilinde bulunan kütüphanelerden yararlanılmıştır. WEKA, açık kaynaklı bir uygulama olup, kodlar dinamik bağlantı kütüphanesi haline geliştirilmiş ve uygulama içinde kullanılmıştır. Daha önce de belirtildiği üzere WEKA yapay öğrenme uygulaması olup sınıflamayı; tecrübe edilmiş, öğrenilmiş örnek veriler üzerinden yapmaktadır. Tez uygulaması için belirlenmiş seminer, sempozyum, konferans, kongre, panel, çalıştay sınıflarına ait örnek web sayfaları ile bir eğitim dokümanı oluşturulmuştur. Bu eğitim dokümanı ve eğitim dokümanı hazırlanırken kullanılmış web sayfaları Ek – D ve Ek – E’de verilmektedir. Geliştirilen uygulama Ek – D’de gösterilen eğitim seti dokümanı yardımı ile indirilen web sayfalarını

sınıflandırmaktadır. Bu sınıflandırmayı yaparken önce de bahsedilmiş en yakın k komşu ve karar ağacı algoritmalarını kullanmaktadır. Yapılan sınıflama işleminde iki ayrı algoritma kullanıldığından, yapılacak ölçümler için bu algoritmalara ait iki ayrı sonuç dokümanı oluşturulmuştur. Bu dokümanların oluşturulmasında “Kumeleme” adında Visual Basic.NET programlama dili kullanılarak geliştirilmiş diğer bir yardımcı uygulamanın içerisinde bulunan kodlar ile oluşturulmuştur. Bu uygulamanın kaynak kodları EK – K’ de gösterilmektedir. Geliştirilen uygulamada, karar ağacı algoritması ile belirlemiş olduğu sınıfı, incelenen web sayfasının adresini ve bu web sayfasının etiketlerinden ayrılmış metni ile veri tabanındaki ilgili tabloya kaydedilmektedir. Web sayfalarının etiketlerinden ayrılmış metnin veri tabanına kaydedilmesinin sebebi, arama yapacak kullanıcılara aradıkları sitenin içeriğinde de arama yapabilmelerini sağlamaktır. Geliştirilen uygulamada karar verme aşamasında kullanılan en yakın k komşu ve karar ağacı algoritmaları ile ilgili kodlar Ek – F ve Ek – G’ de görülmektedir. Bu algoritmalar belirtilen veri yolunda bulunan “Egitim.arrf” dokümanındaki eğitim satırları yardımı ile sınıflandırma yapılmıştır. Verilen eklerden de görüldüğü gibi bu iki fonksiyon gerçek sayı tipinde değer döndürmektedirler. Bu değerlerin karşılıkları eğitim dokümanında belirlenmiş sınıfların ismidir. 0 indisli öznitelik dizisinin karşılığı “seminer”, 1 indisli öznitelik dizisinin değeri “sempozyum”, 2 indisli dizi elemanının değeri “konferans”, 3 indisli dizi elemanın değeri “kongre”, 4 indisli dizi elemanın değeri “panel”, 5 indisli dizi elemanın değeri “çalıştay” ve 6 indisli dizi elemanının değeri de “diğer” dir. “Diğer” sınıfının bulunmasındaki amaç, indirilmiş olan sayfanın içeriğinin, belirtilen ve arama sonuçlarının görüntülenmesinde büyük rol oynayan sınıfların hiçbirine ait olmadığını belirtmektir.

Daha önce de belirtildiği gibi indeksleme işleminden önce, indirilen web sayfasının içeriği etiketlerinden ayrıştırıldıktan ve sayfanın sınıfı belirlendikten sonra veri tabanına kaydedilmelidir. Sayfa hakkındaki bilgilerin veri tabanına kaydedilmesi için geliştirilen uygulamaya ait veri tabanında; bulunan bağlantılar, bu bağlantıların temizlenmiş metin bilgisi, bağlantıların sınıfı, dikey arama motorunu kullanıp bağlantıların ilgililikleri hakkında geri bildirimde bulunan kullanıcıların yapmış olduğu ilgililik belirlemesinin değeri, bağlantıların bulunma önceliği, kullanıcı adları ve şifreleri tablolarda tutulmaktadır. Bulunan bağlantılar, bu bağlantıların metin kısmı, bağlantıların

sınıfı, dikey arama motorunu kullanıp bağlantıların ilgililikleri hakkında geri bildirimde bulunan kullanıcıların yapmış olduğu ilgililik belirlemesinin değeri, bağlantıların bulunma önceliği değerleri linkler tablolarında tutulmaktadır. Bu tablolar bahsedilen bağlantıların bulunma önceliklerine göre adlandırılmıştır. Kullanıcıların belirlemiş oldukları kullanıcı adı ve şifre değerleri de kullanıcı tablosunda tutulmaktadır. Şu ana kadar bahsi geçmemiş olan “Okul_Anasayfalar” tablosu, dikey arama motorunun giriş değerleri olan eğitim kurumlarının ana sayfalarının adreslerinin tutulduğu bir tablo olup, kullanıcıların yapmış oldukları aramalarda bu tablodan herhangi bir bağlantı görüntülenmemektedir. Dikey arama motorunun veri tabanı katında kullanılan tablolar ve alanların adları Şekil 8.2’ de gösterilmektedir.



Şekil 8.2. Veri Tabanı Tabloları

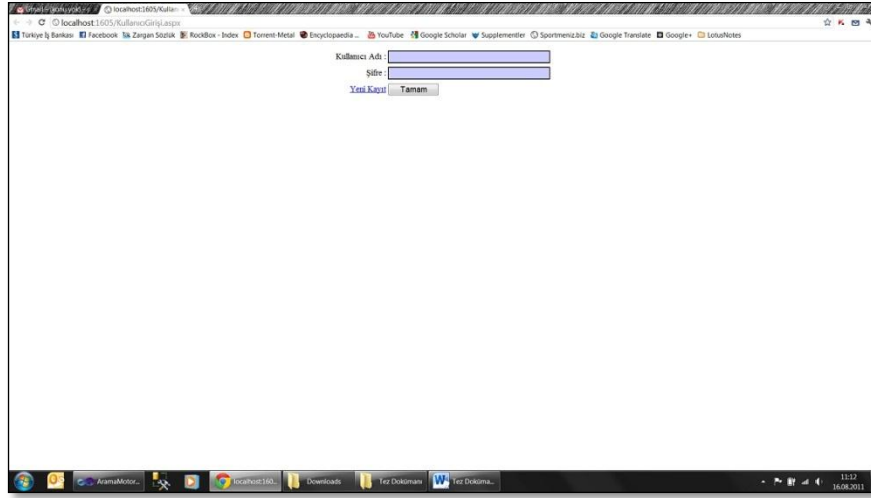
Şekil 8.2’ de gösterilen veri tabanı tablolarında “Bulunma_Oncelik” kolonu, sayısal veri tipinde olup bulunan ve kaydedilen linklerin okul ana sayfasından kaçınıcı defada bulunduđu bilgisini içermektedir. “Link” alanı, metin veri tipinde olup bulunan bağlantının web adresi bilgisini tutmaktadır. “Text” alanında indirilen web sayfasının, web sayfası içeriğinden ayrıştırılmış hali kaydedilmekte olup, bu kolonun değeri metin veri tipindedir. “Karar” alanında, yapılan sınıflandırmanın değeri tutulmakta olup bu değer de metin veri tipindedir. Son olarak “İlgililik” alanı, kullanıcıların yapmış olduđu aramaların görüntülenen sonuçlarla ilgili olup olmadığı bilgisinin saklandığı alandır. Bu alan sayısal veri tipindedir ve görüntülenen bağlantı ilgili ise 1, ilgisiz ise 0 olarak veri tabanında tutulur.

8.2. Kullanıcı Arayüzü

Genel olarak bir arama motoru iki yüzden oluşur denebilir. İlk ve kullanıcının görmediği ve genellikle de hakkında çok bilgisi olmadığı; kullanıcıların arama yapacakları bilgilerin, internette bulunan web sayfalarının ziyaret edilerek ve onların indekslenip veri tabanına kaydedildiği kısım olan ağ örümceği kısmıdır. İkinci kısım ise kullanıcıların indekslenen web sayfaları arasından arama yapabilmelerini sağlayan kullanıcı arayüzü kısmıdır. Bu bölümde geliştirilen arama motorunun kullanıcı arayüz kısmından söz edilecektir. Kullanıcı arayüzü kısmı, ASP.NET uygulaması olup sunucu tarafında C#.NET programlama dili kullanılarak geliştirilmiş internet üzerinde çalışan bir arayüzdür. Bu arayüzün geliştirilme amacı, ağ örümceği tarafından depolanmış bilgilerin, arama yapacak kullanıcılar tarafından aramalarını gerçekleştirebilmeleridir.

Kullanıcı arayüzü, dikey arama motoru ile arama yapacak kullanıcıları, kullanıcı girişi sayfası ile karşılamaktadır. Giriş sayfası Şekil 8.3’ te görüntülenmektedir. Bu sayfanın bulunmasındaki amaç, dikey arama motorunu ile arama yapacak kullanıcıların bilinçli birer kullanıcı olduklarından emin olup, ileriki

aşamalarda bize doğru geri bildirimlerde bulunmaları için fırsat tanımaktır. Böylece dikey arama motorunu kullanacak kişiler sistemde kullanıcı adı ve şifresi tanımlı olan kişiler olmak durumundadırlar. Eğer bir kullanıcı; dikey arama motorunu ilk defa kullanacak ise; giriş sayfasında bulunan “Yeni Kullanıcı” butonuna basarak kayıt oluşturması için kullanıcı kayıt sayfasına yönlendirilmektedir. Şekil 8.4’ te bu sayfa görüntülenmektedir. Yeni kullanıcı sayfasına ilerleyen bölümlerde daha detaylı değinilecektir. Kullanıcı giriş sayfasında, istenilmeyen karakter kontrolü yapılmıştır. İstenilmeyen karakter kontrolünün amacı, kullanıcının bu karakterleri kullanarak sisteme izinsiz şekilde giriş yapmasını önlemektir. Bu istenilmeyen karakter kontrolü java script kullanılarak yapılmış olup ve kaynak kodları Ek – I’ da verilmiştir.



Şekil 8.3. Kullanıcı Giriş Sayfası

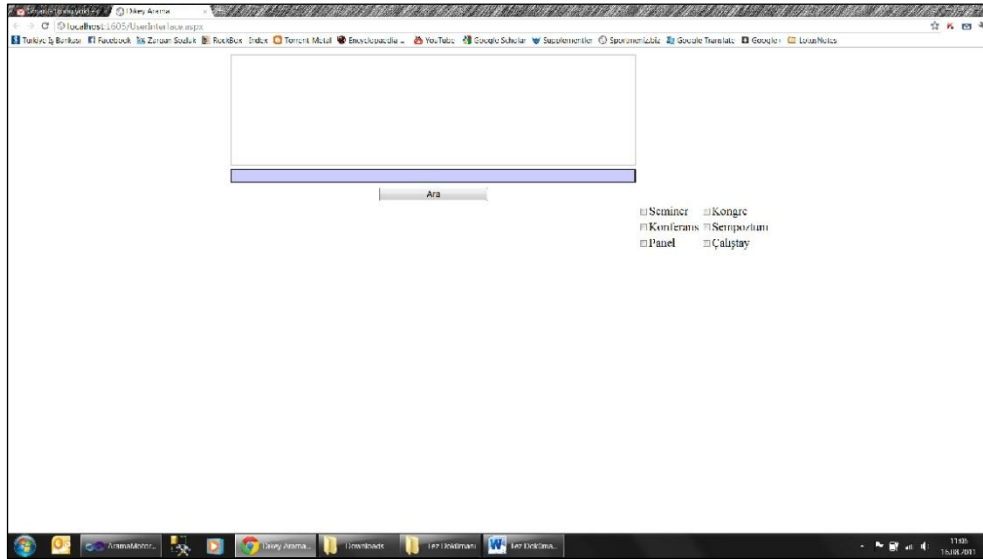
Görüntülenen arama sonuçlarının, aranan konu ile ilgili olup olmadıkları bilgisinin, dikey arama motorunu kullanan kişi tarafından geri bildirim ile sağlanacağı bilinmektedir. Bu aşamada; kullanıcılardan, üyelik sistemine giriş yapmış olmalarını, dolayısı ile sistemde kayıtlarının bulunması beklenmektedir. Eğer kullanıcının sistemde

kaydı yok ise sisteme geri bildirimde bulunamaz. Kullanıcı kayıt sayfası, Şekil 8.4' te gösterilmektedir. Kullanıcı kayıt sayfasında; kullanıcının, kullanıcı adını, şifresini belirlediği ve elektronik posta adresini girdiği alanlar bulunmaktadır. Bu alanlarda, güvenlik amaçlı istenilmeyen karakter kontrolü, elektronik posta adresinin doğru formatta girilip girilmediği kontrolü ve ilk girdiği şifrenin ikincisi ile aynı olup olmadığı kontrolleri java script ile yapılmıştır. Java script kaynak kodu Ek – J' de gösterilmektedir. Kullanıcı bu kontrolleri geçtikten sonra belirlemiş olduğu kullanıcı adı daha önce başka bir kullanıcı tarafından alınmamış ise belirlemiş olduğu bilgiler veri tabanına kaydedilmektedir. Eğer belirlemiş olduğu kullanıcı adı sistemde mevcut ise yeni bir kullanıcı adı belirlemesi için uyarılmaktadır.

Şekil 8.4. Kullanıcı Kayıt Sayfası

Belirli özel bir konu veya konularda arama yapmak, dikey arama motorlarının ana amacı olduğundan dolayı geliştirilmiş dikey arama motoru da seminer, kongre, konferans, sempozyum, panel, çalıştay ile ilgili haberleri ve bu belirtilen haberlerin konuları hakkında arama yapabileceğini kullanıcılara sunmaktadır. Dikey arama motorunun kullanıcı arayüzü Şekil 8.5' te gösterilmektedir. Kullanıcıların arama yapmak istedikleri haberlerin etkinlik sınıfını veya sınıflarını Şekil 8.5' te görülen kontrol kutuları alanlarının yardımı ile yapabilmektedirler. Buna ek olarak, seçtikleri

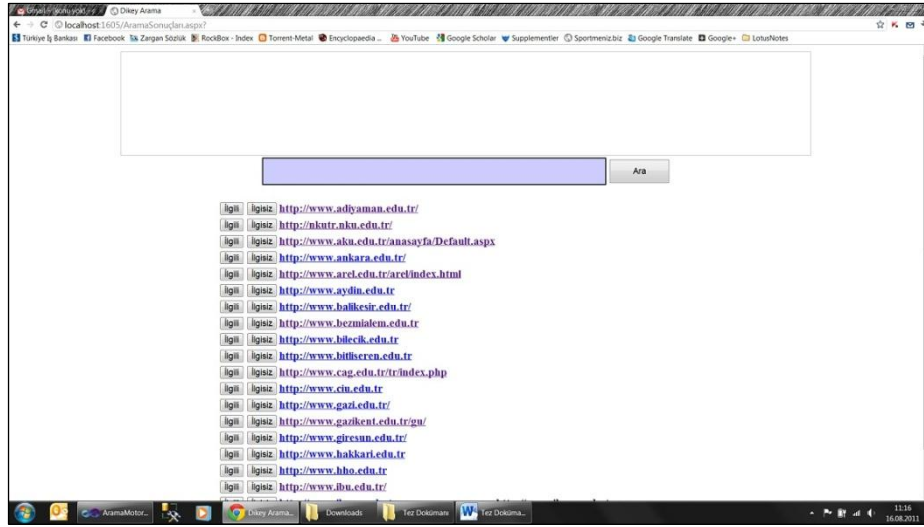
haber sınıflarının içerisinde istedikleri bir konu ile ilgili olan haberleri görüntülemeleri için de metin kutusu alanı bulunmaktadır. Arama sonuçlarını eğer arama kriterleri ile ilgili bilgi veri tabanında mevcut ise, “Ara” butonuna bastıktan sonra görüntüleyebilmektedirler. Kullanıcının görüntülenmesini talep edecek herhangi bir etkinlik seçilip seçilmemiş olduğu ve arama metin kutusu alanına bir giriş yapıp yapılmamış olduğu kontrolü, kullanıcı tarafında yazılmış olan java script ile yapılmıştır. Bu kontrolün yapıldığı java script kaynak kodları Ek – H’ de görüntülenmektedir.



