

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

739842

İNTERNET ÜZERİNDE ÇALIŞAN
BİR DOĞAL DİL İŞLEME UYGULAMASI:
SORU CEVAPLAMA SİSTEMİ

- 139842 -

Bilgisayar Müh. M. Fatih AMASYALI

FBE Bilgisayar Mühendisliği Anabilim Dalında
Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Prof. Dr. Oya KALIPSIZ

Doç. Dr. Coşkun Sönmez

Yrd. Doç. Dr. Banu Dizi

İSTANBUL, 2003

TE. YÖNETİM KURULU
DOKÜMANLARI MERKEZİ

İÇİNDEKİLER

Sayfa

KISALTMA LİSTESİ	iv
ŞEKİL LİSTESİ	v
ÇİZELGE LİSTESİ	vi
ÖNSÖZ.....	vii
ÖZET	viii
ABSTRACT	ix
1. GİRİŞ.....	1
1.1 Doğal Dil Nedir?.....	1
1.2 Dilbilgisi Kuramları.....	2
1.3 Doğal Dil İşleme ve Uygulamaları.....	4
1.4 Dünya Bilgisi	6
1.5 WordNet	7
1.6 Arama motorları.....	8
1.7 Soru cevaplama sistemleri	9
1.8 Soru cevaplama sistemlerinin ilgili olduğu diğer alanlar	10
2. ÖNCEKİ ÇALIŞMALAR	11
2.1 TREC yarışması ve kapsamı.....	11
2.2 TREC veritabanında ve Internet’te aramanın farkları	12
2.3 TREC için tasarlanan sistemler	13
2.3.1 Quantum	13
2.3.2 Shapaqa.....	14
2.3.3 SMART IR	15
2.3.4 WEBCLOPEDIA.....	15
2.4 Internet üzerinde çalışan soru cevaplama sistemleri	17
2.4.1 MULDER	18
2.4.2 AnswerBus.....	20
2.4.3 QuASM.....	22
2.4.4 AskMSR	24
3. TASARLANAN SİSTEM	25
3.1 Sisteme Genel Bakış	25
3.1.1 Sorunun Tipinin Belirlenmesi	26
3.1.2 Kullanıcı Sorusunun Çözümlemesi	27
3.1.3 Kullanıcı Sorusunun Çoklaştırılması.....	28
3.1.4 Arama Motorundan Sonuçların Alınması.....	29
3.1.5 Sonuç sayfalarından aday cümlelerin seçilmesi	30
3.1.6 Aday cümlelerin puanlanması	31

3.1.7	Cevapların kullanıcıya verilmesi	37
3.2	Sistemde kullanılan araçların ayrıntılı incelenmesi	38
3.2.1	Kelimeleri gövde ve eklerine ayıran ayırıştırıcı	38
3.2.2	Şablonlar Veritabanı	39
3.2.3	Kullanıcı sorularının toplanacağı ortak veritabanı	41
3.3	Sistemin Gerçekleştirilmesi	42
3.4	Şablonların Oluşturulma Metodu	42
3.4.1	Nedir sorularının şablonlarının oluşturulması	42
3.4.2	Nedir haricindeki soruların şablonlarının oluşturulması	43
4.	SİSTEMİN UYGULAMA SONUÇLARI.....	45
5.	SONUÇ ve ÖZET	63
6.	GELECEKTEKİ ÇALIŞMALAR İÇİN ÖNERİLER.....	64
KAYNAKLAR.....		65
ÖZGEÇMİŞ.....		66



KISALTIMA LİSTESİ

TREC	Text REtrieval Conference
SMART IR	Smart Information Retrival
MRR	Mean Reciprocal Rank
MG	Managing Gigabytes
HTML	Hyper Text Markup Language
FYE	Fiil yapım eki
IYE	İsim yapım eki
Ytk	yeterlilik kipi
ol	olumsuzluk eki
SFE	sıfat fiilimsi eki
ç	çoğul eki
ys	yardımcı ses
ieç1	iyelik eki 3. çoğul şahıs
ayr	ayrılma hali (“-den” hali)
bul	bulunma hali (“-de” hali)
tm	tamlayan eki (“-in” eki)
ku	koruyucu ünsüz
ef_dgz	ek fiil görülen geçmiş zaman eki
şet1	1. tekil şahıs eki
yon	yönelme hali (“-e” hali)
ef_gez	ek fiil geniş zaman eki
kh	kaynaştırma harfi
gez	geniş zaman
dgz	görülen geçmiş zaman eki
mgz	öğrenilen geçmiş zaman eki
şz	şimdiki zaman eki
ZFE	zarf fiilimsi eki
şeç1	1. çoğul şahıs ek
bli	belirtme hal eki (“-i” hal eki)
iet3	3. tekil iyelik şahıs eki
IFE	isim fiilimsi eki

ŞEKİL LİSTESİ

Şekil 1.1 Türkçe için derin yapıdan yüzeysel yapıya geçiş.....	3
Şekil 1.2 İngilizce için derin yapıdan yüzeysel yapıya geçiş.....	3
Şekil 1.3 WordNet'in sonuç sayfası(http://www.cogsci.princeton.edu/cgi-bin/webwn)	8
Şekil 2.1 Shapaqa sisteminin arabirimi(http://ilk.kub.nl/~antalb/abvi/week3/shapaqa.html)..	14
Şekil 2.2 Webclopedia sisteminin değerlendirme sonuçları(Hovy, 2001)	17
Şekil 2.3 MULDER'in diğer sistemlerle karşılaştırılması (Kwok, 2000)	20
Şekil 2.4 AnswerBus sisteminin genel yapısı (Zheng, 2002).....	20
Şekil 2.5(a) Tablonun sayfadaki görünümü	23
Şekil 2.5(b) Tablonun QuASM tarafından saklanış biçimi (Pinto, 2000).....	23
Şekil 3.1 Tasarlanan sistemin çalışma akışı	26
Şekil 3.2 Sistemin arabirimi ve verdiği cevap örnekleri	37
Şekil 3.3 Sistemin kelimeleri gövde ve eklerine ayırmasında izlediği yollar	39
Şekil 4.1 Sistemin değerlendirilmesi için tasarlanan arabirim	45
Şekil 4.2 Soru türlerinin MRR puanlarının gösterimi-1	59
Şekil 4.3 Soru türlerinin MRR puanlarının gösterimi-2	60
Şekil 4.4 Soru türlerinin MRR puanlarının gösterimi-3	60
Şekil 4.5 Soru türlerinin MRR puanlarının gösterimi-4	61
Şekil 4.6 Soru türlerinin MRR puanlarının gösterimi-5	61



ÇİZELGE LİSTESİ

Çizelge 1.1 Doğal dil işleme araçları ve uygulamaları arasındaki ilişki. (Daelemans,2002).....	5
Çizelge 1.2 İngilizce için örnek doğal dil işleme araçları	5
Çizelge 1.3 Soru cevaplama sistemlerinin ilgili olduğu alanlar	10
Çizelge 2.1 Soru Cevaplama sistemleri ve cevap formatları (Zheng, 2002).....	11
Çizelge 2.2 TREC-8,9,10'dan alınmış soru ve cevap örnekleri	12
Çizelge 2.3 İnternet üzerindeki aramayla, TREC veritabanında arama yapmanın farkları (Buchholz, 2002).....	13
Çizelge 3.1 Soru kelimelerinden soru türünün belirlenmesinde kullanılan tablodan bir kesit.	26
Çizelge 3.2 Cümlelerdeki kelimelerin ayrıştırılması ve etiketlenmesi örnekleri	27
Çizelge 3.3 Kelimelerin etiketlemesinde kullanılan anahtar kelimelere örnekler.....	28
Çizelge 3.4 Kullanıcı sorusunun eklerinden arındırılması örnekleri.....	29
Çizelge 3.5 Cevapların nerede aranacağına dair karşılaştırma.....	29
Çizelge 3.6 Sonuç sayfalarından aday cümlelerin seçilmesi örnekleri	30
Çizelge 3.7 Aday cümlelerin eşleştirme puanlarından örnekler.....	31
Çizelge 3.8 Aday cümlelerin sıralama puanlarından örnekler	32
Çizelge 3.9 Bileşen sorularında aranan şablonların listesi	33
Çizelge 3.10 Aday cümlelerin aldıkları puanlar	33
Çizelge 3.11 Kelimelerin çözümlenme örnekleri	39
Çizelge 3.12 Şablonlar veritabanından bir kesit.....	40
Çizelge 3.13 Nedir soru türü şablonları.....	43
Çizelge 3.14 Zaman şablonlarının elde edilmesinde kullanılan cümlelerden örnekler	44
Çizelge 4.1 Soru ve doğru cevap sayılarının soru türlerine göre dağılımı	47
Çizelge 4.2 Soru türlerine göre MRR puanlarının dağılımı	48
Çizelge 4.3 Puanlama türlerine göre MRR puanları	49
Çizelge 4.4 TREC yarışmasının en iyileri(MRR puanları)	49
Çizelge 4.5 Sisteme sorulan sorulara ve sistemin cevaplarına örnekler.....	50
Çizelge 4.6 MRR ve doğru cevap oranlarının karşılaştırılması.....	62

ÖNSÖZ

Bilginin değeri arttıkça bilgiye erişim metotları da doğal olarak artmaktadır. Bugün bilgiye erişimde en çok kullanılan yollardan biri arama motorlarıdır ancak bu arama motorları sadece site arama motorlarıdır. Bu tezde bir cevap arama motoru geliştirilmiştir. Geliştirilen cevap arama motoru kullanıcının günlük konuşma diliyle sorduğu sorunun cevabını bir arama motoruna bağlanarak cevaplamaya çalışan bir sistem olarak geliştirilmiştir.

Bir soru cevaplama sisteminin amacı kullanıcıyı istediği bilgiye daha hızlı erişirmek ve harcadığı emeği daha aza indirmektedir.

Arama motorları, cevaplar yerine cevapları içeren dokümanların adreslerini bulurlar. Kullanıcı yine gelen dokümanların içinde ihtiyacı olan bilgiyi aramak zorundadır. Kullanıcıları bu işlemlerden kurtarmak amacıyla soru cevaplama sistemleri geliştirilmiştir.

İngilizce için birçok örneği bulunmasına rağmen Türkçe için Internet üzerinden soru cevaplama sistemi geliştirilmemiştir. İşte bu çalışmada bir ilk olarak bu tür bir sistem geliştirilmeye çalışılmıştır.

Çalışmam boyunca beni destekleyen; tez yöneticim Prof. Dr. Oya Kalıpsız'a , desteğini ve ilgisini her zaman hissettiğim hocam Yrd. Doç. Dr. Banu Diri'ye, tüm destek ve yardımlarından dolayı Seyhan İşler'e, çalışma ve oda arkadaşlarıma ve özellikle aileme destekleri ve yardımları için teşekkür ederim.

ÖZET

Bilgiye erişimin en yaygın ve en kolay yollarından biri Internet, Internet'teki bilgiye erişimin en kolay yolu ise arama motorlarıdır. Ancak arama motorlarından istenilen bilgiye erişmek için önce uygun anahtar kelimeler seçmeli, daha sonra sonuç sayfasını yada sonuç sayfalarındaki bağlantılardaki sayfaları incelenmelidir. Tüm bu işlemler kullanıcı tarafından yapılmak zorundadır. Oysa kullanıcıların belirli tipteki soruları için tüm bu işlemleri otomatik hale dönüştürmek mümkündür ve İngilizce için birçok sistem geliştirilmiştir. Kullanıcısından sorusunu doğal dillerde kabul edebilen ve kullanıcıya cevabı bulması için sayfa adresleri listesi yerine cevabın kendisini verebilen sistemlere soru cevaplama sistemleri denmektedir.

Bu çalışmada; Türkçe için doğal dille çalışan bir soru cevaplama sistemi gerçekleştirilmiştir. Sistem öncelikle kullanıcısının doğal dille sorduğu sorusunu arama motoru sorgusuna çevirmekte ve arama motorunun sonuç sayfasından yada bağlantılarındaki sayfalardan olası cevap cümlelerini seçmektedir. Olası cevap cümlelerini çeşitli kriterlere göre puanlandırıp en yüksek puanı alan ilk beş cümle kullanıcıya iletilmektedir. Sistem 524 adet soruyla değerlendirilmiş ve arama motorunun sonuç sayfası kullanıldığında soruların yaklaşık %43'üne, sonuç sayfasının bağlantılarındaki sayfalar kullanıldığında %60'ına ilk beş cümle içinde doğru cevap verebilmiştir.

Anahtar kelimeler: Doğal Dil İşleme, Soru cevaplama sistemleri.

ABSTRACT

The easiest and common way to achieve information is Internet. The easiest way to achieve information on Internet is using a search engine. But user most do extra processes (selection of keywords, research on result pages and web sites) while using a search engine. These extra processes can be done automatically for some kind of queries. The systems, that accepts user queries in natural language and results in only answer not a document list, are named Question-Answering Systems. The instances of these systems are exist for English.

In this study, a Question-Answering system is built for Turkish. The system converts user question in natural language to search engine query and sends these queries to a search engine and receives result page. In result page and the web pages which web addresses are in result page, the answer candidate sentences are selected. The candidate answers are ranked according to four criteria. The five candidates which have the highest scores are the answers. The system is examined with 524 questions. 43% of questions are answers correctly in first five sentences when the result page of search engine was used. 60% of questions are answers correctly in first five sentences when the real pages (at the result page's links) was used.

Keywords: Natural Language Processing, Question Answering Systems.

1. GİRİŞ

Son yıllarda İnternet arama motorları ve bilgi dizinleri önemli adımlar atmış olsalar da, bilginin İnternet’te verimli bir şekilde yerleştirilmesi problemi çözümden uzaktır. Bu çalışmada İnternet üzerindeki bilgiye bir kişiye soru sorar gibi (doğal dille) sorular sorarak erişimi sağlayan bir sistem geliştirilmiştir.

Sistemin odaklandığı soru türleri, “ Dünyadaki ikinci en yüksek dağ hangisidir? ” , “ Uzaya giden ilk insan kimdir? ” gibi gerçek sorulardır. Her ne kadar bu çeşit sorular, arama motorlarında bulunan sınırlı sorgulama alt kümesini temsil etseler de, cevabı bulabilmek için kullanıcı tarafından harcanacak zamanın önemli derecede azaltılması olanağını içermektedirler. Çalışmanın amacı, kullanıcıyı bir dizi arama motoru sonuç sayfalarındaki sorunun cevabını potansiyel olarak içeren bir çok web sitesini okumaya zorlamak yerine, kullanıcıya cevabı içeren cümleler veren bir sistem oluşturmaktır.

Kullanıcı sorusunun kesin bir cevabı olduğu durumlarda (ör: Plüton’u kim keşfetti? Polonya’nın başkenti neresidir?) puanlanmış bir cevap listesi, puanlanmış bir doküman adres listesinden çok daha kullanışlıdır.

Sistemin amacını gerçekleştirmek üzere kullanılan modüller, değerlendirmenin daha etkili bir biçimde tasarlanması için ayrı ayrı geliştirilmiştir. Bu sayede her bir modülün sistemin performansına etkileri daha kolay anlaşılmıştır.

Birinci bölümün devamında konu ile ilgili diğer alanlara ait giriş niteliğinde bilgiler verilmiş, İkinci bölümde ise bu konuda daha önceden yapılmış olan çalışmalardan bir kısmı anlatılmıştır. Üçüncü bölümde, tasarlanan sistemin çalışma akışı ve modülleri ayrıntılı bir biçimde sunulmuş, dördüncü bölümde ise sistemin farklı puanlama türlerine göre sistemin performansı sunulmuş ve değerlendirilmiştir. Beşinci bölümde ise yapılan çalışmadan çıkarılan sonuçlar verilmiş ve son bölümde gelecek çalışmalar için önerilerde bulunulmuştur.

1.1 Doğal Dil Nedir?

Dil, başlıca üç işlevi olan toplumsal bir varlıktır. İlki iletişim kurma, ikincisi bilgi aktarma üçüncüsü ise düşünceyi geliştirmedir(Erkman,1998).

Dil konusunda yapılan çalışmalarda tüm dillerde geçerli olan özellikler bulunmaya çalışılmıştır. Örneğin nesneleri adlandırma, bunları sıfatlarla niteleme, bu nitelediğimiz şeylere fiillerle bir oluş, bir durum yada bir davranış yakıştırma, olayları zaman içinde sıralama, olayları belli bir bakış açısına göre yansıtmaya tüm dillerde bulunan ortak özelliklerdir

(Erkman, 2000).

İşte tüm dilleri kapsayan ortak özelliklere dilbilim(linguistic), tek tek dillerin özelliklerine ise o dilin dilbilgisi(gramer) denmektedir.

Dilbilgisi kapsamına sesbilgisi dediğimiz, ses dizgesi sistemi, anlambilim dediğimiz anlamlar dizgesi, biçimbilim dediğimiz sözcüklerin yapısını belirleyen dizge, cümlelerin yapısını belirleyen söz dizim dizgesi ve sözcük dağarcığı, yani sözlükte dahildir (Fromkin, 1993).

1.2 Dilbilgisi Kuramları

Çağdaş dilbilimin kurucusu sayılan Saussure, dilbilimini diğer bilim dallarının uzantısı olmaktan kurtarmıştır. Ancak Saussure dile çok genel bir açıdan bakar, ardından gelen dilbilimciler, dili katmanlara ayırmışlar ve daha dar alanlarda dili incelemişlerdir. Bu çalışmaların sonucunda dünyadaki dillerin birçok ortak noktası olduğu gözlenmiştir.

Dilin katmanları olarak sesbilgisi, sözlükbilim, anlambilim, söz dizim ve edim bilim belirlenmiştir.

Sesbilgisi: İnsanların çıkarabildikleri ses sayısı 35 civarındadır. Bu sesleri her dilde farklı kurallar dahilinde birbirine ekleyerek sözcükleri meydana getiririz. Örneğin Türkçe’de hiçbir sözlük iki ünsüz harf ile başlamaz. İngilizce’de 'mb' ile başlayan sözcük bulunmaz. Seslerin her türlü özelliklerini kapsayan dilbilgisi katmanı bu ses katmanıdır.

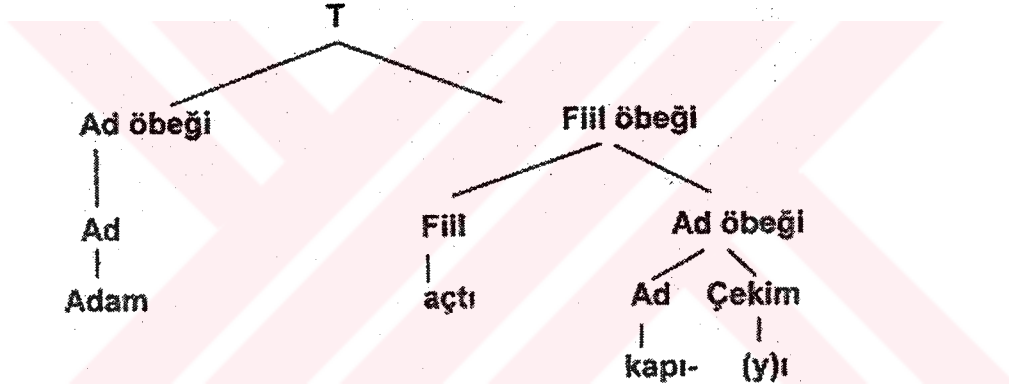
Sözlükbilim ve anlambilim: Seslerden meydana gelen sözcüklerin incelenmesi sözlükbilimin ve anlambilimin kapsamına girer. Diller arasında yapılan anlambilimsel karşılaştırmalarda, sözcüklerin anlam alanlarının tam olarak birbirine uymadığı görülmüştür. Örneğin almak sözcüğü hem bir yerde duran şeyi almak, hem de fiyatını ödeyerek almak anlamlarında kullanılır. Oysa İngilizce’de iki anlam içinde birbirleri yerine kullanılmayan farklı birer sözcük vardır. (take, buy).

Söz dizimi: Sözcüklerin birbirine eklenmesiyle cümlelerin nasıl oluştuğunu araştırır. Sözcükler her dilde birbirine farklı bağlanırlar. Türkçe’de bu bağlanma çoğu durumda çekim ekleriyle yapılır. Biçimbilim(morphology) ekleri inceleyen bilim dalıdır.

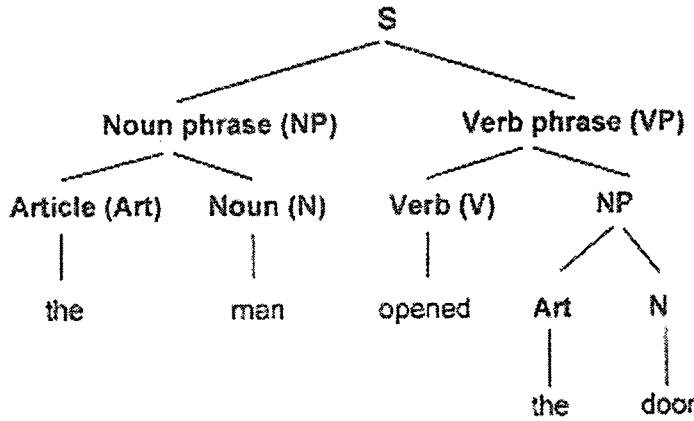
Söz dizim kuramının öncüsü Chomsky’dır. Chomsky diller arasındaki benzerlikleri söz dizim düzenlerinde aramıştır. Zaten söz dizim üzerine çalışanların tümü bir cümlelerin öğeleri arasındaki gizli bağları ortaya çıkarmaya çalışmışlardır. Chomsky ve onu izleyen üretici-dönüşümsel dilbilgisi ekolü, diyagramlar çizerek cümle içerisindeki bu hiyerarşik bağlantıları

göstermişlerdir. Chomsky'e göre bu örtük bağlantılar dile getirmek istediğimiz düşünce aşamasında hemen her dilde aynıdır. Düşünce aşamasındaki bu alt düzleme Chomsky derin yapı(deep structure) adını verir. Ancak düşünceyi dışa vurma aşamasında, diller arasındaki farklar belirlenir. Her dil kendi diziliş kurallarını dayatır. Chomsky, işte bu dışa vurma aşamasına da düzeysel yapı(surface structure) adını verir. Dışa vurma aşamasına geçmek için, insanlar önce düşünce üretir, sonra bu düşüncelerini(önergelerini) konuştukları dile uygun olarak cümlelere dönüştürürler. Üretici-dönüşümsel(generative-transformational) adı buradan gelmektedir (Erkman, 2000).

Chomsky'ye göre bir cümle bir ad öbeği ve bir fiil öbeğinden oluşur. Her öbeğin kendi içinde bir yönetim şeması vardır. Örneğin derin yapıda, bir adama bir kapıyı açma eylemini yakıştırıyorsak derin yapıdan yüzeysel yapıya geçişimiz Türkçe için şekil 1.1, İngilizce için şekil 1.2'deki gibi olacaktır(Erkman, 2000).