Şekil 8.5. Kullanıcı Arayüzü

Kullanıcı, bahsedilen kontroller bakımından gerekeni yaptıktan sonra arama sonuçları sayfasına yönlendirilir. Bu sayfada, kullanıcının yapmış olduğu arama sonuçlarının veri tabanında karşılıklarını görüntüler. Veri tabanına sayfaların linkleri, etkinlik sınıfı ve web sayfası içeriğini metin olarak kaydedildiğinden, kullanıcıya bu bilgilerin sadece bağlantı bilgisi görüntülenir. Bu sayfada anlatılanlara ek olarak kullanıcıdan, görüntülenen bağlantıların aradığı konu ve etkinlik sınıfı ile ilgili olup

olmadığı hakkında geri bildirimde bulunması istenmektedir. Bu geri bildirim, ağ örümceğinin sınıflandırma aşamasındaki etkinliğini öğrenmede yol göstermesi bakımından büyük etki sağlayacağı düşünülmektedir. Arama sonuçlarının görüntülediği sayfa, Şekil 8.6' da gösterilmektedir.



Şekil 8.6. Arama Sonuçları Sayfası

9. SONUÇLAR VE ÖNERİLER

Geliştirilen dikey arama motorunun başarımının, WEKA yazılım aracı ile Ek- D' de verilen eğitim seti üzerinden değerlendirilmesi yapıldığında;

Ek- D' de verilen eğitim setini oluşturan örnekler,

- Seminer
- Sempozyum
- Konferans
- Kongre
- Panel
- Çalıştay
- Diğer

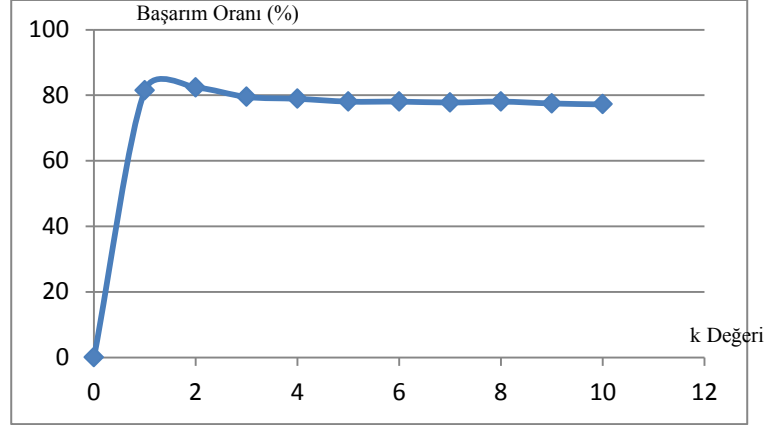
Olmak üzere 7 farklı sınıfta, her birinden ellışer adet olacak şekilde seçilmiştir. Eğitim setini oluşturan bu örneklerin eşit miktarda seçilmesinin sebebi, bir doküman türünün diğer türe baskın gelip sınıflandırmayı kötüleştirmesini engellemektir.

Geliştirilen dikey arama motorunda web sayfaları içeriklerini sınıflandırmak için kullandığımız en yakın k komşu ve karar ağacı algoritmalarından önceki bölümde bahsedilmiştir. Bu iki algoritma kullanılarak varsayılan parametrelerle eğitim seti üzerinden alınan başarım oranı sonuçları çizelge 9.1' de verilmiştir.

Çizelge 9.1. Sınıflama algoritmaları başarım oranı sonuçları

Sınıflama Algoritması	Başarım oranı(%)
J48 (Karar Ağacı algoritması)	86,33%
IBk (en yakın komşu algoritması)	81,48%

Çizelge 9.1' de görüldüğü gibi, J48 Karar ağacı algoritması ile yapılan sınıflandırma yaklaşık %5 oranında farkla en yakın k komşu algoritmasından daha başarılıdır. k en yakın komşu algoritmasında, k değeri 1 olarak alınmış, yani sadece en yakın komşuluğa bakılmıştır. k değeri yükseltilerek yapılan deneylerde en yüksek başarım oranı k değerinin 2 olduğu durumda alınmış olsa da, bu değer %82,33 oranında kalarak yine karar ağacı algoritmasından daha düşük başarıma sahip olmuştur. k değerinin 10' a kadar yükseltilmesi ile yapılan deneylere ilişkin başarım grafiği Şekil 9.1' de görülmektedir. Grafikten de anlaşılacağı üzere, k için optimum değer 2 olduğu görülmektedir.



Şekil 9.1. En yakın k komşu algoritmasında k değerine göre olan değişim

J48 karar ağacı sınıflandırma algoritması için de çeşitli parametre değişiklikleri denenmesine karşın, en iyi sonucun varsayılan parametrik değerler ile elde edildiği görülmüştür.

Sınıflandırma algoritmaları kullanılarak oluşturulan sınıflandırıcı araçların karşılaşılabileceği çeşitli sorunlar vardır. Bunlardan biri, eldeki verinin herhangi bir sınıf tipinde fazla veri içermesinden kaynaklı olarak oluşan aşırı uyum (overfitting) sorunudur. Bu sorunu en başından engellemeye çalıştığımız için sınıflar arasında dengeli bir dağılım olmasına dikkat edilmiştir. Ancak yine de her sınıflandırıcının başarımı nesnel bir ölçütle aşırı uyuma karşı denetlenmelidir. Bu amaçla kullandığımız ölçüt de kapa istatistiği[16] olmuştur. Her iki algoritmada da en yüksek başarımın elde edildiği parametrik sınıflandırıcılar için kapa değerleri çizelge 9.2’ de verilmiştir.

Çizelge 9.2. Sınıflama algoritmaları kapa değeri

Sınıflama Algoritması	Kapa değeri
J48 (Karar Ağacı algoritması)	0,84
IBk (en yakın komşu algoritması)	0,79

Çizelge 9.2’ deki kapa değerlerine bakıldığında veri kullanılarak yapılan öğrenme işleminin oldukça başarılı olduğu görülmektedir. Kapa değeri üzerine yapılan çalışmalar 0,4 üzerindeki her kapa değerinin[17] veride öğrenilebilir bir örüntü

olduğunu gösterdiğini vurgulamaktadır. 0,8 civarındaki her iki kappa değeri de bu açıdan oldukça başarılıdır.

Doküman sınıflandırmamızda önceden 7 tür belirlediği belirtilmişti. Bu türlerden hangilerinde en yüksek başarımlar elde edildiğine bakarak, ileriki çalışmalarda diğer türlere ilişkin başarımların yükseltilmesine yönelik çalışmalar da yapılabilir. Çizelge 9.3’ de J48 ve IBk sınıflandırıcıları ile elde edilen karmaşıklık matrisleri verilmiştir.

Çizelge 9.3. Sınıflama algoritmaları karmaşıklık matrisleri

a	b	c	d	e	f	g	<-	TAHMİN (J48)
47	0	0	0	1	1	1	a	Seminer
0	48	0	1	1	0	0	b	Sempozyum
0	1	43	2	3	2	0	c	Konferans
0	2	4	42	2	0	0	d	Kongre
2	2	1	2	43	0	0	e	Panel
1	2	4	0	1	39	3	f	Çalıştay
0	1	5	1	0	2	41	g	Diğer

a	b	c	d	e	f	g	<-	TAHMİN (IBk)
46	0	1	0	1	1	1	a	Seminer
2	40	2	1	0	1	4	b	Sempozyum
1	1	42	1	1	1	4	c	Konferans
0	5	2	39	2	0	2	d	Kongre
0	1	2	0	45	0	2	e	Panel
2	1	3	1	1	38	4	f	Çalıştay
3	0	6	0	1	1	39	g	Diğer

Karmaşıklık matrisleri her sınıf için ayrı ayrı incelendiğinde, J48 sınıflandırıcısının “panel” türü hariç diğer türlerde IBk sınıflandırıcısından daha iyi başarımlar gösterdiği görülmektedir. Sadece bu tür açısından bakıldığında da her iki sınıflandırıcı arasındaki başarımların farkı göz ardı edilebilir seviyededir. Yine her iki sınıflandırıcı için de en düşük ortak başarımların çalıştay türünde görülmektedir. Bu tür için özel bir çalışma yapılması gerektiği açıktır.

Çizelge 9.4. Özelliklerin kazanç değerleri

Kazanç değeri	Özellik
0.49727	sempozyum
0.46283	seminer
0.43413	panel
0.43171	kongre
0.36131	çalıştay
0.34637	konferans
0.14784	tarih
0.1262	katılım
0.10995	bilimsel
0.08517	bildiri
0.03962	oturum
0.02156	sunum
0.01874	etkinlik
0.01219	toplanti
0.01047	diğer
0.00439	ulusal
0.00272	uluslar arası

Veri üzerinde incelenmesi gereken bir diđer kısım da, veriyi oluřtururken kullanılan özellik kümesindeki her bir elemanın sınıflandırmaya olan katkısıdır. Bunu gerekleřtirmek için bilgi kazancı (information gain) olarak isimlendirilen bir ölçütten faydalanılır. Kullandığımız özelliklere ait bilgi kazancı deđerleri izelge 9.4' te verilmiřtir.

Kullanılan özelliklerin bilgi kazançlarına bakıldığında da türlere karřılık gelen yani sempozyum, seminer, panel gibi özelliklerin yüksek bilgi sağladıkları görölmektedir. Bu gibi özelliklerin yüksek bilgi sağlaması zaten beklenen bir durumdur. Ancak, bu özelliklere ek olarak kullanılan diđer özelliklerin de bilgi kazancı sağladığının görölməsi önemli bir durumdur. Çünkü bilgi kazancı deđerleri 0 olan bir özellik yoktur ve kullanılan her özelliğin sınıflandırma başarısına olumlu yönde katkı sağladığı söylenebilir. Kullandığımız eğitim verisine ilişkin bu bilgiler ışığında, oldukça yüksek başarımlı bir doküman türü ayırıcı geliřtirdiğimiz açıka görölmektedir.

Yukarıda yapılan deđerlendirmeler dođrultusunda bakıldığında dikey arama motoru geliřtirilmesinde makine öğrenmesi kullanmak başarılı sonuçlara ulařtırmaktadır. Bu alıřmada seçilmiř özellikler yerine bir başka konu için özellikler belirlenip bu özelliklere bađlı eğitim dokümanı oluřturularak farklı konular için dikey arama motoru geliřtirilebilir.

Kaynaklar

- [1] http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_repo rt.pdf
- [2] http://hmi.ucsd.edu/howmuchinfo_research_report_consum.php
- [3] <http://www.cs.waikato.ac.nz/ml/weka/>
- [4] Y. Tonta, Y. Bitirim, H. Sever, 2002."Türkçe Arama Motorlarında Performans Değerlendirme". Total Bilişim.
- [5] <http://www.w3.org/>
- [6] <http://searchenginewatch.com/article/2065173/How-Search-Engines-Work>
- [7] http://www.uniroma2.it/didattica/prog_web/deposito/search_engine.pdf
- [8] M. Peshave, 2002. "How Search Engines Work And A Web Crawler Application". University of Illinois at Springfield Department of Computer Science.
- [9] S. Batsakis, E. G.M. Petrakis, E. Milios, 2009. "Improving The Performance of Focused Web Crawlers", Technical University of Crete Department of Electronic and Computer Engineering
- [10] Jude Mc Glinchey, 1999. "Vertical Search Engines". Intelligent Systems Research Centre University of Ulster Faculty of Engineering.
- [11] Alpaydın E., 2011. "Yapay Öğrenme". Boğaziçi Yayınları.
- [12] J. Graupmann, M. Biwer, C. Zimmer, P. Zimmer, M. Bender, M. Theobald, G. Weikum, 2004. "COMPASS: A concept – based Web Search Engine for HTML, XML and Deep Web Data". Max-Planck Institute for Computer Science.
- [13] M.Diligenti, F.M. Coetzee, S.Lawrence, C.L. Giles, M. Gori, 2000. "Focused Crawling Using Context Graphs". NEC Research Institute.
- [14] Yajun Du, Haiming Li, Zheng Pei, Hong Peng, 2005. "Intelligent Spider's Algorithm of Search Engine Based on Keyword".
- [15] Shanmugasundaram Hariharan, 2009. "Search Results Optimization & Summarization – Case study on Google". Georgian Electronic Scientific Journal: Computer Science and Telecommunications.
- [16] J. Cohen, 1960. "A coefficient of agreement for nominal scales". Educational and Psychological Measurement Vol. XX, No. 1.
- [17] Landis, J.R., Koch, G.G., 1977, "The measurement of observer agreement for categorical data", Biometrics. Vol. 33, 159-174

ÖZGEÇMİŞ

Kişisel

Adı Soyadı : Can RAZBONYALI

Doğum Tarihi : 14.12.1984

Doğum Yeri : Ankara

Eğitim

2009 – 2011	Trakya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümü
2003 – 2007	Maltepe Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü
2002 – 2003	Özel Marmara Koleji
1997 – 2002	Burak Bora Anadolu Lisesi

İş Deneyimi

2011 –	BYM Epistem, Lotus Notes Uzman Yardımcısı
2007 – 2008	Tradesof, Yazılım Uzman Yardımcısı

Dil

İyi derecede İngilizce (Okuma, Yazma, Konuşma)

EK A – AĞ ÖRÜMCEĞİ

```

using System;
using System.Collections;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using System.Threading;
using System.Data.SqlClient;
using System.Net;
using System.IO;
using System.Text.RegularExpressions;
using System.Diagnostics;
using AramaMotoru.Class;
using Kumele;
using weka;
using System.Data.OleDb;

namespace AramaMotoru
{

publicpartialclassForm1 : Form
{
publicdouble TreeDegerlendir()
{

double sonuckod = 0;
try
{
// Eğitim Basladi.
// insts içinde %100 tüm eğitim satırları var
weka.core.Instances insts = new weka.core.Instances(new
java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya Üniversitesi\\Tez ve
Seminer\\Tez\\CANDATA\\Egitim.arff"));
insts.setClassIndex(insts.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
// Burada Sınıflama Algoritması seçiliyor. Tree seçmek en etkin olabilir, daha iyi sınıflar.
weka.classifiers.Classifier cl = new weka.classifiers.trees.J48();
// burada öğrenirken işlenmesi gereken eğitim satırlarını rasgele sırala yapıldı.
weka.filters.Filter myRandom = new weka.filters.unsupervised.instance.Randomize();
myRandom.setInputFormat(insts);
insts = weka.filters.Filter.useFilter(insts, myRandom);
// Rasgele Sırala bitti.
weka.core.Instances train = new weka.core.Instances(insts, 0, insts.numInstances());
cl.buildClassifier(train);
// Eğitim bitti.

//string Metin = "seminer ve konferans var konferans konferans toplanti sempozyum";
Kumele.KumeClass MetinKumele = new Kumele.KumeClass();
MetinKumele.AttributeAdeti = 17;