Şekil 1.1 Türkçe için derin yapıdan yüzeysel yapıya geçiş.



Şekil 1.2 İngilizce için derin yapıdan yüzeysel yapıya geçiş.

Chomsky'ye göre tüm tümleçler fiil öbeğine bağlıdır, dolayısıyla fiil öbeğinin içinde yer alırlar. Bu diyagram yüzeysel yapıdaki sırayı tam olarak göstermez. Türkçe de –i tümleci, fiilden önce söylenecektir(adam kapıyı açtı), oysaki İngilizce'de fiilden sonra söylenecektir(the man opened the door).

Ayrıca yüzeysel yapıda da cümle kuruluş türleri fazla olmadığı için hemen hemen tüm dillerde cümledeki öğeler aşağıdaki üç temel dizilişe uyarlar.

Özne- Yüklem-Tümleç

Özne- Tümleç -Yüklem

Yüklem-Özne- Tümleç

1.3 Doğal Dil İşleme ve Uygulamaları

Doğal dil işleme; insanların kendi aralarında iletişim için kullandıkları dilin bilgisayarlar yardımıyla çeşitli amaçlarla işlenmesidir. Aşağıda çeşitli doğal dil işleme uygulamalarından örnekler verilmiştir.

Metin Sınıflandırma: Bir metni önceden belirlenmiş kategorilerden hangisine ait olduğunu bulan sistemlerdir.

Yazar Tanıma: Bir metnin yazarının bulan sistemlerdir.

Dil Tanıma: Bir metnin hangi dille yazıldığını belirleyen sistemlerdir.

Konu Belirleme: Bir metnin hangi konuda yazılmış olduğunu belirleyen sistemlerdir.

Bilgi Çıkarımı: Bir metinden daha önceden belirlenmiş şablonların doldurulmasını gerçekleştiren sistemlerdir.

Özet Çıkarımı: Bir metnin özetini çıkaran sistemlerdir.

Soru Cevaplama: Kullanıcısının sorduğu sorulara cevap bulan sistemlerdir.

Doğal dilleri işlemek için birçok araç geliştirilmiştir. Doğal Dil işleyen yazılım araçlarıyla, uygulamada nerelerde kullanıldıkları çizelge 1.1'de gösterilmiştir. Yukarıdan aşağıya doğru kelimelerin anlamlarına olan ihtiyaç artmaktadır.

Çizelge 1.1 Doğal dil işleme araçları ve uygulamaları arasındaki ilişki. (Daelemans,2002)

NLP araçları	METİN	Uygulamalar
Morfolojik Analiz		Yazım Denetleme
Kelimelerin Türlerini Bulma		Grammer Denetleme
Birliktelikleri Bulma(Tamlama vb.)		Doküman Erişim
Sözdizimsel Analiz		Doküman Sınıflandırma
Kelimeler Arası İlişkileri (öğeleri) Bulma		Bilgi Çıkarımı
Etiketleme		Özet Çıkarımı
Şüpheli Durumların Çözümlemesi		Soru Cevaplama
Anlamsal Analiz		Diyalog Sistemleri
Zamir Çözümleme		Otomatik Tercüme
Söylem Analizi	ANLAM	

İngilizce için birçok Doğal Dil İşleme aracı mevcuttur. Çizelge 1.2’de bunlardan bir kısmının adresleri verilmiştir.

Çizelge 1.2 İngilizce için örnek doğal dil işleme araçları

Senga: Information Retrieval Software	http://www.senga.org
GATE: General Architecture for Text Engineering	http://www.gate.ac.uk/
Banter Technology	http://www.banter.com
AFGL Project: Affix Grammars over a Finite Lattice	http://www.cs.kun.nl/agfl/
Alembic Workbench	http://www.mitre.org/technology/alembic-workbench/
Thistle	http://www.ltg.ed.ac.uk/software/thistle/demos/index.html
Smart Tutorial	http://pi0959.kub.nl:2080/Paai/Onderw/Smart/hands.html
Parser Servlet	http://www.teemapoint.net/nlpdemo/servlet/ParserServlet

1.4 Dünya Bilgisi

Sadece sözdizimsel analiz yapan bir sisteme “Ali bugün okulda tahtaya kalktı.” cümlesi girildiğinde; sistem “Ali tahtaya kalktı mı? “ sorusuna rahatlıkla yanıt verebilecektir. Fakat “Ali bugün okula gitti mi?” sorusuna, beş yaşında bir çocuk bile yanıt verebilecekken, sistem yanıt veremeyecektir.

Bilgisayar sistemleri dünyanın en iyi satranç şampiyonlarını bile yenerken, hesaplamalar konusunda uzmanları bile oldukça gerilerde bırakırlarken, bazı konularda(yukarıda bahsedilen örnek gibi) beş yaşında bir çocuğa dahi yetişmemektedirler. Bunun sebebi bizlerin yaşadığı dünya hakkında , bizler hakkında hiçbir şey bilmemeleridir. Bilgisayarlara böylesi bilgiler (dünya bilgileri) aktarıldığında bizlere daha da çok yardımcı sistemler olacaklardır.

◆ Örneğin böylesi gündelik bilgilere sahip olan bir web arama motoruna

“kedim hasta.” ifadesini girdiğimizde olası sonuçlar incelenirse:

İnsanlar ev hayvanlarını severler.

Kedi bir ev hayvanıdır.

İnsanlar sevdikleri canlıların mutlu olmasını isterler.

Canlılar hastayken mutlu olmazlar.

İnsanlar ev hayvanlarının sağlıklı olmasını isterler.

Kedi hasta ise sağlıklı değildir.

Hayvanları hasta durumundan veterinerler kurtarır.

Ev hayvanları, hayvandır.

Kedi hayvandır.

O halde kullanıcıma veteriner adresleri gerekiyor.

Kullanıcımın yakınındaki veteriner adreslerini ona vermeliyim.

Kullanıcım şu an x'te bulunmakta.

ve ekranda : X'e yakın yerlerdeki veteriner adresleri görünür.

Dünya bilgisi'nin yapay zeka'nın gelişimindeki önemi çok büyüktür. Bundan sonra daha etkili gelişmelerin olabilmesi için bu bilgilerin bir şekilde bilgisayarlar tarafından da

kullanılmasını sağlamak yapay zeka topluluğunun ana hedefleri arasında yer almaktadır.

Bu bilginin toplanmasında bugüne kadar birçok deneme gerçekleştirilmiştir. En büyük veri tabanına sahip üçü:

-CYC, Doug Lenat (yaklaşık 3 milyon önermeden oluşan uzmanlarca 10 yıldan beri veri girilen bir veritabanıdır. Erişim için : <http://www.cyc.com>).

-Open Mind, Push Singh (yaklaşık yarım milyon önermeden oluşan, Web sitesi sayesinde herkesin ortak katkısıyla 1999'dan beri veri girilen bir veritabanıdır. Erişim için: <http://openmind.media.mit.edu>).

-Thought Treasure, Eric Mueller (35,000 İngilizce, 21,000 Fransızca kelime ve deyiş, 21,000 kavram ve bu kavramlara ait 51,000 önerme içeren bir veritabanıdır. Erişim için : <http://www.signiform.com/tt/htm/tt.htm>).

1.5 WordNet

WordNet birçok doğal dil işleme programının kullandığı bir sözlüktür. Diğer sözlüklerden farklı olarak çok daha iyi yapılandırılmıştır ve bu sayede içerdiklerinin bilgisayarlar tarafından işlenmesi mümkün hale gelmiştir. WordNet sistemi kelimeleri, anlamlarını, eşanlamlılarını, ilişkili olduğu kelimeleri ve ilişki türlerini, kelimelere erişim için çeşitli yazılımları içerir. Şekil 1.3'te WordNet'te Lincoln kelimesinin arama sonuçları gösterilmiştir.

WordNet 1.7.1 Search

Search word:

Overview for "Lincoln"

The **noun** "lincoln" has 3 senses in WordNet.

1. Lincoln, Abraham Lincoln, President Lincoln -- (16th President of the United States; saved the Union during the Civil War and emancipated the slaves; was assassinated by Booth (1809-1865))
2. Lincoln, capital of Nebraska -- (capital of the state of Nebraska; located in southeastern Nebraska; site of the University of Nebraska)
3. Lincoln -- (long-wooled mutton sheep originally from Lincolnshire)

Search for of senses

- ☒ Show glosses
☐ Show contextual help

Şekil 1.3 WordNet'in sonuç sayfası(<http://www.cogsci.princeton.edu/cgi-bin/webwn>)

WordNet sadece kelime anlamlarını değil kelimelerin eşanlamlılarını, varlık hiyerarşisi içindeki yerlerini ait oldukları sınıfı ve altlarında bulunan sınıfları da göstermektedir. Bir kelimenin istenirse sadece kendisine ait özellikleri istenirse üst sınıflarından mirasla aldığı özellikleri de gösterilmektedir. Örneğin bir kedinin sadece kendine ait olan miyavlama gibi bir özelliği ve üst sınıflarından mirasla aldığı hücreleri olma(canlı olduğu için), hareket edebilme(hayvan olduğu için) gibi özelliklerde bilgi olarak sunulmaktadır. Daha fazla bilgi <http://www.cogsci.princeton.edu/~wn/> adresinden edinilebilir.

Türkçe için böylesi bir kaynağın bulunmaması Türkçe Doğal Dil işleyenler için büyük bir engel olarak ortaya çıkmaktadır. Özellikle anlambilimsel çalışma yapanlar için kelimelerin anlamlarına erişmek ve bilgisayarla işlemek her zaman büyük bir problem olmuştur.

1.6 Arama motorları

Arama motorlarının bir an için yok olduğu düşünüldüğünde arama motorlarının Internet'e

eriřim için(dolayısıyla bilgiye eriřim) kullanılan en önemli kaynak olduđu anlařılır. Arama motorları olmadığında web sayfalarının adresleri ancak etraftaki kiřilerden öğrenilebilirdi. Bu durumda Internet'teki bilginin ne kadar az bir kısmına erişilebileceđi açıkça ortaya çıkar.

Arama motorlarının çoğunun içerisinde her kelime için hangi dokümanda ve kaçar kere geçtiğinin tutulduđu bir veritabanı mevcuttur.

Böylesi bir yapı kullanıcının arama sorgusuyla ilgili metinlere kolayca erişmesini sağlamaktadır. Bir kelimenin bir doküman içerisinde kaç kere geçtiđi, dokümanların sorguyla olan ilgisinin derecesini de bulmakta kullanılır.

Bu yapıya bakılarak genelde bir sitenin arama motoruna kaydının ařağıdaki adımları içerdiđi söylenebilir.

- Sitenin sayfasındaki kelimelerin ve kaçar defa geçtiklerinin bulunması.
- Veritabanında eđer bir kelime varsa geçtiđi adreslere yeni sitenin adresinin ve frekansının eklenmesi
- Yoksa veritabanına bu kelimenin, yeni sitenin adresteki frekansının ve site adresinin eklenmesi

Daha karmařık arama motoru veritabanlarında kelimelerin bir site içinde nerede geçtikleri de tutulur. Kullanıcı sorgusundaki kelimelerin site içinde birbirine yakınlıđı, sitelerin birbiriyle karřılařtırılmasında kullanılan bir yöntemdir.

1.7 Soru cevaplama sistemleri

Internet'teki bilgi patlamasının sonucu olarak dođal dille çalışan soru cevaplama sistemleri büyük bir potansiyel olarak görülmeye bařlandı. Soru cevaplama sistemlerinin ilk geliřtirilen ilk örnekleri genelde, sınırlı konularda çalışmak üzere tasarlanmışlardı.