```

```

    MetinKumele.AttributeAta();
    MetinKumele.KelimeAdetHesapla(MetinTxt.Text.ToString());
    MetinKumele.EgitmenSatiriOzellikYazdir();
    weka.core.Instances test = new
    weka.core.Instances(new java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\Test.arff"));
    test.setClassIndex(test.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
    weka.core.Instance currentInst = test.instance(0);
double predictedClass = cl.classifyInstance(currentInst);
    sonuckod = predictedClass;
// MessageBox.Show(predictedClass.ToString());
    }
catch (java.lang.Exception ex)
    {
        ex.printStackTrace();
    }
    MessageBox.Show("Hata");
    }
return sonuckod;
    }
public double IbkDegerlendir()
    {
double sonuckod = 0;
try
    {
// Egitim Basladi.
// insts içinde %100 tüm eğitim satırları var
    weka.core.Instances insts = new weka.core.Instances(new
    java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya Üniversitesi\\Tez ve
    Seminer\\Tez\\CANDATA\\Egitim.arff"));
    insts.setClassIndex(insts.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
// Burada Sınıflama Algoritması seçiliyor. Tree seçmek en etkin olabilir, daha iyi sınıflar.
    weka.classifiers.Classifier cl = new weka.classifiers.lazy.IBk();
// burada öğrenirken işlenmesi gereken eğitim satırlarını rasgele sırala yapıldı.
    weka.filters.Filter myRandom = new weka.filters.unsupervised.instance.Randomize();
    myRandom.setInputFormat(insts);
    insts = weka.filters.Filter.useFilter(insts, myRandom);
// Rasgele Sırala bitti.
    weka.core.Instances train = new weka.core.Instances(insts, 0, insts.numInstances());
    cl.buildClassifier(train);
// Egitim bitti.
    weka.core.Instances test = new weka.core.Instances(new
    java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya Üniversitesi\\Tez ve
    Seminer\\Tez\\CANDATA\\Test.arff"));
    test.setClassIndex(test.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
    weka.core.Instance currentInst = test.instance(0);
double predictedClass = cl.classifyInstance(currentInst);
    sonuckod = predictedClass;
// MessageBox.Show(predictedClass.ToString());
    }
catch (java.lang.Exception ex)
    {
        ex.printStackTrace();
    }
return sonuckod;
    }

```

```

SqlConnection conn = newSqlConnection("Data Source=Razbon-PC\\SQLEXPRESS;Initial
    Catalog=AramaMotoru;Integrated Security=SSPI");
SqlConnection connInsert = newSqlConnection("Data Source=Razbon-PC\\SQLEXPRESS;Initial
    Catalog=AramaMotoru;Integrated Security=SSPI");
public Form1()
    {
        InitializeComponent();
    }
privatevoid btnStartCrawl_Click(object sender, EventArgs e)
    {
        prgbİndirilecek.Minimum = 0;
        prgbİndirilecek.Maximum = 100000;
        prgbİndirilecek.Value = 100000;
        prgbİndirilen.Minimum = 0;
        prgbİndirilen.Maximum = 100000;
        prgbİndirilen.Value = 0;
        prgbSüreç.Minimum = 0;
        prgbSüreç.Maximum = 100000;
        prgbSüreç.Value = 0;
        int bulunanLinkler_Öncelik = 1;
        OleDbConnection connExcel = newOleDbConnection("Provider= Microsoft.ACE.OLEDB.12.0;Data
            Source=D:\\CR_Doküman\\Yüksek Lisans\\Trakya Üniversitesi\\Tez ve Seminer\\Tez\\Web
            Sayfaları\\Toplu Web Adresleri.xlsx; Extended Properties=\"Excel 12.0;HDR=YES;\"");
        OleDbDataReader dr;
        OleDbCommand commExcel = newOleDbCommand("SELECT * FROM [Sheet1$]", connExcel);
        StreamWriter topluSonucTree, topluSonucIBK;
        connExcel.Open();
        dr = commExcel.ExecuteReader();
        while (dr.Read())
            {
                // MessageBox.Show(dr[0].ToString().Trim());
                LinkLbl.Text = dr[0].ToString().Trim();
                LinkLbl.Refresh();
                spider crawldata = newspider();
                RichTextBox rtxtHTML1 = newRichTextBox();
                rtxtHTML1.Text = crawldata.sayfaIndir(dr[0].ToString().Trim(),bulunanLinkler_Öncelik);
                if (crawldata.last_error != null)
                    {
                        rtxtHTML1.Text = crawldata.last_error;
                    }
                //if
                label1.Text = "Karakter Sayısı : " + rtxtHTML1.Text.Length.ToString();
                //ayrıştırma işlemi
                HTMLparser ayrıştır = newHTMLparser();
                MetinTxt.Text = ayrıştır.parse(rtxtHTML1.Text.ToString());
                MetinTxt.Refresh();
                // Metnin Tipine karar verme işlemi başladı.
                double sonuc1, sonuc2;
                sonuc1 = TreeDegerlendir();
                sonuc2 = IbkDegerlendir();
                Kumele.KumeClass MetinKumele = new Kumele.KumeClass();
                MetinKumele.AttributeAta();
                topluSonucTree = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
                    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\TopluSonucTree.txt", true);
                topluSonucIBK = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
                    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\toplusonucIBK.txt", true);
                if (sonuc1 == sonuc2)

```



```

{
    SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1);
    SinifiLbl.Refresh();
    SqlCommand commLink_Ekle = newSqlCommand("INSERT INTO Okul_AnaSayfalar "
        + "VALUES(" + (int)bulunanLinkler_Öncelik + "," +
        LinkLbl.Text.ToString() + "," +
        SinifiLbl.Text.ToString() + ")")
        , conn);

    conn.Open();
    commLink_Ekle.ExecuteNonQuery();
    conn.Close();
    topluSonucTree.Write(SinifiLbl.Text.ToString());
    topluSonucTree.WriteLine();
    topluSonucTree.Close();
    topluSonucIBK.Write(SinifiLbl.Text.ToString());
    topluSonucIBK.WriteLine();
    topluSonucIBK.Close();
}
elseif (sonuc1 == 6 || sonuc2 == 6)
{
    SinifiLbl.Text = MetinKumele.AttributeDondur(6);
    SinifiLbl.Refresh();
    SqlCommand commLink_Ekle = newSqlCommand("INSERT INTO Okul_AnaSayfalar "+
        "VALUES(" + (int)bulunanLinkler_Öncelik + "," +
        LinkLbl.Text.ToString() + "," +
        SinifiLbl.Text.ToString() + ")")
        , conn);

    conn.Open();
    commLink_Ekle.ExecuteNonQuery();
    conn.Close();
    topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
    topluSonucTree.WriteLine();
    topluSonucTree.Close();
    topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
    topluSonucIBK.WriteLine();
    topluSonucIBK.Close();
}
else
{
    SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1) + " Farklı " +
        MetinKumele.AttributeDondur((int)sonuc2);
    SinifiLbl.Refresh();
    SqlCommand commLink_Ekle = newSqlCommand("INSERT INTO Okul_AnaSayfalar "
        + "VALUES(" + (int)bulunanLinkler_Öncelik + "," +
        LinkLbl.Text.ToString() + "," +
        MetinKumele.AttributeDondur((int)sonuc1) + ")")
        , conn);

    conn.Open();
    commLink_Ekle.ExecuteNonQuery();
    conn.Close();
    topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
    topluSonucTree.WriteLine();
    topluSonucTree.Close();
    topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
    topluSonucIBK.WriteLine();
    topluSonucIBK.Close();
}

```

```

    }
    // Metnin Tipine karar verme işlemi bitti.
    prgbİndirilecek.Value = prgbİndirilecek.Value - 1;
    prgbİndirilen.Value = prgbİndirilen.Value + 1;
    prgbSüreç.Value = prgbSüreç.Value + 1;

    }//while
    connExcel.Close();
    SqlCommand commLinkleriAl = newSqlCommand("SELECT Link FROM Linkler1", conn);
    SqlDataReader dr_Linkler;
    conn.Open();
    dr_Linkler = commLinkleriAl.ExecuteReader();
    bulunanLinkler_Öncelik += 1;
    while(dr_Linkler.Read())
    {
        LinkLbl.Text = dr_Linkler[0].ToString().Trim();
        LinkLbl.Refresh();
        spider crawldata = newspider();
        RichTextBox rtxtHTML1 = newRichTextBox();
        rtxtHTML.Text = crawldata.sayfaİndir(dr_Linkler[0].ToString().Trim(),
        bulunanLinkler_Öncelik);
        if (crawldata.last_error != null)
        {
            rtxtHTML.Text = crawldata.last_error;
        }//if
        label1.Text = "Karakter Sayısı : " + rtxtHTML.Text.Length.ToString();
        //ayrıştırma işlemi
        HTMLparser ayrıştır = newHTMLparser();
        MetinTxt.Text = ayrıştır.parse(rtxtHTML.Text.ToString());
        MetinTxt.Refresh();
        // Metnin Tipine karar verme işlemi başladı.
        double sonuc1, sonuc2;
        sonuc1 = TreeDegerlendir();
        sonuc2 = IbkDegerlendir();
        Kumele.KumeClass MetinKumele = new Kumele.KumeClass();
        MetinKumele.AttributeAta();
        MetinTxt.Text = MetinTxt.Text.Replace("", "").Replace("/", "").Replace("//", "").Replace("-",
        "");
        topluSonucTree = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
        Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\TopluSonucTree.txt", true);
        topluSonucIBK = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
        Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\topluSonucIBK.txt", true);
        if (sonuc1 == sonuc2)
        {
            SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1);
            SinifiLbl.Refresh();

            SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler1 " +
            "SET Karar = " +
            SinifiLbl.Text.ToString() + "," +
            "Text=" +
            MetinTxt.Text + " " +
            "WHERE Link=" +
            dr_Linkler[0].ToString() + ""
            , connInsert);

```

```

try
    {
        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(SinifiLbl.Text.ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
        topluSonucIBK.Write(SinifiLbl.Text.ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
catch(Exception hata)
    {
        commKararGuncelle.Dispose();
        connInsert.Close();
        connInsert.Dispose();
    }
}
elseif (sonuc1 == 6 || sonuc2 == 6)
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur(6);
        SinifiLbl.Refresh();
        SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler1 " +
            "SET Karar = " +
            SinifiLbl.Text.ToString() + "," +
            "Text = " +
            MetinTxt.Text + " " +
            "WHERE Link = " +
            dr_Linkler[0].ToString() + """,
            connInsert);

        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
        topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
else
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1) + " Farklı " +
        MetinKumele.AttributeDondur((int)sonuc2);
        SinifiLbl.Refresh();
        SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler1 " +
            "SET Karar = " +
            MetinKumele.AttributeDondur((int)sonuc1) + "," +
            "Text = " +
            MetinTxt.Text + " " +
            "WHERE Link = " +
            dr_Linkler[0].ToString() + """,
            connInsert);

        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
    }
}

```

```

connInsert.Close();
topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
topluSonucTree.WriteLine();
topluSonucTree.Close();
topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
topluSonucIBK.WriteLine();
topluSonucIBK.Close();    }

prgbİndirilecek.Value = prgbİndirilecek.Value - 1;
prgbİndirilen.Value = prgbİndirilen.Value + 1;
prgbSüreç.Value = prgbSüreç.Value + 1;
} //while
conn.Close();
SqlCommand commLinkler2A1 = newSqlCommand("SELECT Link FROM Linkler2", conn);
SqlDataReader dr_Linkler2;
conn.Open();
dr_Linkler2 = commLinkler2A1.ExecuteReader();

bulunanLinkler_Öncelik += 1;
while (dr_Linkler2.Read())
{
    LinkLbl.Text = dr_Linkler2[0].ToString().Trim();
    LinkLbl.Refresh();
spider crawldata = newspider();
RichTextBox rtxtHTML1 = newRichTextBox();
    rtxtHTML.Text = crawldata.sayfaIndir(dr_Linkler2[0].ToString().Trim(),
bulunanLinkler_Öncelik);
if (crawldata.last_error != null)
    {
        rtxtHTML.Text = crawldata.last_error;
    } //if
    label1.Text = "Karakter Sayısı : " + rtxtHTML.Text.Length.ToString();
//ayrıştırma işlemi
HTMLparser ayrıştır = newHTMLparser();
    MetinTxt.Text = ayrıştır.parse(rtxtHTML.Text.ToString());
    MetinTxt.Refresh();
// Metnin Tipine karar verme işlemi başladı.
double sonuc1, sonuc2;
    sonuc1 = TreeDegerlendir();
    sonuc2 = IbkDegerlendir();
    Kumele.KumeClass MetinKumele = new Kumele.KumeClass();
    MetinKumele.AttributeAta();
    MetinTxt.Text = MetinTxt.Text.Replace("","").Replace("/","").Replace("/", "").Replace("-",
""");
    topluSonucTree = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\TopluSonucTree.txt", true);
    topluSonucIBK = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\topluSonucIBK.txt", true);
if (sonuc1 == sonuc2)
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1);
        SinifiLbl.Refresh();
SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler2 " +
        "SET Karar = " +
        SinifiLbl.Text.ToString() + "," +
        "Text = " +

```

```

        MetinTxt.Text + "" " +
        "WHERE Link = ""      +
        dr_Linkler2[0].ToString() + """"
        , connInsert);

    connInsert.Open();
    commKararGuncelle.ExecuteNonQuery();
    connInsert.Close();
    topluSonucTree.Write(SinifiLbl.Text.ToString());
    topluSonucTree.WriteLine();
    topluSonucTree.Close();
    topluSonucIBK.Write(SinifiLbl.Text.ToString());
    topluSonucIBK.WriteLine();
    topluSonucIBK.Close();
    }
elseif (sonuc1 == 6 || sonuc2 == 6)
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur(6);
        SinifiLbl.Refresh();
        SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler2 "      +
            "SET Karar = ""      +
            SinifiLbl.Text.ToString() + ""," +
            "Text = "" +
            MetinTxt.Text + "" " +
            "WHERE Link = ""      +
            dr_Linkler2[0].ToString() + """"
            , connInsert);

        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
        topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
else
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1) + " Farklı " +
            MetinKumele.AttributeDondur((int)sonuc2);
        SinifiLbl.Refresh();
        SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler2 " +
            "SET Karar = ""      +
            MetinKumele.AttributeDondur((int)sonuc1) + ""," +
            "Text = "" +
            MetinTxt.Text + "" " +
            "WHERE Link = ""      +
            dr_Linkler2[0].ToString() + """"
            , connInsert);

        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
    }

```

```

        topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
    prgbİndirilecek.Value = prgbİndirilecek.Value - 1;
    prgbİndirilen.Value = prgbİndirilen.Value + 1;
    prgbSüreç.Value = prgbSüreç.Value + 1;
} //while
conn.Close();
SqlCommand commLinkler3Al = newSqlCommand("SELECT Link FROM Linkler3", conn);
SqlDataReader dr_Linkler3;
conn.Open();
dr_Linkler3 = commLinkler3Al.ExecuteReader();
bulunanLinkler_Öncelik += 1;
while (dr_Linkler3.Read())
{
    LinkLbl.Text = dr_Linkler3[0].ToString().Trim();
    LinkLbl.Refresh();
spider crawldata = newspider();
RichTextBox rtxtHTML1 = newRichTextBox();
    rtxtHTML.Text = crawldata.sayfaIndir(dr_Linkler3[0].ToString().Trim(),
bulunanLinkler_Öncelik);
if (crawldata.last_error != null)
    {
        rtxtHTML.Text = crawldata.last_error;
    } //if
    label1.Text = "Karakter Sayısı : " + rtxtHTML.Text.Length.ToString();
//ayırıştırma işlemi
HTMLparser ayırıştır = newHTMLparser();
    MetinTxt.Text = ayırıştır.parse(rtxtHTML.Text.ToString());
    MetinTxt.Refresh();
// Metnin Tipine karar verme işlemi başladı.
double sonuc1, sonuc2;
sonuc1 = TreeDegerlendir();
sonuc2 = IbkDegerlendir();
Kumele.KumeClass MetinKumele = new Kumele.KumeClass();
MetinKumele.AttributeAta();
MetinTxt.Text = MetinTxt.Text.Replace(" ", "").Replace("/", "").Replace(" ", "").Replace("-",
""");
    topluSonucTree = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\TopluSonucTree.txt", true);
    topluSonucIBK = newStreamWriter("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
    Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\topluSonucIBK.txt", true);
if (sonuc1 == sonuc2)
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1);
        SinifiLbl.Refresh();
SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler3 " +
        "SET Karar = " +
        SinifiLbl.Text.ToString() + "," +
        "Text = " +
        MetinTxt.Text + " " +
        "WHERE Link = " +
        dr_Linkler3[0].ToString() + ""
        , connInsert);

        connInsert.Open();