TREC-8(Text REtrieval Conference) ile birlikte, “gerçek dünya verileriyle”, “konu bağımsız soru cevaplama sistemleri” birbiriyle karřılařma olanağı buldu.

Günümüzde kullanılan arama motorları aranılan bilgiyi içeren yada ilgili olabilecek dokümanların kendilerini veya yerlerini kullanıcılarına sunmaktadırlar. Kullanıcı karřısına sunulan puanlanmış bir listeden ihtiyaç duyduđu bilgiyi çıkarma işlemini kendi başına yapmak zorundadır. Bu işlem yine (göreceli olarak) büyük miktarda dokümandan bilgi çıkarımı işlemine dönüşmektedir.

Kullanıcı sorusunun kesin bir cevabı olduğu durumlarda; bu işlemde kullanıcıyı kurtarmak mümkün gözükmemektedir. Günümüzde bu işlemi(kullanıcıya sadece ihtiyacı olan bilgiyi verme) gerçekleştirecek araçlara ihtiyaç duyulmakta olup Internet'teki bilgi miktarı arttıkça bu ihtiyaç da artacaktır.

1.8 Soru cevaplama sistemlerinin ilgili olduğu diğer alanlar

Soru cevaplama sistemlerini incelerken konuyla yakından ilgili olan diğer konularla beraber incelemenin faydalı olduğu görülmüştür. Soru cevaplama sistemlerinin yakından ilgili olduğu diğer alanlar ve kısa açıklamaları Çizelge 1.3'te verilmiştir.

Çizelge 1.3 Soru cevaplama sistemlerinin ilgili olduğu alanlar

Alan	Açıklama
Doğal Dille İlişkisel Veritabanlarını Sorgulama	Yapısı belli bir ilişkisel veritabanı üzerinde SQL yerine doğal dille sorgulanma yapılmasına imkan veren sistemlerdir.
Bilgiye Erişim Sistemleri	Kullanıcının istediği konudaki dokümanları yada adreslerini bulan sistemlerdir. Tüm arama motorları bu tür sistemlere örnektir.
Bilgi Çıkarım Sistemleri	Kullanıcının önceden tanımladığı şablonlar ile dokümanları tarayarak dolduran sistemlerdir. Şablonlar bir veri tabanı da olabilir. Örneğin firmaların web sitelerinden firma profillerini bulup daha önceden yapısı verilmiş olan bir veri tabanı tablosuna yazma işlemi bu tür bir işlemdir.
Okuduğunu Anlama	Sisteme tek bir metin verilir; yine bu metinle ilgili sorulara cevap verilmesi istenmektedir.
Soru-Cevap İkili İlişkileri İçeren Veritabanlarında Cevap Arama	Soru-cevap ikilileri içeren veritabanlarında(ör: sıkça sorulan sorular), web sitelerinde(ör: yardım masası) kullanıcının sorduğu soruya en yakın olan sorunun bulunup oradaki cevabın kullanıcıya iletilmesi işlemidir.

2. ÖNCEKİ ÇALIŞMALAR

Türkçe ile yapılmış bir çalışma olmamasına karşın İngilizce ile çalışan birçok çalışma mevcuttur. Çizelge 2.1’de bu sistemlerden bazıları ve kullanıcıya verdikleri cevapların formatları gösterilmektedir.

Çizelge 2.1 Soru Cevaplama sistemleri ve cevap formatları (Zheng, 2002)

Sistem	Cevap Formatı
AnswerBus	Cümleler
AskJeeves	Dokümanlar
IONAUT	Paragraflar
LCC	Cümleler
Mulder	Cevaplar
QuASM	Dokümanlar
START	Çeşitli
Webclopedia	Cümleler

2.1 TREC yarışması ve kapsamı

TREC konferansları sekizincisinden itibaren konu bağımsız soru cevaplama sistemlerinin birbiriyle karşılaştırıldığı bir yarışmaya dönüşmüştür. Yarışmanın amacı büyük kapsamlı metinlerde yer alan cevaplara erişimdir.

Yarışmaya katılan sistemlere 1,000,000 dokümandan oluşan (3GB) gazete makaleleri ve her sene değişen miktarda (200-700 adet) soru verilmektedir.

Yarışmacılardan istenen her soru için 5 adet sıralaması yapılmış cevaplar listesidir. Cevapların 50 yada 250 karakterden oluşması istenmektedir.

Yarışmacıların değerlendirilmesi ise uzman kişilerce yapılmaktadır. Uzmanlar sistemlerin verdikleri cevap listesinde gerçek cevabın olup olmadığına karar vermektedirler. Cevabın 5 elemanlı listenin neresinde olduğuna göre sistemler her sorudan bir puan almakta ve sonuçta alınan puanların ortalaması yarışmacının puanını ortaya çıkarmaktadır.

Cevaplara MRR(Mean Reciprocal Rank) puanların verilmesi işleminde izlenen yol: cevap birinci sırada yer alıyorsa 1, ikinci sırada yer alıyorsa 1/2, üçüncü sırada yer alıyorsa 1/3, dördüncü sırada yer alıyorsa 1/4, beşinci sırada yer alıyorsa 1/5, listede yer almıyorsa 0 puan almaktadır.

Çizelge 2.2’de TREC-9 ve TREC-10 sorularından ve cevaplarından örnekler verilmiştir.

Çizelge 2.2 TREC-8,9,10’dan alınmış soru ve cevap örnekleri

TREC soru- cevap örnekleri	
Soru1	How far is it from Denver to Aspen?
Cevap1	The Aspen/Snowmass area is about 200 miles southwest of Denver.
Soru2	What county is Modesto, California in?
Cevap2	Modesto in Stanislaus County.
Soru3	Who was Galileo?
Cevap3	Galileo Galilei, the astronomer
Soru4	What is an atom?
Cevap4	Atoms, long considered to be the smallest units of matter.
Soru5	What was the name of the first Russian astronaut to do a spacewalk?
Cevap5	Aleksei A. Leonov
Soru6	Where is Belize located?
Cevap6	Central America
Soru7	How much folic acid should an expectant mother get daily?
Cevap7	400 micrograms
Soru8	What type of bridge is the Golden Gate Bridge?
Cevap8	Suspension
Soru9	What is the population of the Bahamas?
Cevap9	250,000
Soru10	Who is the President of Ghana?
Cevap10	Rawlings
Soru11	How far is Yaroslavl from Moscow?
Cevap11	280 miles
Soru12	What country is the biggest producer of tungsten?
Cevap12	China
Soru13	When did the Jurassic Period end?
Cevap13	144 million years ago

2.2 TREC veritabanında ve Internet’te aramanın farkları

TREC veritabanında arama yapmakla Internet üzerinde arama yapmanın birçok yönden farkları bulunmaktadır. Çizelge 2.3’de ikisinin arasındaki farklılıklar gösterilmiştir.

Çizelge 2.3 Internet üzerindeki aramayla, TREC veritabanında arama yapmanın farkları
(Buchholz, 2002)

	TREC	INTERNET
Doküman sayısı	<1,000,000	<3,000,000,000
Doküman türü	Gazete makaleleri	Her tür
İşlem	Internet'ten bağımsız	Internet'e bağlı
Cevap formatı	50-250 bytes	Her türlü
Yanlış bilgi	Nadir	Çok fazla
Doğru cevabı içeren doküman sayısı	Az	Çok fazla

2.3 TREC için tasarlanan sistemler

Bugüne kadar TREC yarışmaları için birçok sistem geliştirilmiştir ve halende geliştirilmeye devam edilmektedir. Bunlardan birçoğu incelenmiş ve 4 tanesinin(Quantum, Shapaqa, QuASM, AskMSR) çalışma mantıklarına değinilmesi uygun görülmüştür.

2.3.1 Quantum

TREC veritabanı üzerinde çalışan Luv Plamondon ve Lelia Kosseim tarafından 2002 yılında geliştirilen bir soru cevaplama sistemidir.

Sistem, kullanıcı sorusunu analiz ederek uygun bir çıkarım fonksiyonu seçmektedir. Çıkarım fonksiyonları cevapların dokümanlar içinde nasıl yer aldığını göstermektedir. Çıkarım fonksiyonlarını $C=f(p,q)$ şeklinde ifade edersek p paragraf, q soru sınıfı, C ise olası cevaplar kümesidir. Paragraf OKAPI (Robertson, 1998) adlı bir bilgi erişim motoru kullanılarak elde edilmiştir. OKAPI'ye soru gönderildiğinde soruyla ilgili en çok 30 adet paragraf geri döndürmektedir.

Çıkarım fonksiyonlarında etiketler, eş anlamlı kelimeler kullanılmıştır. Etiketler Alembic(<http://www.cogsci.ed.ac.uk/~amyi/mate/alembic.html>) adlı bir sistemle; eş anlamlılar ise Wordnet kullanılarak bulunmuştur.

Çıkarım fonksiyonlarına örnek olarak tanım, adet, mesafe, özellik, kişi ve yer verilebilir..

Eğer cevap TREC yarışmasının istediği 50 karakterden daha kısa ise bulunan cevap ortada olacak şekilde sağdan ve soldan 50 karaktere erişilinceye kadar genişletilmiştir.

Olası cevapların puanlanmasında iki adet puanlama yöntemi kullanılıp bunların toplamı olası cevabın puanı olmaktadır. İlki çıkarım fonksiyonunun uyuma göre verdiği puan, ikincisi de OKAPI'nin olası cevabın geçtiği paragrafa verdiği puandır.

Sistem 492 adet TREC sorusuyla test edilmiş ve MRR dan 0.223 almıştır.

2.3.2 Shapaqa

Şekil 2.1'deki gibi bir arabirime sahip olan sistem, anlaşılacağı gibi tam bir doğal dille çalışmamaktadır. Kullanıcıdan sorusunu öğelerine ayırmasını istemekte ve daha sonra cevapları kullanıcıya döndürmektedir.

Maximum content length:

Specific URL: or

Use: ☐ AltaVista (slow, whole sentences) or ☒ Google (fast, partial sentences)

Number of documents per time: Start at document:

Include: ☒ Result summary ☒ Single results ☒ Non-results ☐ Not found, too long etc.

☐ Instances ☐ Chunked

☒ Grammatical Relations ☐ Chunk Types

Who/What	?
did	invented
whom/what	radium
when	
where	
why/what for	
how	
how much	
about	
as	

Şekil 2.1 Shapaqa sisteminin arabirimi(<http://ilk.kub.nl/~antalb/abvi/week3/shapaqa.html>)

Örneğin Shapaqa'nın arabirimi kullanılarak yukarıda radyumu kimin bulduğu sorulmuştur. Sistem cümleleri öğelerine ayırarak sorunun öğelerine ve öge yapısına(ağaç yapısına) en

yakın olan cevapları getirmektedir. Ancak tezin yazımı sırasında Shapaqa'nın sistemindeki bir problem yüzünden cevap alınamamıştır. TREC-8'in 200 sorusunda test edilen sistemin MRR sonucu 0.28'dir.

2.3.3 SMART IR

Steven Abney, Michael Collins ve Amit Singhal tarafından TREC dokümanları üzerinde cevap bulması için tasarlanan sistem temel olarak 5 modülden(İlgili paragrafların seçilmesi, olası cevapların elde edilmesi, olası cevapların sınıflandırılması, sorunun sınıflandırılması, cevapların puanlanması) oluşmaktadır (Abney, 2000). İlgili paragrafların TREC veritabanından seçilmesi işleminde Eşitlik 2.1 deki formül kullanılmıştır.

$$Paragrafpuan_i = \frac{1}{4}S_{i-1} + \frac{1}{2}S_i + \frac{1}{4}S_{i+1} \quad (2.1)$$

S_i soru cümlesindeki kaç kelimenin i . paragrafta yer aldığını göstermektedir. Formülden de anlaşılacağı gibi her paragrafın puanı kendinden öncekine ve sonrakine etki etmektedir. En yüksek puanı olan 50 paragraf ikinci modüle gönderilmektedir. İkinci ve üçüncü modülde, ilk modülden gelen kelimeler etiketlenmektedir. Etiketlere örnek olarak kişi, yer, şirket, tarih, sayı ve süre verilebilir. Dördüncü modülde içerdiği soru kelimelerine göre sorunun neyi sorduğu belirlenmektedir; diğer bir ifadeyle soru sınıflandırılmaktadır. Sorunun sınıfı ile kelimelerin sınıfları(etiketleri) aynı kümedirler. Son modülde ise olası cevaplar puanlanmaktadır. İlk olarak sorunun ve kelimenin etiketinin aynı olup olmadığına bakılmaktadır. Daha sonra her sınıf için özel olarak geliştirilmiş uygunluk ölçütleriyle cevaplar puanlanmaktadır.

Sistemin TREC-8 sorularındaki MRR puanı 50 karakterlik cevaplarda 0.261, 250 karakterlik cevaplarda ise 0.545'tir.

2.3.4 WEBCLOPEDIA

Webclopedia (Hovy, 2001) soru cevaplama sistemi, sonuçlarının doğruluğunu(kesinliğini) geliştirmek için yazımsal ve anlamsal bilginin kullanımını sağlamaktadır. Daha önceki soru cevaplama sistemleri de bu şekilde dış bilginin(dünya bilgisinin) kullanılması gereğini açıkça ortaya koymuştur. Örneğin, "Bant genişliği nedir? " gibi tanım türü sorularda, sistem cevap yığığında tanımları aramaya başlamadan önce, terimsel açıklamada kullanılan kelimeleri seçmek için WordNet kullanmaktadır.

Webclopedia soruları cevaplamak için iki temel kaynak kullanmaktadır. İlki dilbilgisi, ikincisi

de dünya bilgisidir. İlki doğal dil işleme araçlarından oluşmakta; ikincisi ise 140 adet soru tipi ve bunlara ait şablonlardan, ayrıca 10,000 düğümden oluşan kavramlar hiyerarşisinden oluşmaktadır.

Soru türleri Internet'ten indirilmiş yaklaşık 17,000 adet soru-cevap ikililerinin analizinden elde edilmiştir. Soru tiplerine ait şablonlar ise yine bu soru-cevap ikililerinden otomatik olarak çıkarılmış ve belli ağırlık değerleri verilmiştir.

Şablonların otomatik olarak elde edilmesi işlemi aşağıdaki adımlardan oluşmaktadır:

- Örnek soru-cevap ikilileri alınmış ((Mozart-1756)(Gandhi-1869)...) ve Altavista'ya sorgu olarak gönderilmiştir.
- İlk 1000 sonuç sayfasındaki bu ikililerin beraber geçtiği cevap cümlelerinde en uzun eşleşen parçalar bulunmuştur. Örneğin 'A B C D E F', 'B C D N', 'S L B C D' cümleleri için 'B C D' ifadesi en uzun eşleşen parçadır.
- Cevap cümlelerindeki en uzun eşleşen parçadaki soru kelimesi ;<NAME>, cevap kelimesi ise <Answer> ile etiketlenmiştir.

Örneğin doğum yılı soran yukarıdaki ikililer için

- 1.00 <NAME> (<Answer> -)
- 0.87 <NAME> was born on < Answer > , in
- 0.87 <NAME> was born on < Answer >
- 0.86 <NAME> was born on < Answer > in
- 0.86 <NAME> was born on < Answer >

şeklinde birçok şablon elde edilmiştir. Şablonların başlarındaki ağırlık değerleri ise şablonun geçtiği doğru cevabı içeren cümle sayısının, şablonun geçtiği cevap cümlesi sayısına bölümünden elde edilmiştir.

Soru türleri ve bunlara ait şablonların tümüne

http://www.isi.edu/natural-anguage/projects/webclopedia/Taxonomy/taxonomy_toplevel.html

adresinden erişilebilir.

Webclopedia diğer soru cevaplama sistemlerine benzer bir mimariye sahiptir. Aşağıda sistemin modülleri ve kısaca anlatımları verilmiştir.

Soru Ayırıştırma: CONTEX adlı ayırıştırıcı kullanılarak, sorunun yazımsal ve anlamsal analizini yapılmış ve soru türü belirlenmiştir.

Sorgu Formasyonu: WordNet kullanılarak soru çeşitli formlara dönüştürülür.

Doküman Seçimi: MG(Managing Gigabytes) adlı bilgi erişim motoru kullanılarak sorgularla ilgili en yüksek puan alan N sayıda doküman döndürür.

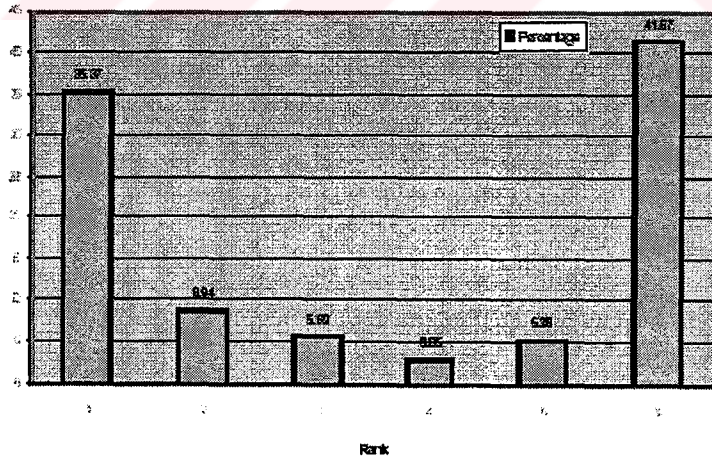
Dokümanlardan Cümlelerin Seçimi ve Sıralanması: Her doküman için, $k \ll N$ olacak şekilde k adet cümle, soru ve onun genişletilmiş sorgu kelimelerini kullanan ve her cümleye bir puan veren bir formülle seçilir. Sonuç cümleleri aldıkları puanlara göre sıralanır.

Cümle Ayırıştırma: CONTEX kullanılarak en yüksek puanlı 300 cümleyi öğelerine ayrılır.

Tam Cevaba Erişim: Her aday cümlelerin ögesel yapısı, sorunun ögesel yapısı ile karşılaştırılarak cümlelerin içinden cevap seçilir.

Cevapların Sıralanması: Aday cevapların bulunduğu cümlelerin puanları karşılaştırılarak en yüksek puanlı ilk 5 cevap kullanıcıya iletilir.

Webclopedia TREC-10 yarışmasına katılmış ve 0.435 MRR puanı almıştır. Yarışmadaki ortalama MRR puanı 0.234'tür. Şekil 2.2'de sistemin cevaplara hangi sırada doğru cevap verdiğinin yüzdeleri gösterilmiştir.



Şekil 2.2 Webclopedia sisteminin değerlendirme sonuçları(Hovy, 2001)

2.4 İnternet üzerinde çalışan soru cevaplama sistemleri

İnternet üzerinde, konu bağımsız çalışan bir soru cevaplama sistemlerine olan ihtiyaç önceki

bölümlerde belirtilmişti. İnternet'teki bilgi artışına paralel olarak bu bilgiye erişecek araçlarında önemi artmıştır. Bu bölümde son yıllarda tasarlanmış olan İnternet üzerinde çalışan soru cevaplama sistemleri anlatılacaktır.

2.4.1 MULDER

MULDER web üzerinde kullanılabilen ilk genel amaçlı ve tam otomatikleştirilmiş soru cevaplama sistemidir(Kwok, 2000). MULDER, kullanıcı tarafından gelen çoklu aramaları otomatik olarak bir arama motoruna(Google) yönlendirmekte ve arama motorunun çıktısını özetlemektedir. MULDER, cevapları elde etmek için, çeşitli araçlar kullanmaktadır.

MULDER'da kullanıcının sorusu öncelikle PC-KIMMO(Antworth 1990) ve MEIParser(Charniak 1999) adlarındaki sözcüksel ayrıştırıcılardan geçirilir. Daha sonra sorunun ağaç yapısı çıkarılır. Sorudaki soru kelimelerine bakılarak sorunun türü belirlenir. Sistemde isim, sayı ve zaman olmak üzere üç temel soru türü vardır. MEIParser kullanılarak sorudaki kelimeler arası ilişkiler belirlenir. Sorunun sınıflandırılmasında WordNet'ten de yararlanılmaktadır. Soru sınıflandırıldıktan sonra üzerinde çeşitli dönüşümler yapılarak birden fazla arama motoru sorgusu elde edilmektedir.

Soru üzerindeki dönüşümler 3 farklı yöntemle yapılmaktadır.

- Fiil dönüşümleri: Örneğin 'When did Nixon visit China?(Nixon Çin'i ne zaman ziyaret etti?)' sorusunun cevabında büyük olasılıkla 'Nixon visited China ...' tarzı bir ifade bulunacağından visit→ visited olarak dönüştürülmüştür.
- Sorgu genişletmesi: WordNet kullanılarak sıfatlara uygun isimler bulunur. Örneğin 'How tall is Mt.Everest?(Everest ne kadar uzundur?)' sorusu için 'tall' sıfatına uygun bir isim olan 'heigth' kelimesi WordNet'ten bulunup sorgu 'the heigth of Everest ...' şekline getirilmiştir.
- Dönüşümsel gramer: İngilizce için geçerli olan Dönüşümsel Gramer Kuralları(Akmajian, 1975) kullanılarak soruların yeni halleri üretilmiştir. Dönüşümsel gramer bir cümle için uygun dönüşümleri ve bu dönüşümlerin anlam üzerindeki etkilerini açıklamaktadır.

Kullanıcı sorusunu çeşitli aşamalardan geçirdikten sonra elde edilen arama motoru sorguları Google'a gönderilmektedir. Google'ın seçilme nedeni olarak içerik olarak diğer arama motorlarından çok daha fazla hacme sahip olması ve site puanlama metodunun (Henzinger,

1999] diğerlerine oranla çok daha başarılı olması gösterilmiştir.

MULDER olası cevapları bulmak için içinde arama motoruna gönderilen sorgulardaki kelimeleri içeren cümleleri belirlemekte ve onlara da puan vermektedir.

Puanlama işlemi için öncelikle bir ön çalışma yapılmıştır. İnternet'ten 100,000 doküman indirilmiş ve bu dokümanlar içindeki kelimelerin ters doküman frekansı(inverse document frequency) Z doküman sayısını, K kelimeyi içeren doküman sayısını belirtmek üzere $w_i = Z/K$ şeklinde hesaplanmıştır. Her bir cümle için puanı ise Eşitlik 2.2 ve 2.3'deki şekilde hesaplanmıştır.