```

```

        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(SinifiLbl.Text.ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
        topluSonucIBK.Write(SinifiLbl.Text.ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
    elseif (sonuc1 == 6 || sonuc2 == 6)
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur(6);
        SinifiLbl.Refresh();
        SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler3 " +
            "SET Karar = " +
            SinifiLbl.Text.ToString() + "," +
            "Text = " +
            MetinTxt.Text + " " +
            "WHERE Link = " +
            dr_Linkler3[0].ToString() + ""
            , connInsert);

        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
        topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
    else
    {
        SinifiLbl.Text = MetinKumele.AttributeDondur((int)sonuc1) + " Farklı " +
            MetinKumele.AttributeDondur((int)sonuc2);
        SinifiLbl.Refresh();
        SqlCommand commKararGuncelle = newSqlCommand("UPDATE Linkler3 " +
            "SET Karar = " +
            MetinKumele.AttributeDondur((int)sonuc1) + "," +
            "Text = " +
            MetinTxt.Text + " " +
            "WHERE Link = " +
            dr_Linkler3[0].ToString() + ""
            , connInsert);

        connInsert.Open();
        commKararGuncelle.ExecuteNonQuery();
        connInsert.Close();
        topluSonucTree.Write(MetinKumele.AttributeDondur((int)sonuc1).ToString());
        topluSonucTree.WriteLine();
        topluSonucTree.Close();
        topluSonucIBK.Write(MetinKumele.AttributeDondur((int)sonuc2).ToString());
        topluSonucIBK.WriteLine();
        topluSonucIBK.Close();
    }
    prgBİndirilecek.Value = prgBİndirilecek.Value - 1;

```

```
        prgBİndirilen.Value = prgBİndirilen.Value + 1;
        prgBSüreç.Value = prgBSüreç.Value + 1;
    }//while
    conn.Close();
    if (bulunanLinkler_Öncelik == 3)
    {
        MessageBox.Show("İşlem Bitti");
        Application.Exit();
    }
    }//btnStartCrawl_Click
private void btnAyrıştır_Click(object sender, EventArgs e)
{
    HTMLparser ayrıştır = new HTMLparser();
    MetinTxt.Text = ayrıştır.parse(rtxtHTML.Text.ToString());
    SinifiLbl.Text = "Karakter Sayısı : " + MetinTxt.Text.Length.ToString();
    }//btnAyrıştır_Click
}
}
```


EK B – spider.cd Sınıfı

```

using System;
using System.Collections;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using System.Threading;
using System.Data.SqlClient;
using System.Net;
using System.IO;
using System.Text.RegularExpressions;
using System.Diagnostics;
namespace AramaMotoru.Class
{
publicclassspider
{
publicstring URL;
publicstring default_url;
publicstring SaltText;
publicstring HTML;
publicstring last_error;
publicBoolean ozelDurum;
publicWebResponse response;
publicEncoding last_encoding;
publicint error_cnt = 0;
privateHashtable linkler = newHashtable();
ArrayList arrayLinkler = newArrayList();
privateWebResponse GetResponse(string url)
{
try
{
HttpWebRequest request;
request = (HttpWebRequest)WebRequest.Create(url);
request.KeepAlive = false;
request.Method = "GET";
request.AllowAutoRedirect = false;
if (ozelDurum == false)
{
request.ContentType = "application/x-www-form-urlencoded";
request.MediaType = "text/html;charset=utf-8";
request.UserAgent = "MSIE 7.0";
}
return request.GetResponse();
}
catch (WebException e)
{
last_error = "Web sayfası açılmadı: " + e.Message;
returnnull;
}
}
}

```

```

    }
}
publicstring sayfaIndir(string URL,int bulunanLinklerOncelik)
{
    p: WebResponse response = GetResponse(URL);
    if (response != null)
    {
        if (last_encoding == null)
            last_encoding = Encoding.Default;
        StreamReader reader = newStreamReader(response.GetResponseStream(), last_encoding);
        //icerik deęerinin ierisinde uygun encoding yapılıř HTML kodu var
        string icerik = reader.ReadToEnd();
        if (error_cnt == 2)
        {
            last_error = "2 kere aynı sayfa istendi...";
            return"";
        }
        if (icerik.Contains("◆"))
        {
            error_cnt++;
            last_encoding = Encoding.Default;
            //icerik = sayfaIndir(URL);
            goto p;
        }
        if (icerik.Contains("Ä") || icerik.Contains("Ă"))
        {
            error_cnt++;
            last_encoding = Encoding.UTF8;
            //icerik = sayfaIndir(URL);
            goto p;
        }
        linkleri_Tablola(icerik, URL, bulunanLinklerOncelik);

        return TurkceKarakterDuzelt(icerik);
    }
    elsereturn"";
}

publicvoid crawl()
{
    StreamReader sr;
    last_encoding = Encoding.Default;
    response = GetResponse(URL);
    sr = newStreamReader(response.GetResponseStream(), last_encoding);
    last_encoding = Encoding.Default;
    HTML = sr.ReadToEnd();
}
privatestring TurkceKarakterDuzelt(string ayrık)
{
    ayrık = ayrık.Replace("&#252;", "ü");
    ayrık = ayrık.Replace("&#220;", "Ū");
    ayrık = ayrık.Replace("&#231;", "ç");
    ayrık = ayrık.Replace("&#199;", "Ç");
    ayrık = ayrık.Replace("&#351;", "ş");
    ayrık = ayrık.Replace("&#350;", "Ş");
    ayrık = ayrık.Replace("&#286;", "Ğ");
}

```

```

ayrık = ayrık.Replace("&#287;", "ğ");
ayrık = ayrık.Replace("&#305;", "ı");
ayrık = ayrık.Replace("&#304;", "İ");
ayrık = ayrık.Replace("&#246;", "ö");
ayrık = ayrık.Replace("&#214;", "Ö");
ayrık = ayrık.Replace("&uuml;", "ü");
ayrık = ayrık.Replace("&Uuml;", "Ü");
ayrık = ayrık.Replace("&ccedil;", "ç");
ayrık = ayrık.Replace("&Ccedil;", "Ç");
ayrık = ayrık.Replace("&ouml;", "ö");
ayrık = ayrık.Replace("&Ouml;", "Ö");
ayrık = ayrık.Replace("Ý", "İ");
ayrık = ayrık.Replace("Đ", "Ğ");
ayrık = ayrık.Replace("Đ", "Ş");
ayrık = ayrık.Replace("đ", "ğ");
ayrık = ayrık.Replace("ý", "ı");
ayrık = ayrık.Replace("đ", "ş");
return ayrık;
}
public Boolean linkleri_Tablola(string icerik, string baslamaURL, int bulunanLinkler_Öncelik)
{
SqlConnection conn = new SqlConnection("Data Source=Razbon-PC\\SQLEXPRESS;Initial
Catalog=AramaMotoru;Integrated Security=SSPI");
string anaURL = FindBaseUrl(baslamaURL);
int link_sayısı = 0;
string pattern = "href=\"(.*)\"";
Regex exp = new Regex(pattern, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
MatchCollection matchList = exp.Matches(icerik);
for (int i = 0; i < matchList.Count; i++)
{
Match match = matchList[i];
if (match.Value.Length > 0)
{
string web_page = match.Groups[1].Value;
web_page = web_page.ToLower();
linkler.Add(link_sayısı++, web_page);
}
}
int n = link_sayısı;
string contain;
// popup varsa çekmek için
string pattern2 = @"javascript:pop_up\('(.*?)'";
Regex exp2 = new Regex(pattern2, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
matchList = exp2.Matches(icerik);
for (int i = 0; i < matchList.Count; i++)
{
Match match = matchList[i];
if (match.Value.Length > 0)
{
string web_page = match.Groups[1].Value;
web_page = web_page.ToLower();
contain = anaURL + web_page;
if (web_page.Length != 0 && linkler.ContainsValue(contain) == false)
{

```

```

if (web_page != "")
    {
        linkler.Add(link_sayısı++, anaURL + web_page);
    }
}

// bitti
// href çift tırnaklı çekmek için
string pattern3 = @"http://.*?/";
Regex exp3 = new Regex(pattern3, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
matchList = exp3.Matches(icerik);
for (int i = 0; i < matchList.Count; i++)
    {
Match match = matchList[i];
if (match.Value.Length > 0)
    {
string web_page = match.Groups[1].Value;
web_page = web_page.ToLower();
contain = anaURL + web_page;
if (web_page.Length != 0 && linkler.ContainsValue(contain) == false)
    {
if (web_page != "")
        {
            linkler.Add(link_sayısı++, anaURL + web_page);
        }
    }
    }
}

// bitti
// href tek tırnaklı çekmek için
pattern = "href=\"(.*)\"";
exp = new Regex(pattern, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
matchList = exp.Matches(icerik);
for (int i = 0; i < matchList.Count; i++)
    {
Match match = matchList[i];
if (match.Value.Length > 0)
    {
string web_page = match.Groups[1].Value;
web_page = web_page.ToLower();
linkler.Add(link_sayısı++, web_page);
    }
}

// Bitti
/// java script ajaxsayfa cekmek için
pattern = "ajaxsayfa.\"(.*)\"";
exp = new Regex(pattern, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
matchList = exp.Matches(icerik);
for (int i = 0; i < matchList.Count; i++)
    {

```

```

Match match = matchList[i];
if (match.Value.Length > 0)
    {
string web_page = match.Groups[1].Value;
    web_page = web_page.ToLower();
    linkler.Add(link_sayısı++, web_page);
    }
}
// bitti

/// java script window open cekmek için
pattern = "window.open.\\(.*?)\\";
exp = new Regex(pattern, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
matchList = exp.Matches(icerik);
for (int i = 0; i < matchList.Count; i++)
    {
Match match = matchList[i];
if (match.Value.Length > 0)
    {
string web_page = match.Groups[1].Value;
    web_page = web_page.ToLower();
    linkler.Add(link_sayısı++, web_page);
    }
}
// bitti
for (int i = 0; i < linkler.Count; i++)
    {
if (linkler[i].ToString().Contains("../") == true ||
    linkler[i].ToString().Trim().Contains("../") == true ||
    linkler[i].ToString().Trim().Contains("../..") == true)
        {
            linkler[i] = linkler[i].ToString().Replace("../", "").Replace("../..", "").Replace("../..", "");
        }
}
if (anaURL.Substring(anaURL.Length - 3, 3) == "/tr")
    {
for (int i = 0; i < linkler.Count; i++)
    {
if (linkler[i].ToString().Length > 3)
        {
if (linkler[i].ToString().Substring(0, 3) == "tr")
            {
                linkler[i] = linkler[i].ToString().Remove(0, 3);
            }
        }
    }
}
else
    {
        linkler[i] = "";
    }
}
StreamWriter sw = new StreamWriter("C:\\Users\\RazboN\\Desktop\\linkler.txt", true);
for (int i = 0; i < linkler.Count; i++)
    {
string kurum=null;
string[] kurumAltIsmi;
if (anaURL.Contains("http")==true&& anaURL.Contains("www")==true)

```

```

        {
            kurum = anaURL.Substring(11, anaURL.Length - 11);
        }
if (anaURL.Contains("http") == true&& anaURL.Contains("www") == false)
    {
        kurum = anaURL.Substring(7, anaURL.Length - 7);
    }
if (kurum.Contains("/"))
    {
        kurumAltIsmi = kurum.Split(newchar[] { '/' }, StringSplitOptions.RemoveEmptyEntries);
        kurum = kurumAltIsmi[0].ToString();
    }
    kurumAltIsmi = kurum.Split(newchar[] { '.' }, StringSplitOptions.RemoveEmptyEntries);
if (kurumAltIsmi.Length==4)
    {
        kurum = kurumAltIsmi[1].ToString() + '.' +
            kurumAltIsmi[2].ToString() + '.' + kurumAltIsmi[5].ToString();
    }
if (linkler[i].ToString().Contains(".com.") == false&&
    linkler[i].ToString().Contains(".edu.") == false&&
    linkler[i].ToString().Contains(".gov.") == false&&
    linkler[i].ToString().Contains(kurum) == false&&
    linkler[i].ToString().Contains("http://") == false&&
    linkler[i].ToString().Contains("www") == false&&
    linkler[i].ToString().Contains(".com") == false)
    {
if (linkler[i].ToString() != "")
    {
if (linkler[i].ToString().Substring(0, 1) == "/")
    {
if (kurum == "pau.edu.tr")
    {
if (linkler[i].ToString().Length <= 20)
    {
        linkler[i] = "http://www." + kurum + linkler[i] + "/default.aspx";
    }
elseif(linkler[i].ToString().Contains("/pau/") == true)
    {
        linkler[i] = linkler[i] + "/default.aspx";
    }
else
    {
        linkler[i] = "http://www." + kurum + linkler[i];
    }
else
    {
        linkler[i] = baslamaURL.Remove(baslamaURL.Length - 1, 1) + linkler[i];
    }
    }
else
    {
if (baslamaURL.Substring(baslamaURL.Length - 1) == "/")
    {
        linkler[i] = baslamaURL + linkler[i];
    }
    }
    }
    }

```

```

    }
else
    {
        linkler[i] = baslamaURL + "/" + linkler[i];
    }
}
}
// Kurum dışındaki ilgisiz yerleri çekmeyi engellemek için bu satır yazıldı.
if (linkler[i].ToString().Contains(kurum) == false)
    {
        linkler[i] = "";
    }
elseif (linkler[i].ToString().Contains("http://") == false ||
        linkler[i].ToString().Contains("www") == false)
    {
        if (linkler[i].ToString().Contains(".com.") == false&&
            linkler[i].ToString().Contains(".edu.") == false&&
            linkler[i].ToString().Contains(".gov.") == false&&
            linkler[i].ToString().Contains(kurum) == false&&
            linkler[i].ToString().Contains("http://") == true&&
            linkler[i].ToString().Contains("www") == false)
            {
                linkler[i] = linkler[i].ToString().Insert(7, anaURL.Remove(0,7));
            }
        elseif (linkler[i].ToString().Contains(".com.") == false&&
            linkler[i].ToString().Contains(".edu.") == false&&
            linkler[i].ToString().Contains(".gov.") == false&&
            linkler[i].ToString().Contains(kurum) == false&&
            linkler[i].ToString().Substring(0, 1) == "/")
            {
                linkler[i] = linkler[i].ToString().Remove(0, 1);
                linkler[i] = anaURL + linkler[i];
            }

        elseif (linkler[i].ToString().Contains("http://") == false&& linkler[i].ToString().Contains("www") ==
            false)
            {
                if (linkler[i].ToString().Substring(0, 1) == "/")
                    {
                        linkler[i] = kurum + linkler[i];
                    }
                else
                    {
                        linkler[i] = kurum + "/" + linkler[i];
                    }
            }
    }
else
    {
        if (linkler[i].ToString().Contains("tr") == false&&
            FindBaseUrl(baslamaURL).ToString().Contains("sabanciuniv") == true)
            {
                linkler[i] = linkler[i].ToString() + "tr/";
            }
    }
}