$w_i \rightarrow$ cümlelerdeki kelimeleri;

$n \rightarrow$ cümle içindeki sorgu kelimelerinin sayısını;

$d_1, \dots, d_{n-1} \rightarrow$ n adet sorgu kelimesinin arasındaki kelime bazındaki mesafeyi göstermektedir.

$$D(s) = \frac{\sqrt{d_1^2 + \dots + d_{n-1}^2}}{(n-1)} \quad (2.2)$$

$$S(s) = \frac{\sum_{i=1}^n w_i}{D(s)} \quad (2.3)$$

En yüksek puanı alan cümlelerin içinde cevap türüne uygun olan kelimeler yada kelime grupları sonuç cevabının bulunması için tekrardan puanlanırlar. $k_1 \dots k_n c_1 \dots c_m a_1 \dots a_p$ şeklindeki bir cümledeki a_i cevabının puanı;

$k_i \rightarrow$ sorgu kelimeleri

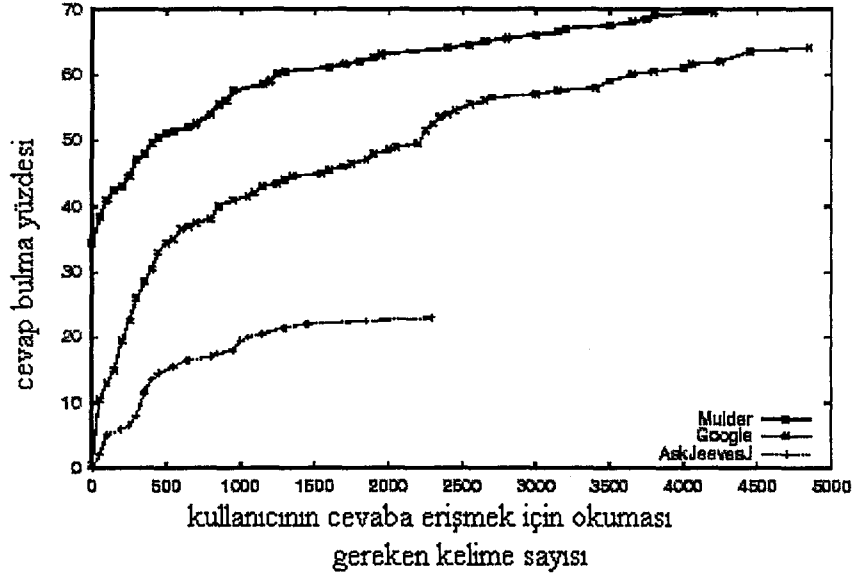
$a_i \rightarrow$ cevap kelimelerini

$c_i \rightarrow$ diğer kelimeleri göstermek üzere eşitlik 2.4'teki formülle bulunur.

$$K_L = \frac{w_1 + \dots + w_n}{m} \quad (2.4)$$

MULDER'ın performansı kullanıcının cevaba erişmek için okuyacağı kelime sayısı ile ölçülmüştür.

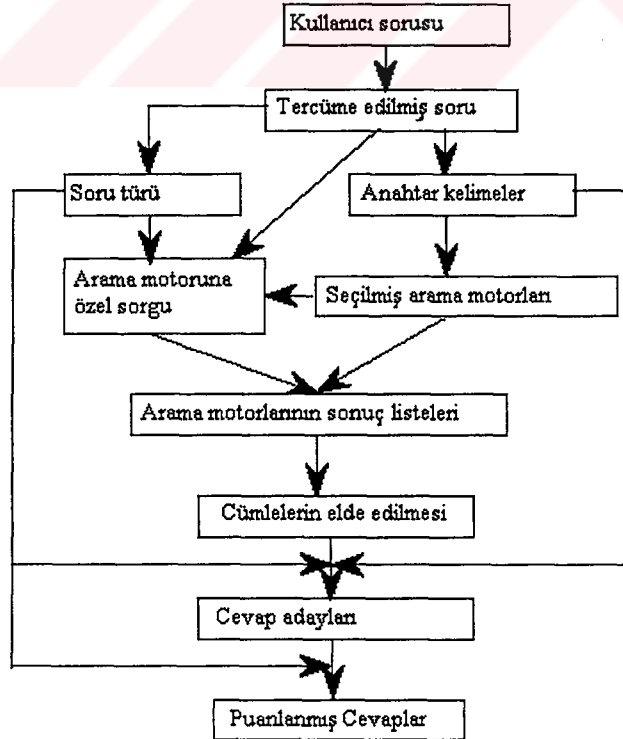
Şekil 2.3'te 200 adet TREC-8 sorusu için MULDER, Google ve benzer bir sistem olan AskJeeves'in karşılıklı değerlendirilmesi verilmiştir.



Şekil 2.3 MULDER'in diğer sistemlerle karşılaştırılması (Kwok, 2000)

2.4.2 AnswerBus

Zhiping Zheng tarafından 2002 yılında gerçekleştirilen sistem kullanıcıdan İngilizce, Almanca, Fransızca, İtalyanca ve Portekizce sorular kabul edebilmektedir. Şekil 2.4'te AnswerBus(www.answerbus.com)'ın genel yapısı verilmiştir.



Şekil 2.4 AnswerBus sisteminin genel yapısı (Zheng, 2002)

AnswerBus ilkönce kullanıcının dilini basit bir modülle bulmakta, eğer İngilizce'den farklı bir dil ise Altavista'nın BabelFish(<http://babel.altavista.com>) adlı çeviri modülünü kullanarak İngilizce'ye çevirmektedir. Daha sonra soru türünü belirlemekte ve her soru türüne göre farklı bir arama motoru seçerek ona uygun sorgular hazırlamaktadır. Örneğin günlük olaylar için YahooNews'un, Google'dan daha iyi bir tercih olacağı belirtilmiştir.

Seçilen arama motorlarına gönderilen sorguya cevap olarak geri dönen sayfalar içerisinde cevabın olması muhtemel cümleleri seçmekte ve puanlamaktadır.

Hangi soruya hangi arama motorunun daha iyi cevap vereceğini bulmak için önceden cevaplanmış 2000 TREC sorusu alınmış ve değişik arama motorlarına gönderilmiştir. Bu sayede arama motorlarının hangi sorulara doğru cevap verdiği bulunmuştur. 2000 sorudaki her kelime indekslenerek kelimelerin geçtiği soruların, kaç tanesine, hangi arama motorundan doğru cevap alındığı çıkarılır.

Örneğin *nasıl* kelimesi için Google 20 doğru cevap bulurken Yahoo 12 doğru cevap bulmuşsa *nasıl* kelimesini içeren sorulara Google'ın daha iyi cevap bulduğu söylenebilir.

Kullanıcı sorusunu arama motoru sorgusuna dönüştürürken 3 temel yöntem kullanılmıştır. Birincisi edat, zamir ve bağlaç tarzı kelimelerin silinmesi, ikincisi sıkça kullanılan kelimelerin silinmesi ve sonuncusu ise kelimeler üzerindeki değişimlerdir. Örneğin

“When did the Jurassic Period end?” sorusundaki “end”, “ended”e dönüştürülmektedir.

Öncelikle dokümanlardaki cümleler bir filtreden geçirilmiş ve eşitlik 2.5'teki formülle elenmiştir.

$$q \geq \lfloor \sqrt{Q-1} \rfloor + 1 \quad (2.5)$$

Eşitlik 2.5'teki formülde q cümledeki sorgu kelimelerinin sayısını; Q ise sorgudaki sorgu kelimelerinin sayısını göstermektedir.

Filtreden geçebilen cümlelerin puanlanmasında temel olarak 3 teknik kullanılmıştır. Birincisi cümleyi soru türüne göre etiketlemektir. Bu yapılırken fiyat soran bir soru için cümlelerdeki fiyat bildiren ifadeler bulunmakta ve buna göre bir puan verilmektedir. Hızın arttırılması için sadece sorunun cevabı olabilecek etiketlerle cümleler etiketlenmektedir. Aksi takdirde gereksiz birçok etiket bulunacaktır.

İkinci yöntem olan zamir çözümleme için sadece birbirini takip eden cümlelerden eğer birinde

zamir varsa diğerine de aldığı puanın bir kısmı verilmektedir. Üçüncü yöntemde ise cevabın arama motoru tarafından kaçınıcı olarak getirildiğine göre puan verilmektedir.

Sonuç olarak AnswerBus, 200 TREC sorusunun %68'ine ilk 5 cevap içinde doğru cevap vermiştir. Ortalama cevap verme süresi ise 7.20 saniyedir. Ortalama cevap uzunluğu 141 karakterden oluşmaktadır.

2.4.3 QuASM

Amerika'daki istatistiksel veritabanlarını sorgulamak için geliştirilmiş bir sistemdir(Pinto, 2000). Sistem www.fedstats.gov adresindeki yaklaşık 170,000 web sitesinde arama yapmaktadır. İşlemin hızlı olması için tüm web siteleri yerel bir veritabanına indirilmiştir. Veritabanında sayfalar orjinal halleriyle değil; sistemce etiketlenmiş halleriyle saklanmaktadır. Her sayfa işlenerek içinde konu bütünlüğü olduğu düşünülen parçalara ayrılmaktadır. Bu yapılırken HTML etiketlerinden yararlanılmıştır. Örneğin kalın yada büyük puntolarla yazılmış kelimelerin başlık belirttiği düşünülerek bir sonraki başlığa kadarki bölüm, başlıktaki konuyla ilgilidir şeklindeki kurallarla konulara ayırma işlemi gerçekleştirilmiştir.

QuASM'nin çalışması diğer soru cevaplama sistemlerine benzemektedir. Kullanıcının girdiği soru incelenerek türü belirlenir. Bu işlem Naive Bayes metodu kullanılarak yapılmıştır. Soru ve türlerinden oluşan bir veritabanında kelimelerin ve kelime ikililerinin hangi tür sorularda bulunduklarının olasılıkları($P(w_i|C)$) ile çıkarılmıştır. Sorunun hangi sınıfa ait olduğu ise eşitlik 2.6'teki formülle bulunmuştur.

$$P(Q | C) = P(w_1 | C) * P(w_2 | C) * ... * P(w_n | C) \quad (2.6)$$

QuASM'nin diğer soru cevaplama sistemlerinden temel farkı tablolardaki bilgiye de erişim özelliği sağlamasıdır. Tablonun başlığını yada başlıklarını tablonun içindeki bilgilerle ilişkilendirerek tablo başlığındaki bir soru sorulduğunda tablonun içindeki bilgileri de kullanıcıya iletebilmektedir. Şekil 2.5(a) ve Şekil 2.5 (b)'te bir metin tablosunun orijinal hali ve QuASM'da saklanan hali gösterilmiştir.

Onions: Area Planted and Harvested by Season, State, and United States, 1996-98									
Season		Area Planted				Area Harvested			
and									
State		1996	1997	1998	1996	1997	1998		
		Acres							
Spring									
AZ	:	2,100	2,100	2,500	1,900	2,100	2,500		
CA	:	10,100	9,900	7,000	9,600	9,600	6,800		
GA	:	16,000	16,200	15,000	14,700	15,800	13,900		
TX	:	15,300	12,400	12,000	13,000	9,800	11,400		
Total	:	43,500	40,600	36,500	39,200	37,300	34,600		

Şekil 2.5(a) Tablonun sayfadaki görünümü

```

<QA_SECTION><QA_METADATA>
<TITLE> Onions: Area Planted and Harvested by Season, State, and
United States, 1996-98 </TITLE>
<CAPTIONS> Season : Area Planted : Area Harvested and :-----
: State : 1996 : 1997 : 1998 : 1996 : 1997 : 1998 :
Acres Spring : </CAPTIONS>
<ROW> AZ :</ROW>
<COLUMN> Area Planted ----- 1997 Acres
</COLUMN>
</QA_METADATA>
<TITLE> Onions: Area Planted and Harvested by Season, State, and
United States, 1996-98 </TITLE>
<CAPTIONS> Season : Area Planted : Area Harvested and :-----
: State : 1996 : 1997 : 1998 : 1996 : 1997 : 1998 :
Acres Spring : </CAPTIONS>
<ROW> AZ :</ROW>
<COLUMN> Area Planted ----- 1997 Acres
</COLUMN>
2,100 </QA_SECTION>

```

Şekil 2.5(b) Tablonun QuASM tarafından saklanması biçimi (Pinto, 2000)

Sistem 177,670 web sayfası içinde cevabı bulunan, sistemi gerçekleştirenlerce belirlenmiş 73 adet istatistiksel soru ile test edilmiş ve işlenmiş sayfalar üzerinde 0.253; işlenmemiş orijinal sayfalar üzerinde 0.209 MRR puanı almıştır. Ortalama cevap verme süresi işlenmiş sayfalarda 8.1 saniye iken orijinal sayfalarda 137.5 saniye olarak ölçülmüştür.