```

```

    }
    for (int i = 0; i < link_sayısı; i++)
    {
        if (baslamaURL.ToLower() == linkler[i].ToString().ToLower())
        {
            linkler.Remove(i);
        }
        elseif (linkler.Contains(linkler[i].ToString()) == true)
        {
            linkler.Remove(i);
        }
    }
    arrayLinkler.AddRange(linkler.Values);
    arrayLinkler.Sort();
    for (int j = 0; j < arrayLinkler.Count - 1; j++)
    {
        for (int i = j + 1; i < arrayLinkler.Count; i++)
        {
            if (arrayLinkler[j].ToString().ToLower() == arrayLinkler[i].ToString().ToLower() ||
                arrayLinkler[j].ToString().ToLower() == anaURL.ToLower())
            {
                arrayLinkler[i] = "";
            }
        }
    }
    if (arrayLinkler[i].ToString() != "")
    {
        if (arrayLinkler[i].ToString().Contains(".gif") == true ||
            arrayLinkler[i].ToString().Contains(".GIF") == true ||
            arrayLinkler[i].ToString().Contains(".jpg") == true ||
            arrayLinkler[i].ToString().Contains(".JP") == true ||
            arrayLinkler[i].ToString().Contains(".Jp") == true ||
            arrayLinkler[i].ToString().Contains(".jpeg") == true ||
            arrayLinkler[i].ToString().Contains(".xml") == true ||
            arrayLinkler[i].ToString().Contains("mailto") == true ||
            arrayLinkler[i].ToString().Contains("#") == true ||
            arrayLinkler[i].ToString().Contains(".CSS") == true ||
            arrayLinkler[i].ToString().Contains("javascript:") == true ||
            arrayLinkler[i].ToString().Contains("JAVASCRIPT:") == true ||
            arrayLinkler[i].ToString().Contains("Javascript:") == true ||
            arrayLinkler[i].ToString().Contains(".CSS") == true ||
            arrayLinkler[i].ToString().Contains(".css") == true ||
            arrayLinkler[i].ToString().Contains("image.") == true ||
            arrayLinkler[i].ToString().Contains(".pdf") == true ||
            arrayLinkler[i].ToString().Contains("this") == true ||
            arrayLinkler[i].ToString().Contains("/eng") == true ||
            arrayLinkler[i].ToString().Contains(".doc") == true ||
            arrayLinkler[i].ToString().Contains("webmail") == true ||
            arrayLinkler[i].ToString().Contains("target") == true ||
            arrayLinkler[i].ToString().Contains("href") == true ||
            arrayLinkler[i].ToString().Contains(".ico") == true ||
            arrayLinkler[i].ToString().Contains("erasmus") == true ||
            arrayLinkler[i].ToString().Contains("mail") == true ||
            arrayLinkler[i].ToString().Contains("posta") == true ||
            arrayLinkler[i].ToString().Contains("/en") == true ||
            arrayLinkler[i].ToString().Contains("kutuphane") == true ||
            arrayLinkler[i].ToString().Contains("ogrenci") == true
        )
    }

```



```

        || arrayLinkler[i].ToString().Contains("ftp") == true ||
        arrayLinkler[i].ToString().Contains("ogr") == true
        || arrayLinkler[i].ToString().Contains("redirect") == true ||
        arrayLinkler[i].ToString().Contains("login") == true)

        arrayLinkler[i] = string.Empty;
    }//if
} //for
} //for
linkler.Clear();
int linkSayısı = 0;
for (int i = 0; i < arrayLinkler.Count; i++)
{
    if (arrayLinkler[i].ToString() != "")
    {
        if (arrayLinkler[i].ToString().Contains("jump") == true&&
            arrayLinkler[i].ToString().Substring(0,1).ToLower() == "r")
        {
            arrayLinkler[i] = arrayLinkler[i].ToString().Remove(0, 22);
            linkler.Add(linkSayısı, arrayLinkler[i]);
            linkSayısı++;
        }
    }
    else
    {
        linkler.Add(linkSayısı, arrayLinkler[i]);
        linkSayısı++;
    }
}
for (int i = 0; i < linkSayısı; i++)
{
    if (arrayLinkler[i].ToString().Contains(".gif") == true ||
        arrayLinkler[i].ToString().Contains(".GIF") == true
        || arrayLinkler[i].ToString().Contains(".jpg") == true || arrayLinkler[i].ToString().Contains(".JP") == true
        || arrayLinkler[i].ToString().Contains(".Jp") == true || arrayLinkler[i].ToString().Contains(".jpeg") == true
        || arrayLinkler[i].ToString().Contains(".xml") == true ||
        arrayLinkler[i].ToString().Contains("mailto") == true
        || arrayLinkler[i].ToString().Contains("#") == true ||
        arrayLinkler[i].ToString().Contains(".CSS") == true
        || arrayLinkler[i].ToString().Contains("javascript:") ||
        arrayLinkler[i].ToString().Contains("JAVASCRIPT:") == true
        || arrayLinkler[i].ToString().Contains("Javascript:") == true ||
        arrayLinkler[i].ToString().Contains(".CSS") == true
        || arrayLinkler[i].ToString().Contains(".css") == true ||
        arrayLinkler[i].ToString().Contains("image.") == true
        || arrayLinkler[i].ToString().Contains(".pdf") == true ||
        arrayLinkler[i].ToString().Contains("this") == true
        || arrayLinkler[i].ToString().Contains("/eng") == true ||
        arrayLinkler[i].ToString().Contains(".doc") == true
        || arrayLinkler[i].ToString().Contains("webmail") == true ||
        arrayLinkler[i].ToString().Contains("target") == true
        || arrayLinkler[i].ToString().Contains("href") == true ||
        arrayLinkler[i].ToString().Contains(".ico")
        || arrayLinkler[i].ToString().Contains("erasmus") == true ||
        arrayLinkler[i].ToString().Contains("mail") == true

```

```

        || arrayLinkler[i].ToString().Contains("posta") == true ||
        arrayLinkler[i].ToString().Contains("/en") == true
        || arrayLinkler[i].ToString().Contains("kutuphane") == true ||
        arrayLinkler[i].ToString().Contains("ogrenci") == true
        || arrayLinkler[i].ToString().Contains("ftp") == true ||
        arrayLinkler[i].ToString().Contains("theslide") == true)

        arrayLinkler[i] = string.Empty;
else
    {
try
    {
if (arrayLinkler[i].ToString() != "")
    {
SqlCommand commLink_Ekle = newSqlCommand("INSERT INTO Linkler" +
(int)bulunanLinkler_Öncelik +
" VALUES("          +
                                (int)bulunanLinkler_Öncelik + "," +
                                linkler[i].ToString() + ",","")
                                , conn);

        conn.Open();
        commLink_Ekle.ExecuteNonQuery();
        conn.Close();
    }
    }//try
catch (Exception ex)
    {
MessageBox.Show(ex.ToString());
    }
    }//else
    }
sw.Close();
returntrue;
    }
publicstring FindBaseUrl(string URL)
    {
string result = "";
string pattern = @"http://.*?/";
Regex exp = newRegex(pattern, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
MatchCollection matchList = exp.Matches(URL);
if (matchList.Count > 0 && matchList[0].ToString().Contains("sabanciuniv.edu") == true)
    {
Match match = matchList[0];
result = match.Groups[0].Value + "tr/";
    }
else
    {
result = URL;
    }
return result;
    }
}
}

```

EK C – HTMLparser.cs Sınıfı

```

using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Text.RegularExpressions;
using System.Collections;
using System.Net;
using System.IO;
using AramaMotoru.Class;
namespace AramaMotoru.Class
{
classHTMLparser
{
publicstring parse(string html)
{
string ayrıştırılmış;
string pattern = @"<table.*?</table>";
Regex exp = newRegex(pattern, RegexOptions.IgnoreCase | RegexOptions.Multiline |
RegexOptions.IgnorePatternWhitespace);
MatchCollection mch;
mch = exp.Matches(html);
ayrıştırılmış = Regex.Replace(html, "<table.*?</table>", string.Empty, RegexOptions.Singleline |
RegexOptions.Multiline);
ayrıştırılmış = Regex.Replace(html, "<script.*?</script>", string.Empty, RegexOptions.Singleline
| RegexOptions.Multiline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "<style.*?</style>", string.Empty,
RegexOptions.Singleline | RegexOptions.Multiline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "<.*?>", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "{.*?}", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "&nbsp;", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "-->", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "&quot;", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "&raquo;", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "&bull;", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "&copy;", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
ayrıştırılmış = Regex.Replace(ayrıştırılmış, "&gt;", string.Empty, RegexOptions.Multiline |
RegexOptions.Singleline);
return ayrıştırılmış;
}
}
}

```

EK D – Eğitim Seti Dokümanı

@relation Bilimsel

@attribute seminer {TRUE, FALSE}

@attribute sempozyum {TRUE, FALSE}

@attribute konferans {TRUE, FALSE}

@attribute kongre {TRUE, FALSE}

@attribute etkinlik {TRUE, FALSE}

@attribute panel {TRUE, FALSE}

@attribute çalistay {TRUE, FALSE}

@attribute diger {TRUE, FALSE}

@attribute katilim {TRUE, FALSE}

@attribute toplantı {TRUE, FALSE}

@attribute oturum {TRUE, FALSE}

@attribute tarih {TRUE, FALSE}

@attribute bilimsel {TRUE, FALSE}

@attribute sunum {TRUE, FALSE}

@attribute bildiri {TRUE, FALSE}

@attribute ulusal {TRUE, FALSE}

@attribute uluslararası {TRUE, FALSE}

@attribute tur {seminer,sempozyum,konferans,kongre,panel,çalistay,diger}

@data

FALSE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,sempozyum

FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,TRUE,TRUE,FALSE,TRUE,FALSE,sempozyum

FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,TRUE,FALSE,sempozyum

FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,sempozyum

TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,FALSE,sempozyum

FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,TRUE,FALSE,TRUE,TRUE,sempozyum

FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,TRUE,FALSE,sempozyum

FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,sempozyum

FALSE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,TRUE,FALSE,TRUE,FALSE,sempozyum

FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,sempozyum

FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,TRUE,FALSE,TRUE,TRUE,FALSE,sempozyum

FALSE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,TRUE,sempozyum

FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,TRUE,TRUE,TRUE,FALSE,TRUE,FALSE,FALSE,TRUE,diger
TRUE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,diger
FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,FALSE,FALSE,FALSE,FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,FALSE,TRUE,diger

EK E – Eğitim Seti Oluşturmada Kullanılan Web Sayfaları

http://www.anayasa.org/index.php?option=com_content&view=article&id=72%3Aanayasa-platformu-ulusal-cars&catid=57%3Aanayasa-platformu&Itemid=55&lang=tr

<http://www.manset61.com/yerel/1-ulusal-hamsi-calistay-sonuclari-aciklandi.htm>

<http://www.bgst.org/keab/ak20080101.asp>

<http://www.haberajans.com/habereGit.php?haberUrl=http://www.haberler.com/ulusal-medyanin-dogu-guneydogu-algisi-konulu-2355639-haberi/>

<http://www.facebook.com/group.php?gid=116667985010488>

<http://kits.ankos.gen.tr/calistay2010/index.php>

http://www.ttb.org.tr/udek/index.php?option=com_content&view=article&id=184:ulusal-yeterlik-kurulu-iii-lyporu&catid=76:lylar&Itemid=143

<http://www.karasuhaber.com/karasudan-son-haberler/mersin-baligi-geriye-doner-mi-4257.html>

<http://www.timeturk.com/tr/2010/11/06/ulusal-medyanin-dogu-guneydogu-algisi-konulu-calistay-sona-erdi.html>

<http://prizma.dogus.edu.tr/?p=1189>

<http://www.webhaber.com/ulusal-medyanin-dogu-guneydogu-algisi-konulu-2355639-haberi>

<http://www.tugev.net/forum/index.php?topic=611.0>

<http://www.samisener.net/default.asp?page=yaptiklarimiz>

<http://www.millet.gr/haber.php?haberid=23701>

<http://www.turkhukuk sitesi.com/calendar.php?do=getinfo&day=2011-3-8&e=1043&c=1>

http://trese.cs.utwente.nl/TSAD/calistay_program.htm

<http://www.kpssnavi.net/dekanlar-ogretmenler-icin-calistay-baslatti-h729.html>

http://www.sosyalsiyaset.net/documents/25122009_3.asp

http://www.sessiztv.net/index.php?option=com_content&view=article&id=175:ii-ulusal-engell-breyler-cn-fzksel-aktvte-calitayi-11-13-ekm-canakkale&catid=1:son-haberler&Itemid=50

<http://www.antraktsinema.com/haber.php?id=855>

<http://www.geredehaber.com/cevre-kent/242-yerel-bolgesel--cevre-kent/1316-1ulusal-tasarim-bolumleri-urun-calistayi-beypazarinda-basladi.html>

<http://aef.marmara.edu.tr/iysdosya/USOC.pdf>

<http://www.turk.internet.com/portal/yazigoster.php?yaziid=18338>

<http://www.adapostasi.com/haber/14988/turkce-calistay-i-duzenlendi.aspx>

<http://www.portal26.com/haber/4563/ulusal-sanat-calistayi-sona-erdi.html>

<http://www.haberler.com/ankara-genelkurmay-tmmm-den-uluslararası-calistay-haberi/>

<http://avrupaninsinirlaraporlaracilis.blogspot.com/2007/11/altay-uluslararası-iler-yoluyla-yerel.html>

<http://www.tumgazeteler.com/?a=1531584>

http://www.yapi.com.tr/Etkinlikler/sinasosta-uluslararası-calistay_46940.html

http://isparta.cevreorman.gov.tr/isparta/AnaSayfa/resimliHaber/10-09-26/Isparta_Uluslararası%C4%B1_%C3%87al%C4%B1%C5%9Ftay_a_ev_sahipli%C4%9Fi_yap%C4%B1yor.aspx?sflang=tr

<http://hurarsiv.hurriyet.com.tr/goster/ShowNew.aspx?id=15780098>

<http://www.haberpan.com/gkurmaydan-uluslararası-medyaa--c2-91calistay-c2-92i-haberi/>

<http://www.turkpatent.gov.tr/portal/default2.jsp?sayfa=703&haber=1164>

<http://www.turizmhaberleri.com/haberayrinti.asp?ID=18966>

<http://www.lpghaber.com/OMU-Ve-Tso%60dan-Uluslararası-Calistay--haberi-233464.html>

<http://www.bighaber.com/haber/1-iznik-konsil039in-toplandigi-senato-sarayi039nin-yeri-uluslararası-calistay-ile-belirlenecek-364362.html>

http://www.dilbilimi.net/2011_iv_uluslararası_turkiye_turkcesi_agiz_arastirmalari_calistayi.pdf

<http://www.haberortak.com/Haber/Iste-Egitim/04052009/Uluslararası-arkeolojik-calistay-kuruldu.html>

http://turkoloji.cu.edu.tr/SEMPOZYUM/sempozyum_duyuru.php

<http://www.cicr.org.tr/?p=announcements>

<http://engelliler.gen.tr/f53/ozurlulerin-haklarina-iliskin-ilk-uluslararası-calistay-6481/>

<http://www.okuloncesiforum.com/konferans-ve-etkinlik-duyurulari/57863-14-uluslararası-egitimde-yaratıcı-drama-ve-tiyatro-calistay.html>

<http://www.hurriyet.com.tr/gundem/15780098.asp>

<http://www.haberler.com/anadolu-da-helenistik-seramik-baslikli-2276021-haberi/>

<http://www.istanbulburda.com/Birinci-Izник-Konsili'nin-yeri,-uluslararası-calistay-ile-belirlenecek-BURSA-31243>

http://yenicehaber.com/SnUoO/SnUoO/Guzel_Sanatlardan_uluslararası_calistay-content-n-17-330.html

http://www.akampus.com/kampus_online/haber_detay.asp?haber=Dokuz-Eylul-Universitesi---Uluslararası-Arkeolojik-Calistay:-Anadolu-da-Gec-Antik-Donem-Cam-Sanati--M.S.4.--8.-Yuzyillar-&news=1639

http://www.yapitr.com/Etkinlikler/sinasosta-uluslararası-calistay_46940.html

<http://www.orhaz.com.tr/iuc/?p=165>

http://www.sessiztv.net/index.php?option=com_content&view=article&id=204:ii-ulusal-fiziksel-aktivite-caltay&catid=1:son-haberler&Itemid=50

<http://haberguncel.blogspot.com/2011/06/ulusal-iradenin-ustunlugu-mu-cuneyt.html>

<http://www.karaelmashaber.com/yazar/1159-turgut-guvendurustce-kilimliye-ulusal-bakis.html>

<http://www.hasankeyf.gov.tr/haber/haberdetay.asp?ID=188>

http://www.anayol.org/haber_detay.asp?id=2156&uyeid=159

http://www.anafor.org/haber_detay.asp?bolum=3618&uyeid=47

<http://www.bugun.com.tr/kose-yazisi/118142-ulusal-istihdam-stratejisi-ne-zaman-makalesi.aspx>

<http://www.internethaber.com/sirnakta-ulusal-basin-teroru-6455y.htm>

<http://www.capital.com.tr/capital-yilin-ulusal-dergisi-secildi-haberler/14810.aspx>

<http://www.haberci.pro/ulusal-gurur-inonu-hasan-camoglu-yaziyor.html>

http://www.turkcebilgi.com/kose_yazisi_88715_cengiz-aktar-siradaki-ulusal-politika-organik-tarim.html

<http://yenisafak.com.tr/yazarlar/?i=11958&y=ResulTosun>

<http://www.sesverturkiye.com/ulusal-kanallarimiz-ne-kadar-ulusal-9822.html>

<http://belgeselokulu.tr.gg/T.ue.rkiye%26%238217%3Bde-ulusal-gazetelerin-habercilik-etik%26%238217%3Bincelemenmesi.htm>

<http://www.internethaber.com/ulusal-cikarlar-boyle-de-korunur-6055y.htm>

http://www.haberiniz.com/yazilar/koseyazisi23716-Ulusal_Guc_Birligi.html

http://www.kentlob.net/index.php?option=com_content&view=article&id=964:eko-kent-tasarm-ulusal-proje-fikir-yarmas-sonucland&catid=113:yarma-sonuclar&Itemid=311

http://www.hiniskultur.com/haber_detay.asp?haberID=2830

<http://tokathaber.com.tr/haberdetay.asp?ID=10409>

<http://haber.xn--it-yka.net/2011/02/11/ulusal-makina-muhendisligi-ogrenci-zirvesi/>

<http://www.hakanyilmazcebi.com/?sayfa=oku&id=95>

<http://www.haberpan.com/makale/melih-arat/basbakana-ve-milli-egitim-bakanina-acik-mektup-2>

<http://www.ajans04.net/43228-Baskan-Ozgener-spor-m-agrihaberleri.html>

<http://www.tumkoseyazilari.com/yazar/arslan-bulut/28-06-2011-rejimi-yikmak-icin-bazi-yargi-mensuplarini-kullaniyorlar.html>

<http://www.ilk-kursun.com/2010/08/banu-avar-yazdi-ulusal-savunma-turk-ordusu-ve-kukla-tiyatrosu/>

<http://www.tepav.org.tr/tr/kose-yazisi/s/441>

http://www.turkcebilgi.com/kose_yazisi_122204_ozdemir-ince-mersin-uluslararası-muzik-festivali.html

http://www.turkcebilgi.com/kose_yazisi_89956_resul-tosun-uluslararası-camia.html

<http://www.taraf.com.tr/murat-belge/makale-uluslararası-jandarma-miz.htm>

<http://www.haberolcek.com/makale/543/uluslar-arasi-ceza-mahkemesi-hikmet-aksu.aspx>

<http://www.taraf.com.tr/yasemin-congar/makale-eve-donus-icin-uluslararası-plan.htm>

<http://yenisafak.com.tr/Yorum/?i=266673>

<http://ceviribilim.com/?p=4159>

http://www.ipekyolhaber.com/Kose_Yazisi.php?column_id=65

<http://haber.xn--it-yka.net/2011/05/03/itu-yelken-takimi-uluslararası-yarista-3-oldu/>

<http://www.erzurumgazetesi.com.tr/default.asp?page=haber&id=54538>

<http://www.gazeteyarin.com/y938-ted-uluslararası-2egitim-forumu.html>

<http://www.perizade.com/17-uluslararasi-istanbul-caz-festivali/comment-page-1>

<http://www.motoaktuel.com/blog/index.php/marmok-1-uluslararasi-gourmet-riders-festivali>

<http://tr.clearharmony.net/articles/200911/6729.html>

<http://www.yg.yenicaggazetesi.com.tr/yazargoster.php?haber=18269>

<http://www.almanyabulteni.de/lokal-haberler/10-uluslararasi-essen-aturk-kosusu>

<http://www.turk.internet.com/portal/yazigoster.php?yaziid=18426>

<http://www.t24.com.tr/haberdetay/85087.aspx>

<http://www.binifikir.be/news/155/ARTICLE/4157/2009-03-02.html>

<http://yenisafak.com.tr/Saglik/?i=320836>

<http://www.istanbulajansi.com/2/1/11/1782/Uluslararasi-istanbul-Opera-Festivali-.html>

<http://www.skyturk.net/haber/uluslararasi-ceza-mahkemesi--kaddafi-hakkinda-yakalama-karari-cikartti-dunya-5051.html>

<http://www.turklider.org/TR/EditModule.aspx?TabId=3748&mid=32349&ItemId=16568>

<http://www.bugun.com.tr/kose-yazisi/146067-uluslararasi-sistem-ve-ortadogu-makalesi.aspx>

http://www.kircaalihaber.com/?pid=8&id_aktualno=242

<http://www.timeturk.com/tr/2011/06/15/sam-da-ulusal-konferans-cagrisi.html>

<http://www.yuksekovahaber.com/haber/ulusal-konferans-yapilacak-52164.htm>

http://www.hinisaktuel.com/haber_detay.asp?haberID=302

<http://www.kurdistan-aktuel.org/yazarlar/nihat-buldan/1458-ulusal-konferans.html>

<http://www.odtugvo.k12.tr/Ankara/projeler/ICA/ICA%20WEB%20Sitesi/ULUSAL%20KONFERANS.html>

http://www.haberdiyarbakir.com/news_detail.php?id=43119

http://www.ozgur-gundem.com/?haberID=12488&haberBaslik=Amed'de%20Ulusl%20Konferans'%20heyecan%C4%B1&action=haber_detay&module=nuce

<http://bianet.org/bianet/siyaset/124651-talabaniden-kurt-ulusal-konferansi-fikrine-destek>

<http://iaeste.marmara.edu.tr/?p=27>

<http://www.boluweb.com/haber.asp?id=7172>

http://cnnkurd.org/index.php?option=com_content&view=article&id=289

<http://www.haberpan.com/haber/sam-da-ulusal-konferans-cagrisi/>

<http://www.salamnews.org/tr/news/read/36915/iranda-kuran-ccedilevirmenlerinin-ilk-ulusal-konferansi-yapilacak/>

<http://www.haberler.com/ulusal-turizmde-inovasyon-konferansi-2819609-haberi/>

http://www.yuksekovaguncel.com/haber_detay.php?id=21312

http://www.librenews.eu/index.php?style=news&cat_id=3&news_id=23077&sendnews=form

http://career.sabanciuniv.edu/news_detail/36368

http://www.sendika.org/yazi.php?yazi_no=23350

http://konferans.ipv6.net.tr/sunumlar/konferans_sonuc_bildirgesi.pdf

<http://www.lpghaber.com/Erzurum%E2%80%99da--Girisimcilik-Ulusal-Konferansi--haberi-321595.html>

<http://uymk2008.ege.edu.tr/>

<http://www.uyma.net/egitim/yuksekgretim-yeterlilikler-cercevesi-ulusal-konferansi/>

http://www.kizilbayrak.net/index.php?id=178&print=1&no_cache=0&tx_ttnews%5Btt_news%5D=40160

<http://www.sakaryaegitim.somee.com/haberdetay.asp?ID=691>

<http://www.haber3.com/gurcistan-ulusal-bilim-akademisi-ogretim-uyeleri-konferans-verdi-887486h.htm>

http://www.bilgesam.org/tr/index.php?option=com_content&view=article&id=1052:uluslararası-konferans-cada-tuerkiye-calmalar-aktueel-problemler&catid=141:konferanslar&Itemid=227

<http://www.mobilyadergisi.com.tr/default.asp?page=altsayfalar&tur=155&ad2=AH%DEAPLA%20YA%DEAMAK%20%20ULUSLARARASI%20KONGRE>

<http://www.businessankara.com/kongreler/netcadden-uluslararası-konferans-7026.html>

<http://icc.tobb.org.tr/haberduyuru/Uluslararası%C4%B1%20ICC%20Ticaret%20Kurallar%C4%B1%20Konferans%C4%B1%2010-11%20Kas%C4%B1m%202010.pdf>

<http://ihm.politics.ankara.edu.tr/konferans/>

<http://www.ajans5.com/detay/2010/07/21/uluslararası-konferans-konusunda-rum-siyasi-partiler-arasında-anlaşmazlık-var.html>

<http://www.1news.com.tr/guneykafkasya/rusya/20100927043403641.html>

http://www.cdgbim.com/index.php?option=com_content&view=article&id=3119:uluslararası-konferans&catid=1:son-haberler&Itemid=50

<http://www.akintiya-karsi.org/koxuz/node/5410>

http://bayposta.com/wiki/Uluslararası%C4%B1_Platformda_Patent_Koruması%C4%B1'_Konulu_Uluslararası%C4%B1_Konferans_Tamamlandı%C4%B1

<http://www.kongre.net/haberler.php?haber=102>

<http://www.gau.edu.tr/tr/940/haberler/doguakdeniz-ve-kibris-konulu-uluslararası-konferansi-bilim-ve-diplomasi-cevrelerinden-yogun-ilgi-gordu>

[http://www.haberx.com/yeniden_insa_icin_uluslararası_konferans\(17,n,10111999,146\).aspx](http://www.haberx.com/yeniden_insa_icin_uluslararası_konferans(17,n,10111999,146).aspx)

<http://www.inovasyondunyasi.com/icerikap/1634/inovasyon-ve-sinai-mulkiyet-haklari-uluslararası-konferansi-.htm>

<http://www.islamidavet.com/2011/06/24/terorizmle-mucadele-uluslararası-konferansi-basliyor/>

<http://www.haberortak.com/Haber/Seminerler/09022009/YASED-8217den-Uluslararası-Konferans.php>

<http://www.egitimduyurulari.com/ho.asp?id=4384>

<http://www.bthaber.com.tr/?p=2430&sayi=SAYI:760>

http://www.mugladevrin.com.tr/index.php?option=com_content&view=article&id=1972:akdeniz-ky-vakf-uluslararası-konferans-yapıldı-

<http://www.yorumla.net/yurtdisi-haberler/798387-haiti-icin-uluslararası-konferans-toplanacak.html>

http://www.ttb.org.tr/mevzuat/index.php?option=com_content&task=view&id=521&Itemid=36

<http://www.bolgeninsesi.com/son-dakka/1-haber-1/2407-karabuek-ueniversitesi-uluslararası-konferansa-ev-sahipli-yapıyor.html>

http://www.ias.com.tr/enterprise/news/20050704_ITU.html

<http://www.ilericigenclik.org/haberler/birlik-dayanisma-hareketi-uluslararası-konferans-duzenliyor>

<http://www.haberler.com/dau-kktc-de-teknoloji-konulu-uluslararası-haberi/>

<http://www.turkuro2010.org/>

<http://www.kimya2011.com/dosyalar/davetiye.pdf>

http://www.mmo.org.tr/genel/bizden_detay.php?kod=19080&tipi=10216&sube=0

<http://biyokongre18.marmara.edu.tr/>

<http://www.iscisagligikongresi.org/>

<http://www.cmokongre2011.org/kongre-cagrisi.asp>

http://www.kasder.org.tr/index.php?option=com_content&view=article&id=263%3A46-ulusal-noeroloji-kongresi-gercekletirildi&catid=25%3Aulusal-kongre-ve-sempozyumlar&Itemid=75&lang=tr

http://www.kasder.org.tr/index.php?option=com_content&view=article&id=261%3A45-ulusal-noeroloji-kongresi&catid=25%3Aulusal-kongre-ve-sempozyumlar&Itemid=75&lang=tr

<http://www.ulusalturizmkongresi.com/HomeDirectoryDefault.UlusalTurizmKongresi>

<http://www.usimp.org/node/71874>

<http://www.todnet.org/kongre2011/>

<http://www.tkd-online.org/egitim/2005/20051126%20-%20Antalya.pdf>

<http://www.tumkongreler.com/kongre/cinsellik-ve-cinsel-tedaviler-ulusal-kongresi>

<http://www.ftr2011.org/>

<http://www.konferanslar.net/app/viewConfDetail.do?id=63>

<http://www.biyosidal.org/Duyuru/15/iiulusal-biyosidal-kongresi.html>

<http://www.teknoport.com.tr/2009/11/12/universite-sanayi-isbirligi-ulusal-kongresi-2010/>

<http://www.kongreara.com/zemin-mekanigi-ve-temel-muhendisligi-13-ulusal-kongresi.html>

<http://www.tumkongreler.com/kongre/vaskuler-anomaliler-dernegi-birinci-ulusal-kongresi>

http://kongretr.com/2008/06/09/trkiye_stn_yetenekli_zocuklar_2_ulusal_kongresi_2527_mart_2009.html

<http://www.das.org.tr/2009/>

<http://www.ulusalsosyalhizmetlerkongresi.org/>

<http://www.denizhaber.com/HABER/22261/1/kiyi-ve-deniz-alanlari-kongre-trabzon-ktu.html>

http://www.yapi.com.tr/Etkinlikler/dinamikler-2006-proje-yonetimi-ulusal-kongresi-7-8-nisanda-itu-ayazaga-kampusunde_39671.html