2.4.4 AskMSR

Eric Brill, Susan Dumains ve Michelle Banko tarafından 2002 yılında geliştirilen AskMSR kullanıcının girdiği soruları yeniden biçimlendirerek Google'a göndermekte ve Google'ın sonuç sayfasındaki sayfa özetlerinden cevabı bulmaya çalışmaktadır(Brill, 2002). Soruların yeniden biçimlendirilme kuralları ve ağırlıkları elle elde edilmiştir. Web sitelerinin içeriğinde arama yapmak yerine Google'ın ürettiği sayfa özetinde arama yapmak sistemin hızını büyük ölçüde arttırmaktadır.

Sistemde öncelikle sayfa özetlerindeki birli, ikili ve üçlü n-gramlar çıkarılmaktadır. N-gramların frekansları ve sayfayı getiren sorgunun ağırlığına bakılarak n-gramlar puanlandırılmaktadır.

Elde edilen n-gramlar soru tipine göre bir filtreden geçirilmekte ve daha sonra benzer kelimeleri içeren n-gramlar birleştirilmektedir. Örneğin tarih soran bir sorguya karşılık içerisinde sayı içermeyen n-gramlar silinmektedir. N-gramlar'ın birleştirilmesi işleminde ise örneğin 'A B C' ve 'B C D' üçlü n-gramları birleştirilerek 'A B C D' dördümlü n-gramı elde edilmektedir. Birleştirme sonucu elde edilen n-gramın puanı bileşenlerinin puanlarının toplamı olarak hesaplanmaktadır. Birleştirme işlemi birleştirecek n-gram kalmayıncaya kadar sürdürülmektedir.

Sistemin performansı 500 adet TREC-9 sorusunda ölçülmüş ve 0.507 MRR puanı almıştır.

3. TASARLANAN SİSTEM

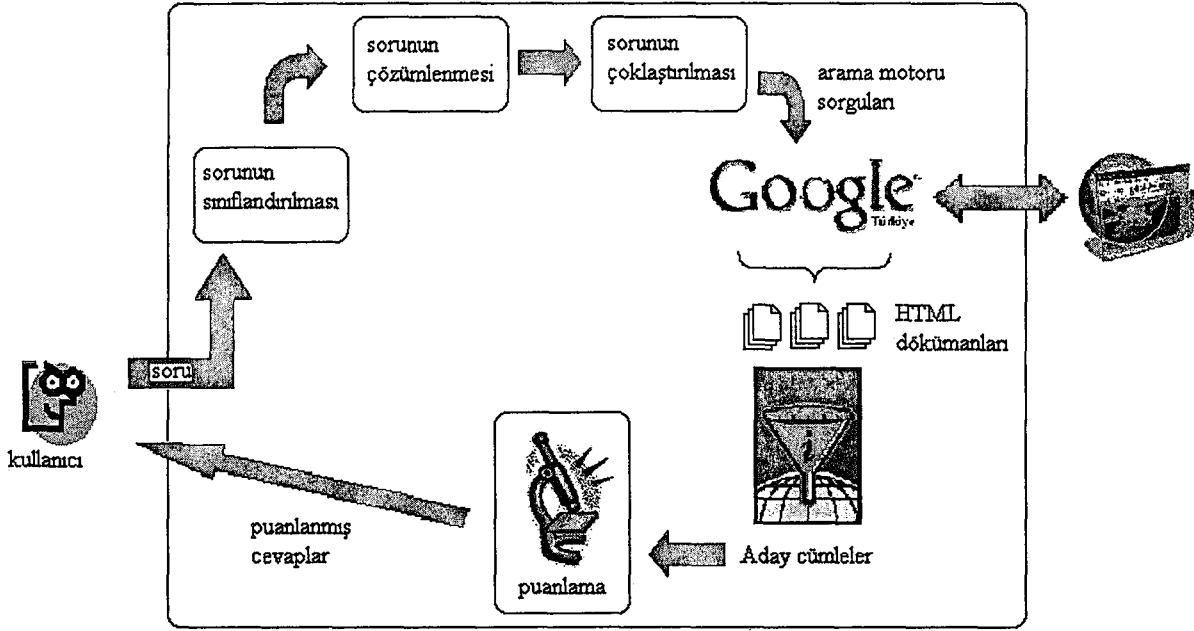
Türkçe için ilk olan bu çalışmada İngilizce için yapılan çalışmalar ayrıntılı olarak incelenmiş ve Türkçe içinde geçerli olabilecek kısımları belirlenmiştir. Birçok sistemin İngilizce için çok sayıda mevcut olan doğal dil işleme araçlarından faydalandığı görülmüştür. Maalesef Türkçe için bu tür araçlar henüz mevcut olmadığından gerekli olan modüller (Türkçe için mevcut olan kelime ayrıştırıcı dışındaki) sistemle birlikte geliştirilmiştir. Ancak diğer sistemlerin kullandıkları bazı modüllerin Türkçe için geliştirilmesi (örneğin en önemli araçlardan biri olan WordNet) başlı başına bir proje olacağından dolayı gerçekleştirilmemiştir.

Geliştirilen sistem her bir modülün performans üzerindeki etkisinin görülebilmesi için istendiği zaman eklenip çıkartılabilecek bir şekilde tasarlanmıştır. Ayrıca sistemin daha sonraki gelişmelere (örneğin Türkçe için bir WordNet) ayak uydurabilmesi için yeni modüller eklenebilir bir yapıda tasarlanmıştır.

Çalışmanın devamında sistemin çalışma akışı çizilmiş ve her bir modülünün ayrıntılı açıklamaları verilmiştir.

3.1 Sisteme Genel Bakış

Tasarlanan sistem diğer soru cevaplama sistemleriyle benzer bir mimariye sahiptir. Sistemin çalışma akışı Şekil 3.1'de gösterilmiştir. Sistem kullanıcının sorduğu soruyu ilk önce sınıflandırmakta ve çözümlemekte; daha sonra çoklaştırarak arama motoru sorgularına dönüştürmektedir. Arama motorundan gelen sonuç sayfalarındaki cümleler bir filtreden geçirilecek aday cümleler belirlenmektedir. Son bölümde de olarak aday cümlelere puan verilerek en yüksek puana sahip ilk 5 cevap kullanıcıya iletilmektedir.



Şekil 3.1 Tasarlanan sistemin çalışma akışı

3.1.1 Sorunun Tipinin Belirlenmesi

Kullanıcının sorduğu sorunun ne tür soru olduğunu anlamak için, içerdiği soru kelimeleri incelenmektedir. Bunun için Çizelge 3.1'deki gibi bir çizelge kullanılmış ve hangi soru kelimelerini içeriyorsa buna karşılık gelen soru türü belirlenmiştir. Sistemde 53 adet soru türü bulunmaktadır.

Çizelge 3.1 Soru kelimelerinden soru türünün belirlenmesinde kullanılan tablodan bir kesit

sorukelime	sorutürü	silinecekkelime
yüzde kaç	oran	kaç
yıl nedir?	yıl	nedir?
yıl kaçtır?	yıl	kaçtır?
yapıtaşları nelerdir?	bileşen	nelerdir?
tarih?	zaman	tarih?
tarih nedir?	zaman	tarih nedir?
tarifi nedir?	nedir	nedir?
şehri hangisidir?	il	hangisidir?
sebebi nedir?	neden	nedir?
sayısı nedir?	sayı	nedir?
nüfusu nedir?	nüfus	nedir?
nüfusu ne kadardır?	nüfus	ne kadardır?
nüfusu kaçtır?	nüfus	kaçtır?

Çizelge 3.1'deki silinecek kelime sütunu, sorudan hangi kelimenin silinerek arama sorgusu yapılacağını gösterir.

3.1.2 Kullanıcı Sorusunun Çözümlemesi

“Bir Türkçe Metindeki Kelimelerin Cümle İçindeki Durumlarına Bakılarak Eklerine Ayrılması” isimli çalışmada(İşler, 2000) geliştirilen ayrıştırıcı kullanılarak kelimelerin gövdeleri, ekleri belirlenmiştir. Ayrıca, ayrıştırıcıya kelimelerin etiketlenmesini sağlayan bir modül eklenmiştir. Etiketler ve örnekleri Çizelge 3.2'de verilmiştir.

Çizelge 3.2 Cümlelerdeki kelimelerin ayrıştırılması ve etiketlenmesi örnekleri

Cümle	Orijinal kelime	Eklerine ayrılmış kelime	Gövde ve eklerin tanımları	Kelimenin görevi
13 Aralık'ta okullar açıldı.	13	13	i[i]	gagun
	Aralık'ta	aralık*da	i[i]*bul	gaay
	okullar	okul*lar	i[i]*ç	
	açıldı	açıl*dı	f[f]*dgz	
29 Mayıs 1453'te okullar açıldı.	29	29	i[i]	ttgun
	Mayıs	mayıs	i[i]	ttay
	1453'te	1453*te	i[i]*blnmyr	ttyil
	okullar	okul*lar	i[i]*ç	
	açıldı	açıl*dı	f[f]*dgz	
29 km yürüdüm.	29	29	i[i]	sayı
	km	km	i[i]	uz
	yürüdüm	yürü*dü*m	f[f]*dgz*set1	
29 km'lik yolu 3 saate aldık.	29	29	i[i]	sayı
	km'lik	km*lik	i[i]*blnmyr	uz
	yolu	yol*u	i[i]*bli	
	3	3	i[i]	sayı
	saate	saat*e	i[i]*yon	zaman
	aldık	al*dı*k	f[f]*dgz*seç1	
Ali üniversiteyi 3 yılda bitirdi.	Ali	ali	ö[i]	özelisim
	üniversiteyi	üniversite*y*i	i[i]*kh*bli	
	3	3	i[i]	sayı
	yılda	yıl*da	i[i]*bul	zaman
	bitirdi	bitir*di	f[f]*dgz	
Ankara ilini gördüm.	Ankara	ankara	ö[i]	özelisim
	ilini	il*i*n*	i[i]*iet3*kh*bli	yer
	gördüm	gör*dü*m	f[f]*dgz*set1	
Bir fil 2500 kilodur.	Bir	bir	sf[sf]	
	fil	fil	i[i]	
	2500	2500	i[i]	sayı
	kilodur	Kilo*dur	i[i]*ef_gez	ag
3 haftada 8 kilo verdim.	3	3	i[i]	sayı
	haftada	hafta*da	i[i]*bul	zaman
	8	8	i[i]	sayı

	kilo	kilo	i[i]	ag
	verdim	ver*di*m	f[f]*dgz*set1	
Nemrut Dağı 23 asır önce yandı.	Nemrut	nemrut	ö[i]	özelisim
	Dağı	dağ*ı	i[i]*bli	yer
	23	23	i[i]	sayı
	asır	asır	i[i]	zaman
	önce	önce	i[i]	
	yandı	yan*dı	i[i]*ef_dgz	
30 metreden atlayan Osman ayağını kırdı.	30	30	i[i]	sayı
	metreden	metre*den	i[i]*ayr	uz
	atlayan	atla*y*an	f[f]*kh*sfe1	
	Osman	osman	ö[i]	özelisim
	ayağını	ayak*ı*n*ı	i[i]*iet3*kh*bli	
	kırdı	kır*dı	i[i]*ef_dgz	

Çizelge 3.2’de verilen örneklerden ‘13 Aralık’ta okullar açıldı.’ ifadesindeki ‘13’; ‘13 Aralık’ gün-ay tamlamasının günü belirten kısmı, ‘Aralık’; tamlamanın ayı belirten kısmıdır. Bu görevleri kelimenin görevleri kısmında ‘gagun’ ve ‘gaay’ ile ifade edilmiştir. ‘okullar’ kelimesi ise *okul* isminden sonra çoğul eki almıştır ve çözümlenmiş halinde ‘i[i]*ç’ ile gösterilmiştir. ‘açıldı’ kelimesi; *açıl* fiiline di’li geçmiş zaman eki gelmesiyle oluşmuştur. ‘29 km’lik yolu 3 saatte aldık.’ ifadesindeki ‘29 km’lik’ tamlamasındaki 29 sayı; *km’lik* uzunluk(uz) olarak belirlenmiştir. Yine ‘3 saatte’ ifadesindeki 3 sayı, *saatte* ifadesi ise zaman olarak belirlenmiştir. ‘2500 kilodur.’ ifadesinde 2500 sayı; *kilodur* ağırlık(ag) olarak belirlenmiştir.