<http://www.psikofarmakoloji2011.org/>
<http://www.ipa.org.tr/TR/IcerikDetay.aspx?ID=40>
<http://www.kokhucrekongresi2011.org/>
<http://www.tumkongreler.com/kongre/36-ulusal-fizyoloji-kongresi-uluslararası-katılımlı>
<http://inccui.anadolu.edu.tr/tr/default.asp>
<http://www.denetimkongre.org/trfiles/default.asp>
<http://www.yemsanayi.com/haber/vi.-ulusal-hayvan-besleme-kongresi-uluslararası-katılımlı,29.06.2011-45.htm>
<http://www.marslogistics.com/logilife/konu.aspx?konu=20>
http://www.erdalzorba.com/haber_detay.php?haber_id=82
<http://www.medimagazin.com.tr/hekim/kongre/tr-uluslararası-yuksekogretim-kongresi-yapılacak-2-20-35291.html>
<http://iubasin.istanbul.edu.tr/duyuru/uluslararası-anayasa-kongresi-h257.htm>
<http://xn--trkoloji-65a.gen.tr/>
<http://www.tumkongreler.com/kongre/uluslararası-katılımlı-1-ulusal-gelisimsel-pediyatri-kongresi>
http://www.metalurji.org.tr/index.php?option=com_content&view=article&id=138:13-uluslararası-metalurji-ve-malzeme-kongres-yapıldı&catid=35:duyurular&Itemid=59
<http://www.tasamasya.org/component/content/article/102-haberler/1145-5-uluslararası-tuerk-asya-kongres-bakanin-sonuc-deklarasyonu.html>
<http://www.haber34.com/uluslararası-cocuk-ve-iletisim-kongresi-1685-haberi.html>
<http://www.canakkaleicinde.com/vii-uluslararası-sivil-toplum-kuruluslari-kongresi-hayirseverlik.html>
<http://v3.arkitera.com/e3710-23-uluslararası-yapi-ve-yasam-kongresi.html>
<http://www.abvizyonu.com/ab/stefan-fule-uluslararası-avrupa-hareketi-kongresi-icin-istanbula-geldi.html>
<http://www.anittahotel.com/son-haberler/7-hititoloji-kongresi>
<http://www.todturkey.org/lang-tr/haberler/326-iv-uluslararası-tuerk-afrika-kongres.html>
<http://www.haberler.com/3-uluslararası-ankara-urojinekoloji-kongresi-2517146-haberi/>
<http://www.turkhukusitesi.com/showthread.php?t=59139>
<http://www.girisimhaber.com/post/2011/04/11/Kobiler-Icin-II-Uluslararası-Girisimcilik-Kongresi-Izmirde.aspx>
<http://www.elektrikhaber.net/haber/642-fuar-ve-kongre-uluslararası-enerji-kongresi.html>
<http://www.balkanskidom.com/showthread.php?p=23793>
<http://www.akcayhaber.web.tr/guncel/1207-Ulusal-Parti-Ali-Emre-Ozsoy-Panelli-Duzenedi.html>
<http://www.serander.net/forum/index.php?topic=2400.0>
<http://yenimedya.wordpress.com/2010/08/19/27-ulusal-bilisim-kurultayiankara-tbd-panel-23-eylul-2010-%E2%80%9C9C-yeni-medyada-nefret-soylemi%E2%80%9D-saat-1045-1200/>
<http://www.bornovagazetesi.com/haberdetay.aspx?hid=655>

http://www.ulusalmuhasebecilerbirligi.com/duyurular/haber_oku.php?haber_id=22

<http://www.mufetder.org.tr/panel2.html>

<http://translate.google.com.tr/translate?hl=tr&langpair=en|tr&u=http://www.sportcentric.com/vsite/vnavsite/page/directory/0,10853,5165-189213-206435-nav-list,00.html>

<http://translate.google.com.tr/translate?hl=tr&langpair=en|tr&u=http://nasbla.org/i4a/pages/index.cfm%3Fpageid%3D4225>

<http://translate.google.com.tr/translate?hl=tr&langpair=en|tr&u=http://www.ncppanel.com/>

<http://www.kizilbayrak.net/rss/arsiv/2010/11/29/artikel/138/istanbulda-ek.html>

<http://www.anayurtgazetesi.com/default.asp?page=haber&id=330394>

[http://www.elestiriyoruz.com/guncel-haberler/basortusu-ve-laiklik-paneli-ulusal-strateji-merkezi-\(usmer\)-chp/?wap2](http://www.elestiriyoruz.com/guncel-haberler/basortusu-ve-laiklik-paneli-ulusal-strateji-merkezi-(usmer)-chp/?wap2)

<http://www.mesop.net/osd/?app=izctrl&archiv=153&izseq=izartikel&artid=608>

<http://www.haberfark.net/detay.asp?id=10333>

<http://www.iscturkey.org/2010/2008/index.php?id=pan>

<http://europistforum2005.blogspot.com/2007/11/panel-ulusal-kltr-ajanslar-stklar-kltr.html>

<http://www.logisticusdergisi.com/haberdetay/8113-DuZCE-KALITE-PANELI.html>

<http://www.saskarahaber.com/v1/haber/712-ardahan-kaisiadamp8217-dan-toplumun-yeri-ve-onemi-paneli.html>

http://www.siemens.com.tr/web/1508-13824-1-1/siemens_turkiye_-_tr/siemens_akademi/haberler/panel_koo_projelerinde_radyoloji_departmanlarinin_isletilmesi

<http://www.haberler.com/kau-de-egitimimizin-ulusal-boyutu-konulu-panel-haberi/>

http://www.kastamonu.edu.tr/Duyurular,Genel_Duyurular_Ayrintili.html?NewsID=882

http://www.kalder.org/index.php?option=com_content&view=article&id=371&Itemid=181&lang=tr

http://www.alanyagc.org/index.php?option=com_content&view=article&id=1346:-qnasl-bir-medya-nasl-bir-duenyaq-paneli&catid=1:haberler&Itemid=546

<http://www.karshaberleri.com/haber/18471/kasiaddan-sivil-toplum-paneli.aspx>

<http://www.tarikakay.com/HaberDetay.aspx?Id=119>

http://www.uzeg.org/index.php?option=com_content&view=article&id=96:uaktan-dil-eitimi-gercekletirildi&catid=42:dernek-haberleri&Itemid=73

<http://www.botso.org.tr/index.asp?rty65=xp5&id=33>

<http://www.haberlink.com/haber.php?query=4538>

<http://www.hukukmerkezi.org/tr/haberler/hmden-haberler/141-uluslararası-panel.html>

<http://www.kampushaber.org/karadeniz-teknik-universitesi/panel-uluslararası-deniz-hukukunda-kiyi-devletinin-gemilere-el-koyma-yetkisinin-sinirlari-31308.html>

<http://www.logisticusdergisi.com/haberdetay/8113-DuZCE-KALITE-PANELI.html>

http://www.bizimelma.org/index.php?/programlar_ve_dersler/mac_os_sistem_pufleri.../leopard_uluslararası_panel_sorunu.html

http://www.habervakti.com/news_details?id=25640

<http://www.theunity.org/files/Uluslararası-Panel-Basında%20İslam-Algisi.pdf>

<http://www.ankaraevents.com/panel-kadinin-siyasete-katilimi-yerel-ve-uluslararası-perspektifler-türk-amerikan-derneği-30-mart-2011/>

http://www.idsb.org/tr/index.php?option=com_content&view=article&id=452:uluslararası-panel-basında-slam-algs&catid=25:basından&Itemid=13

http://www.yalovarehberi.com/haber_detay.asp?haberID=679

<http://www.tumgazeteler.com/?a=4184368>

<http://www.sonpeygamber.info/uluslararası-panel-hz-peygamber-savi-dunyaya-nasil-anlatmalıyız->

<http://www.akvaryumforum.com/forum/balıkca-haberler-80/5888-iklim-değişikliği-ile-ilgili-uluslararası-panel-hakkında-bilgi.htm>

<http://www.uzeg.org.tr/>

<http://www.uzeg.org.tr/>

<http://www.amnesty.org.tr/ai/node/720>

<http://www.pcteknik.net/saglik-haberleri/62332-5-uluslararası-losemili-cocuklar-haftası-programı.html>

<http://www.canakkalecinde.com/6-kucukkuyu-uluslararası-kultur-ve-sanat-festivali.html>

http://www.izmimod.org.tr/index.php?sayfa=2000_12_04&bolum=bultenler&kat=dev

http://insanhaklarimerkezi.bilgi.edu.tr/pages/Activity_full.asp?id=132&r=07.06.2011+22%3A40%3A00

<http://www.rumeliplatformu.com/?p=2703>

<http://www.sondakika.com/haber-fotografarla-kktc-nin-dunu-ve-bugunu-konulu-2808167/>

<http://www.the.org.tr/2011/04/14/dogu-turkistan-paneline-davet/>

<http://www.turcev.org.tr/content.php?conID=64>

http://www.yesilevlerk.com/index.php?option=com_content&view=article&id=18%3Aulusal-ajans-semineri&catid=13%3Adernek-haberler&lang=tr

<http://www.basketligi.com/ulusal-antrenor-gelisim-semineri-sona-erdi/>

http://212.175.131.61/www.isggm.gov.tr/htdocs/article.php?article_id=337

http://www.elbistaninsesi.com/news_detail.php?id=12890

<http://www.facebook.com/event.php?eid=105834909437227>

<http://www.tsf.org.tr/kurullar/merkez-hakem-kurulu/2327-ulusal-hakem-kursu-ve-gelisim-semineri-konyada>

<http://www.resmi-gazete.org/gundem/duyurular/tarim-bakanligi/saglik-ve-bitki-sagligi-onlemleri-ulusal-semineri.html>

http://www.ariokullari.k12.tr/my_documents/haber/8ulusaleko.htm

http://www.hakkariram.com/index.php?option=com_content&view=article&id=110

<http://www.tvf.org.tr/index.php?haber=2851>

http://www.kalder.org/index.php?option=com_content&view=article&id=124&Itemid=54

<http://kalev.test.enroll.com.tr/MenuContent.aspx?id=90>

<http://www.istikbalgazetesi.com/?sec=1&newscatid=7&newsid=55582>

http://www.herkesetiyatro.com/tiyatro_haberi_09_194.html

<http://www.erzincan.bel.tr/index.php?DID=264&Sayfa=20>

http://www.ahsap.org/news_014.html

<http://www.perasanat.com.tr/sofya/pedegoji.html>

<http://www.haberler.com/ulusal-afet-yonetimi-egitim-semineri-yalova-da-haberi/>

<http://www.ulakbim.gov.tr/cabim/yayin/toplanti/>

<http://www.elektrotekno.com/about20282.html>

<http://www.asiyankoleji.com/7/1/haber-detay/eko--okullar-ulusal-koordinatör-ogretmenler-semineri/?mNewsDetail=97>

<http://www.besyo.net/index.php?topic=762.0>

<http://prizma.dogus.edu.tr/?p=301>

<http://www.haberler.com/kutahya-da-1-ulusal-is-semineri-2406458-haberi/>

http://myo.karatekin.edu.tr/index.php?option=com_content&view=article&id=280:uluslararası-seminer-daveti-12-16-mart-2012-helsinki-finlandiya&catid=49:ana-sayfa-duyurular&Itemid=227

<http://www.sondakika.com/haber-merkez-bankasi-ndan-uluslararası-seminer-2612351/>

<http://www.tusiad.org.tr/bilgi-merkezi/raporlar/ab-muktesebati-uluslararası-semineri--is-sagligi-ve-guvenligi-ve-cevre/>

<http://www.ab-ilan.com/ab-uyum-sureci/hakim-ve-savcilar-icin-uluslararası-seminer-001247.html>

<http://www.sigortacigazetesi.com.tr/guncel/118-guncel/1675-tsevden-uluslararası-seminer.html>

<http://siviltoplumhaber.com/ariv/652-romanlara-uluslararası-seminer.html>

<http://www.istanbulburda.com/Uluslararası-SEMINER-16875>

<http://turkish.cri.cn/1101/2011/05/09/162s133068.htm>

<http://www.xing.com/net/stratejikyonyap/fikirler-oneriler-geribildirim-456923/turkiyede-teknolojik-inovasyon-ve-global-ekosistem-uluslararası%C4%B1-semineri-27242643/>

<http://www.hanemiz.com/genel-mimarlik/146681-grenoble-mimarlik-okulu8217nda-8220istanbul8221-konulu-uluslararası-seminer.html>

<http://www.beykoz.edu.tr/tr/iru-academy-uluslararası-semineri-yueksekokulumuz-evsahipliginde-gerceklestirildi>

http://www.kktceeb.com/includes/uploaded/28042010_89.pdf

http://www.sagliklitavuk.org/announces.php/announces_id/62

<http://www.yaraticidrama.org/content/view/174/1/>

http://myo.karatekin.edu.tr/index.php?option=com_content&view=article&id=217:uluslararası-gelecein-leri-konulu-seminere-davet-11-8-mart-2011-helsinki-finlandiya&catid=49:ana-sayfa-duyurular&Itemid=227

<http://www.everybodys-song.net/turkish/seminars.php>

http://www.gazetea24.com/haber/mb_den_uluslararası_seminer.htm

<http://www.sigortacigazetesi.com.tr/guncel/118-guncel/1184-milli-reden-uluslararası-seminer.html>

<http://www.haberalanya.com.tr/root.vol?title=alanya-39da-uluslararası-seminer&exec=page&nid=93414>

<http://www.canakkalecinde.com/gencler-yerel-yonetimde-uluslararası-semineri.html>

<http://www.tff.org/default.aspx?pageID=228&ftxtID=11340>

<http://www.okulonceiyiz.biz/seminer-egitim-ve-konferans-duyurulari/25798-uluslararası-kocluk-sertifika-certified-coach-programi.html>

<http://cyprusyeshler.blogspot.com/2010/11/uluslararası-seminer.html>

<http://www.camlaralti.k12.tr/iv-uluslararası-paylasim-seminerindeki-ortak-paylasimlar-h63.html>

<http://www.fenerbahcegazetesi.com/haberler/haber/201-uluslararası-teknik-direktor-semineri-basladi>

<http://www.facebook.com/topic.php?uid=33150459683&topic=8508>

<http://www.uteg10.org/tr/default.asp?rsm=11150000000>

<http://kirsalturizm.gen.tr/>

<http://16susemp2011.org/>

http://www.emo.org.tr/genel/bizden_detay.php?kod=70047&tipi=10108&sube=6

<http://www.cevremuhendisleri.net/f97/1-ulusal-aritma-camurlari-sempozyumu-acs-2005-a-2135/>

<http://www.medimagazin.com.tr/medimagazin/tr-2-ulusal-kadin-kalp-sagligi-sempozyumu-ankarada-yapildi-676-525-10005.html>

<http://www.ins.itu.edu.tr/muholc/AmacKonu.htm>

http://www.boradair.com/haber/haber_detay.asp?haberID=804

<http://gencadd.org/etkinlikler/genel-merkez/100-gmerkez/227-ulusal-yonetim-semp>

<http://www.gurun.gov.tr/haberdetay.asp?ID=223>

<http://v3.arkitera.com/news.php?action=displayNewsItem&ID=41841>

http://www.tarimmerkezi.com/yazar_kose.php?hid=113

<http://www.siirtmanset.com/universite/ulusal-el-sanatları-sempozyumu-siirtte-basliyor.htm>

<http://matematikcafe.net/k-xxii-ulusal-matematik-sempozyumu.html>

<http://www.cevremuhendisleri.net/f11/ulusal-su-muhendisligi-sempozyumlari-1975/>

<http://www.resmi-gazete.org/gundem/duyurular/devlet-su-isleri/2--ulusal-taskin-sempozyumu-tamamlandi.html>

<http://www.gazetegercek.com/1-ulusal-ilgin-sempozyumu-hazirliklari-tamamlandi.html>

<http://www.okuloncesiforum.com/konferans-ve-etkinlik-duyurulari/73933-ii-ulusal-ogretmenim-sempozyumu-bildiri-basvurusu-daveti.html>

<http://www.operaturkiye.com/wp1/index.php/haberler/ulkemizden/8-ulusal-muzik-sempozyumu-basladi.html/>

<http://www.asisempozyumu.org/>
<http://evrimsempozyumu.org/node/72>
<http://www.speleolojisempozyumu.org/node/27>
<http://www.yerelyonetimler.sakarya.edu.tr/>
<http://www.haksay.org/?q=node/59>
<http://www.ircica.org/islam-medeniyetinde-bagdat-mednetus-selm-uluslararası-sempozyum-tebligleri/irc864.aspx>
http://www.universitehaber.com/article.php?article_id=13919
<http://eyi.pau.edu.tr/>
<http://www.istanbulkutuphaneci.org/node/746>
<http://www.tedaproject.com/TR/belge/1-93673/cevirmenleri-ve-yayincilariyla-turk-edebiyati-iii-ulusl-.html>
<http://www.vanasyanur.net/haber/Bediuzzaman-icin-uluslararası-sempozyum/7314>
<http://www.hossohbet.com/forum/kadin-ve-hukuk/139693-uluslararası-sempozyum-kadin-belleğini-olusturmada-kaynak-sorunu.html>
<http://www.dgtyb.org/news/1206.html>
<http://www.motivakfi.com/default.asp?rsm=270000000000&sayfa=&id=47>
<http://www.memurlar.biz/haber/20110408174643/Osmangaziden-uluslararası-sempozyum.html>
<http://www.turkceogretimi.com/sempozyumlar-konferanslar/iv-uluslararası-dunya-dili-turkce-sempozyumu-mugla-universitesi>
<http://yenisafak.com.tr/Gundem/?i=280367>
<http://www.kadinlarkulubu.com/portal/etiketler/uluslararası-sempozyum>
<http://www.ntvmsnbc.com/id/24970579/>
<http://www.haberhavadis.com/ic-haberler/ydu-ile-turksoy-uluslararası-sempozyum-duzenliyor.htm>
<http://www.gazetegercek.com/gediz-universitesinden-uluslararası-sempozyum-ilklerine-bir-yenisi-eklendi.html>
<http://www.dika.org.tr/?g1=haberler&g2=414&g3=veysel-karaniye-uluslararası-anma>
<http://www.lpghaber.com/Kutahya%60da-Uluslararası-Sempozyum-Ve-Kongre--haberi-526447.html>
<http://www.kenthaber.com/ic-anadolu-nevsehir/Haber/kultur-sanat/Normal/uluslararası-sempozyum/0658a583-b120-420d-ba82-f2f471ed4297>
<http://www.cayfan.net/community/202059-uluslararası-yunus-emre-sempozyumu-sona-erdi.html>
<http://www.nethabercilik.com/haber/yeni-lisan-hareketi-uluslararası-sempozyumu-kocaeli-basladi.htm>
<http://egeweb2.ege.edu.tr/gd/halkilis/bergamasemp/davet.htm>
<http://www.pcteknik.net/edebiyat/58667-klasik-turk-edebiyati-icin-uluslararası-sempozyum.html>
<http://www.ahmetakgunduz.com/news/news-turkish/78-mehmet-akif-ersoy-uluslararası-sempozyumu.html>

EK F – En Yakın k Komşu Algoritması Kullanım Kodları

```

public double IbkDegerlendir()
{
    double sonuckod = 0;
    try
    {
        // Egitim Basladi.
        // insts içinde %100 tüm eğitim satırları var
        weka.core.Instances insts = new weka.core.Instances(new
            java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
                Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\Egitim.arff"));
        insts.setClassIndex(insts.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
        // Burada Sınıflama Algoritması seçiliyor. Tree seçmek en etkin olabilir, daha iyi sınıflar.
        weka.classifiers.Classifier cl = new weka.classifiers.lazy.IBk();
        // burada öğrenirken işlenmesi gereken eğitim satırlarını rasgele sırala yapıldı.
        weka.filters.Filter myRandom = new weka.filters.unsupervised.instance.Randomize();
        myRandom.setInputFormat(insts);
        insts = weka.filters.Filter.useFilter(insts, myRandom);
        // Rasgele Sırala bitti.
        weka.core.Instances train = new weka.core.Instances(insts, 0, insts.numInstances());
        cl.buildClassifier(train);
        // Egitim bitti.
        weka.core.Instances test = new weka.core.Instances(new
            java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
                Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\Test.arff"));
        test.setClassIndex(test.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
        weka.core.Instance currentInst = test.instance(0);
        double predictedClass = cl.classifyInstance(currentInst);
        sonuckod = predictedClass;
        // MessageBox.Show(predictedClass.ToString());
    }
    catch (java.lang.Exception ex)
    {
        ex.printStackTrace();
    }
    return sonuckod;
}

```

EK G – Karar Ağacı Algoritması Kullanım Kodu

```

public double TreeDegerlendir()
{
    double sonuckod = 0;
    try
    {
        // Egitim Basladi.
        // insts içinde %100 tüm eğitim satırları var
        weka.core.Instances insts = new weka.core.Instances(new
            java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
                Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\Egitim.arff"));
        insts.setClassIndex(insts.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
        // Burada Sınıflama Algoritması seçiliyor. Tree seçmek en etkin olabilir, daha iyi sınıflar.
        weka.classifiers.Classifier cl = new weka.classifiers.trees.J48();
        // burada öğrenirken işlenmesi gereken eğitim satırlarını rasgele sırala yapıldı.
        weka.filters.Filter myRandom = new weka.filters.unsupervised.instance.Randomize();
        myRandom.setInputFormat(insts);
        insts = weka.filters.Filter.useFilter(insts, myRandom);
        // Rasgele Sırala bitti.
        weka.core.Instances train = new weka.core.Instances(insts, 0, insts.numInstances());
        cl.buildClassifier(train);
        // Egitim bitti.
        //string Metin = "seminer ve konferans var konferans konferans toplanti sempozyum";
        Kumele.KumeClass MetinKumele = new Kumele.KumeClass();
        MetinKumele.AttributeAdeti = 17;
        MetinKumele.AttributeAta();
        MetinKumele.KelimeAdetHesapla(MetinTxt.Text.ToString());
        MetinKumele.EgitmenSatiriOzellikYazdir();
        weka.core.Instances test = new weka.core.Instances(new
            java.io.FileReader("D:\\CR_Doküman\\Yüksek Lisans\\Trakya
                Üniversitesi\\Tez ve Seminer\\Tez\\CANDATA\\Test.arff"));
        test.setClassIndex(test.numAttributes() - 1); // son alanın tür alan bilgisi olduğunu set ediyor.
        weka.core.Instance currentInst = test.instance(0);
        double predictedClass = cl.classifyInstance(currentInst);
        sonuckod = predictedClass;
        // MessageBox.Show(predictedClass.ToString());
    }
    catch (java.lang.Exception ex)
    {
        ex.printStackTrace();
        MessageBox.Show("Hata");
    }
    return sonuckod;
}

```

EK H – Arama Sayfası JavaScript Kontrolü

```
<scriptlanguage="javascript" type="text/javascript">
function kontrol() {
var a, b, c, d, e, f;
    a = document.getElementById("chckSeminer").value;
    b = document.getElementById("chckKongre").value;
    c = document.getElementById("chckKonferans").value;
    d = document.getElementById("chckSempozyum").value;
    e = document.getElementById("chckPanel").value;
    f = document.getElementById("chckCalistay").value;
if ((a == "on") || (b == "on") || (c == "on") || (d == "on") || (e == "on") || (f == "on")) {
returntrue;
    }
else {
returnfalse;
    }
}
</script>
```

EK I – Kullanıcı Giriş Sayfası JavaScript Kontrolü

```
<scriptlanguage="javascript" type="text/javascript">
function kontrol() {
var istenmeyen = "*|,|\"':<>[]{}`\';()&${}%+~_";
var sifre, kullanıcıAdı;
var i;
    sifre = document.getElementById("txtSifre").value;
    kullanıcıAdı = document.getElementById("txtKullaniciAdi").value;
if (kullanıcıAdı.length == 0 || sifre.length == 0) {
    alert("Alanlar boş geçilemez !");
returnfalse;
    }
for (i = 0; i < kullanıcıAdı.length; i++) {
if (istenmeyen.indexOf(kullanıcıAdı.charAt(i)) != -1) {
    alert("Kullanıcı adında istenmeyen karakter kullanımı !");
returnfalse;
    }
    }
for (i = 0; i < sifre.length; i++) {
if (istenmeyen.indexOf(sifre1.charAt(i)) != -1) {
    alert("Şifrede istenmeyen karakter kullanımı !");
returnfalse;
    }
    }
returntrue;
}
</script>
```

EK J – Kullanıcı Kayıt Sayfası Java Script Kontrolü

```

<scriptlanguage="javascript" type="text/javascript">
function kontrol() {
var emailText = /^[a-zA-Z0-9._-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,4}$/;
var istenmeyen = "*|,|\"<[]{}`\'();& $#%+ -_";
var txtMail;
var sifre1, sifre2, kullanıcıAdı;
var i;
    sifre1 = document.getElementById("txtSifre").value;
    sifre2 = document.getElementById("txtSifreTekrar").value;
    kullanıcıAdı = document.getElementById("txtKullaniciAdi").value;
    txtMail = document.getElementById("txtMail").value;
if (kullanıcıAdı.length == 0 || sifre1.length == 0 || txtMail.length == 0) {
    alert("Alanlar boş geçilemez !");
returnfalse;
    }
for (i = 0; i < kullanıcıAdı.length; i++) {
if (istenmeyen.indexOf(kullanıcıAdı.charAt(i)) != -1) {
    alert("Kullanıcı adında istenmeyen karakter kullanımı !");
returnfalse;
    }
}
for (i = 0; i < sifre1.length; i++) {
if (istenmeyen.indexOf(sifre1.charAt(i)) != -1) {
    alert("Şifrede istenmeyen karakter kullanımı !");
returnfalse;
    }
}
if (sifre1 != sifre2) {
    alert("Girilen şifre,tekrarı ile aynı olmalı !");
returnfalse;
    }
if (emailText.test(txtMail) == false) {
    alert("E-mail hatalı !");
    }
returntrue;
}
</script>

```

EK K – Kullanıcı Kayıt Sayfası Java Script Kontrolü

```

PublicClassKumeClass
StructureKelimeListesiYapi
Public Tur AsBoolean
Public Kelime AsString
Public Adet AsInteger
EndStructure
Private AttributeAdet AsInteger
PublicWriteOnlyProperty AttributeAdeti
Set(ByVal value)
    AttributeAdet = value
EndSet
EndProperty
Public KelimeListesi(100) AsKelimeListesiYapi
Public Attribute(8) AsString
PublicSub AttributeAta()
    KelimeListesi(1).Kelime = "seminer" : KelimeListesi(1).Tur = True
    KelimeListesi(2).Kelime = "sempozyum" : KelimeListesi(2).Tur = True
    KelimeListesi(3).Kelime = "konferans" : KelimeListesi(3).Tur = True
    KelimeListesi(4).Kelime = "kongre" : KelimeListesi(4).Tur = True
    KelimeListesi(5).Kelime = "etkinlik" : KelimeListesi(5).Tur = False
    KelimeListesi(6).Kelime = "panel" : KelimeListesi(6).Tur = True
    KelimeListesi(7).Kelime = "çalıřtay" : KelimeListesi(7).Tur = True
    KelimeListesi(8).Kelime = "diđer" : KelimeListesi(8).Tur = True
    KelimeListesi(9).Kelime = "katılım" : KelimeListesi(9).Tur = False
    KelimeListesi(10).Kelime = "toplantı" : KelimeListesi(10).Tur = False
    KelimeListesi(11).Kelime = "oturum" : KelimeListesi(11).Tur = False
    KelimeListesi(12).Kelime = "tarih" : KelimeListesi(12).Tur = False
    KelimeListesi(13).Kelime = "bilimsel" : KelimeListesi(13).Tur = False
    KelimeListesi(14).Kelime = "sunum" : KelimeListesi(14).Tur = False
    KelimeListesi(15).Kelime = "bildiri" : KelimeListesi(15).Tur = False
    KelimeListesi(16).Kelime = "ulusal" : KelimeListesi(16).Tur = False
    KelimeListesi(17).Kelime = "uluslararası" : KelimeListesi(17).Tur = False
    Attribute(0) = "Seminer"
    Attribute(1) = "Sempozyum"
    Attribute(2) = "Konferans"
    Attribute(3) = "Kongre"
    Attribute(4) = "Panel"
    Attribute(5) = "Çalıřtay"
    Attribute(6) = "Diđer"
EndSub
PublicFunction AttributeDondur(ByVal indis AsInteger) AsString
Return Attribute(indis)
EndFunction
PublicSub KelimeAdetHesapla(ByVal Metin AsString)
Dim kelime AsString
Dim i, uzunluk, sonraki, KelimeSayac AsInteger
Dim onceki AsInteger = 1
    uzunluk = Len(Metin)
For i = 1 To uzunluk
If Mid(Metin, i, 1) = " "Then
    sonraki = i
    kelime = Mid(Metin, onceki, (sonraki - onceki))
For KelimeSayac = 0 To AttributeAdet - 1

```



```

If VarmiKontrol(KelimeListesi(KelimeSayac).Kelime, kelime) Then
    KelimeListesi(KelimeSayac).Adet += 1
EndIf
Next KelimeSayac
    onceki = sonraki
EndIf
Next i
EndSub
PublicSub EgitmenSatiriOzellikYazdir()
Dim KelimeSayac, EnYuksekFrekansliKelimeIndisi AsInteger
Dim TestToplu As System.IO.StreamWriter
basla:
Try
    FileOpen(1, "D:\CR_Doküman\Yüksek Lisans\Trakya Üniversitesi\Tez ve
Seminer\Tez\CANDATA\Test.arff", OpenMode.Output)
    TestToplu = My.Computer.FileSystem.OpenTextFileWriter("D:\CR_Doküman\Yüksek
Lisans\Trakya Üniversitesi\Tez ve Seminer\Tez\CANDATA\TestToplu.txt", True)
    FileOpen(2, "D:\CR_Doküman\Yüksek Lisans\Trakya Üniversitesi\Tez ve
Seminer\Tez\CANDATA\TopluSonucTree.txt", OpenMode.Append)
    FileOpen(3, "D:\CR_Doküman\Yüksek Lisans\Trakya Üniversitesi\Tez ve
Seminer\Tez\CANDATA\topluSonucIBK.txt", OpenMode.Append)
'FileOpen(1, "C:\CANDATA\Test.arff", OpenMode.Output)
Catch
GoTo basla
EndTry
    Print(1, "@relation Bilimsel"& vbNewLine & vbNewLine)
For KelimeSayac = 0 To AttributeAdet - 1
    Print(1, "@attribute "& KelimeListesi(KelimeSayac + 1).Kelime & " {TRUE, FALSE}"&
vbNewLine)
Next
    Print(1, "@attribute tur {")
For KelimeSayac = 0 To AttributeAdet - 1
If KelimeListesi(KelimeSayac).Tur And KelimeSayac > 0 Then Print(1, ",")
If KelimeListesi(KelimeSayac).Tur Then Print(1, KelimeListesi(KelimeSayac).Kelime)
Next
    Print(1, "}")
    Print(1, vbNewLine & vbNewLine)
    Print(1, "@data"& vbNewLine)
For KelimeSayac = 1 To AttributeAdet
If KelimeListesi(KelimeSayac).Adet > 0 Then
If KelimeListesi(KelimeSayac).Adet > KelimeListesi(EnYuksekFrekansliKelimeIndisi).Adet And
KelimeListesi(KelimeSayac).Tur Then
    EnYuksekFrekansliKelimeIndisi = KelimeSayac
EndIf
    Print(1, "TRUE,")
    TestToplu.Write("TRUE,")
    Print(2, "TRUE,")
    Print(3, "TRUE,")
Else
    Print(1, "FALSE,")
    TestToplu.Write("FALSE,")
    Print(2, "FALSE,")
    Print(3, "FALSE,")
EndIf
Next
If EnYuksekFrekansliKelimeIndisi > 0 Then

```

```

        Print(1, KelimeListesi(EnYuksekFrekansliKelimeIndisi).Kelime)
        TestToplu.Write(KelimeListesi(EnYuksekFrekansliKelimeIndisi).Kelime)
Else
    Print(1, KelimeListesi(AttributeAdet - 9).Kelime)
    TestToplu.Write(KelimeListesi(AttributeAdet - 9).Kelime)
EndIf
    TestToplu.WriteLine()
    TestToplu.Close()
    FileClose(1)
    FileClose(2)
    FileClose(3)

EndSub
PublicFunction VarmiKontrol(ByVal Kelime1 AsString, ByVal Kelime2 AsString) AsBoolean
Dim DonenSonuc AsBoolean
    Kelime2 = LCase(Trim(Kelime2))
Dim TutanKarakterAdet = 0, uzunluk AsInteger
If Len(Kelime1) >= Len(Kelime2) Then
    uzunluk = Len(Kelime2)
Else
    uzunluk = Len(Kelime1)
EndIf
If uzunluk > 0 Then
For i = 1 To uzunluk
If Mid(Kelime1, i, 1) = Mid(Kelime2, i, 1) Then
    TutanKarakterAdet += 1
EndIf
Next
If uzunluk = TutanKarakterAdet Then
    DonenSonuc = True
Else
    DonenSonuc = False
EndIf
EndIf
Return DonenSonuc
EndFunction
EndClass

```