Çizelge 3.3’te yer, zaman, ağırlık ve uzunluk etiketlerinin belirlenmesinde kullanılan anahtar kelimeler verilmiştir.

Çizelge 3.3 Kelimelerin etiketlemesinde kullanılan anahtar kelimelere örnekler

[yer]	[zaman]	[ag]	[uz]
ili	dakika	gram	metre
barajı	saat	kilo	km
dağı	gün	kg	kilometre
gölü	hafta	kilogram	santim
şehri	ay	ton	cm
yöresi	yıl	litre	santimetre
bölgesi	asır		mil
kenti			

3.1.3 Kullanıcı Sorusunun Çoklaştırılması

Kullanıcının sorduğu sorudan soru kelimeleri atılarak ve çekim eklerinden ayrılarak iki sorgu

oluşturulmuştur. Çizelge 3.4'te bu işleme örnekler verilmiştir.

Çizelge 3.4 Kullanıcı sorusunun eklerinden arındırılması örnekleri

Kullanıcının Sorusu	1. sorgu Soru kelimeleri atılmış hali	2. Sorgu Eklerinden arındırılmış hali
Asya ile Avrupa'yı ayıran boğazın adı nedir?	Asya ile Avrupa ayıran boğazın adı	Asya ile Avrupa ayır boğaz ad
enzimlerin vücuttaki görevi nedir?	enzimlerin vücuttaki görevi	enzim vücut görev
Deli Dana hastalığının belirtileri nelerdir?	Deli Dana hastalığının belirtileri	Deli Dana hastalık belirti
dünyanın en uzun boylu insanı kimdir?	dünyanın en uzun boylu insanı	dünya en uzun boy insan
su hangi atomlardan oluşur?	su atomlardan oluşur	su atom oluş

Çizelge 3.4'teki kullanıcı soruları kısmındaki soruların soru kelimeleri çıkarılmış halidir.

3.1.4 Arama Motorundan Sonuçların Alınması

Kullanıcı sorusundan elde edilen iki sorgu arama motoruna gönderilmiştir. Aday cevap cümlelerin elde edilmesinde, iki farklı yol kullanılmıştır. Birincisi sadece arama motorunun sonuç sayfasındaki sayfa özetlerindeki cümleleri incelemek; ikincisi ise sonuç sayfasındaki bağlantıların gösterdiği sayfaları incelemektir. Her iki yaklaşımın avantaj ve dezavantajları Çizelge 3.5'te verilmiştir.

Çizelge 3.5 Cevapların nerede aranacağına dair karşılaştırma

Cevabın nerede aranacağı	Avantajları	Dezavantajları
Sadece sayfa özetlerinde arama	Hızlı	Cümleler kesik ve anlamsız dolayısıyla daha az doğru sonuçlar (%43.9)
Sayfaların kendilerinde arama	Cümleler tam ve anlamlı ve dolayısıyla daha doğru cevaplar (%43.9 + %16.6)	Çok yavaş

Sayfaların kendilerinden cevap üretilirken sayfaların orijinal hallerinin yerine Google'ın önbelleğindeki halleri kullanılmıştır. Bunun iki önemli avantajı vardır. Birincisi sayfalardaki bilgilerin indirilmesi daha az zaman alacaktır; ikincisi ise sayfanın arama yapılırkenki (belki

değiştirilmiş yada yayından kaldırılmış) içeriği yerine Google'da indekslendiği hali kullanılarak daha doğru bilgiye erişim sağlanacaktır.

3.1.5 Sonuç sayfalarından aday cümlelerin seçilmesi

Sitelerdeki cümlelerin aday cümleler olabilmesi için bazı filtrelemeler yapılmıştır.

Bir cümlelerin cevap cümlelerinden olabilmesi için aşağıdaki şartları sağlaması gerekmektedir.

- 1- İçinde soru kelimelerinin sayısının en az yarısı kadar soru kelimesi içermeli
- 2- Sorgudaki özel isim olarak etiketlenmiş en son kelimeyi içermeli
- 3- Soru türü cevapta sayısal bir değer beklenen türdensen(yıl, mesafe, oran, ağırlık vb.) içinde sayı bulundurmamalıdır.

Yukarıdaki her üç şarta uygun cümleler aday cevap cümleleri olarak belirlenmiş ve puanlama modülüne gönderilmiştir. Puanlama işlemi zaman alıcı bir işlem olduğundan; filtreleme sayesinde sistemin cevap verme süresi kısaltılmıştır.

Çizelge 3.6'da "Atatürk hangi yıl dünyaya geldi?" sorusu için aday cümlelerin seçiminden örnekler verilmiştir.

Çizelge 3.6 Sonuç sayfalarından aday cümlelerin seçilmesi örnekleri

Cümle	Aday cümle mi	Sebebi
6 Eylül 1944, Pink Floyd grubunun üyelerinden Roger Waters dünyaya geldi.	Hayır	2. Şarta uymuyor.
9 Ekim 1940, Efsane Rock grubunun solisti John Lennon dünyaya geldi.	Hayır	2. Şarta uymuyor.
Kemal Atatürk,1881(Rumi 1296) yılında Selanik'te Koca Kasım Paşa Mahallesi İslahane Caddesi'nde bugün müze olan üç katlı bir evde dünyaya geldi.	Evet	Tüm şartları sağlıyor
1964 yılında İstanbul'da dünyaya geldi.	Hayır	2. Şarta uymuyor.
Muhtar Cem Karaca 5 Nisan 1945 de İstanbul da dünyaya geldi.	Hayır	2. Şarta uymuyor.
ÖMER ÇAKAR 19.02.1950 yılında Sivas'ın Zara ilçesinde dünyaya geldi.	Hayır	2. Şarta uymuyor.
Atatürk'ün babası Ali Rıza Efendi de bu kocacık nahiyesinde dünyaya geldi.	Hayır	3. Şarta uymuyor.
Nilüfer Hatun Ortaokulu ve Taksim Atatürk Lisesi ni bitirdi.	Hayır	1. Şarta uymuyor.
Atatürk'ün babası Ali Rıza Efendi'de bu nahiye de dünyaya geldi.	Hayır	3. Şarta uymuyor.
1881 - Selanik' te dünyaya geldi.	Hayır	2. Şarta uymuyor.

3.1.6 Aday cümlelerin puanlanması

Her cümle için eşleştirme, sıralama ve şablon puanları olmak üzere 3 farklı puan hesaplanmaktadır.

- **Eşleştirme puanının belirlenmesi:** Cümlelerin içerdiği sorgu kelimesi sayısı, o cümlelerin eşleştirme puanı olarak belirlenir. Çizelge 3.7’de “**Boğaz Köprüsü’nün yüksekliği ne kadardır?**” sorusu için aday cümlelerin eşleştirme puanları verilmiştir.

Çizelge 3.7 Aday cümlelerin eşleştirme puanlarından örnekler

Cevap cümlesi	Eşleştirme puanı
Boğaz Köprüsü'nün yüksekliğinin yaklaşık 3 katına eşdeğerdir .	3
64 metre yüksekliğindeki Boğaz Köprüsü'yle, Saipem 7000 .	3
Boğaz Köprüsü'nün denizden yüksekliği ise 60 metre.	3
Mesela Boğaziçi Köprüsü'nün deniz seviyesinden yüksekliği sadece 64 .	3
boyu 165 metre olan Birinci Boğaz Köprüsü'nün .	2

- **Sıralama puanının hesaplanması:** Cümle içindeki sorgu kelimelerinin birbirine yakınlığı, cevabın o cümlede olma olasılığını arttırmaktadır. Bu yüzden sorgudaki kelimelerinin sayısına, cümledeki sorgu kelimelerinin sayısına ve yakınlığına bağlı bir formülle 0-10 arasında bir sıralama puanı hesaplanmıştır. Hesaplama formülü Eşitlik 3.1’de verilmiştir.

$$SP = \left\lfloor \frac{10 * C_k * (C_k - 1)}{Tp * S_k} \right\rfloor \quad (3.1)$$

Eşitlik 3.1’deki C_k cümlelerin içerdiği sorgu kelimesi sayısını, S_k sorgudaki kelime sayısını göstermektedir. Tp ise cümledeki sorgu kelimelerinin birbirlerine kelime bazındaki uzaklıklarının toplamıdır.

Örneğin "A B C D" sorgu cümlesi için "A E F G H B C" cevap cümlesinin sıralama puanı bulunurken öncelikle cevap cümlesindeki sorgu kelimelerinin cümledeki kaçınıcı kelime oldukları bulunmuştur. Örnek için, 1 – 6 – 7 olarak bulunur. Bu dizi sıralı değilse ise

sıralanmaktadır. Sonuç olarak eşitlikteki T_p ifadesi; bu dizideki her rakamın bir sonrakinden farkı bulunduktan sonra bu farkların mutlak değerleri toplanarak elde edilmiştir. Örnek için $T_p = |(1-6)| + |(6-7)| = 6$ olarak bulunur. Örneğin sıralama puanı ise $SP = [10*3*(3-1)]/(6*4) = 2.5$ dir.

Bir başka örnek vermek gerekirse; "A B" sorgu cümlesi için "A B C" cevap cümlesinin sıralama puanı $SP = [10*2*(2-1)]/(1*2) = 10$ olarak bulunmuştur.

Çizelge 3.8'de “Boğaz Köprüsü’nün yüksekliği ne kadardır?” sorusu için aday cümlelerin sıralama puanları verilmiştir.

Çizelge 3.8 Aday cümlelerin sıralama puanlarından örnekler

Cevap cümlesi	Sıralama puanı
64 metre yüksekliğindeki Boğaz Köprüsü’yle, Saipem 7000 arasında yaklaşık 8 metre mesafe kalacak.	10
Fatih Sultan Mehmet Köprüsü’nün güvenli geçiş yüksekliği 64 metre olduğu için bu köprünün altından geçişte bir sorun görülüyor.	2
Gemi, denizden 64 metre yüksekliğinde olan Boğaziçi Köprüsü’nün altından yaklaşık 10 dakikada geçişini tamamladı.	6

- **Şablon puanının belirlenmesi:** Kullanıcının girdiği soru türüne ait şablonlar veritabanından çekildikten sonra cümlede bu şablonların olup olmadığı araştırılır. Eğer varsa o var olan şablona ait ağırlık değeri cümlelerin şablon puanı olarak belirlenir.

Örnek olarak “plastik neden elde edilir? ” sorusunun türü “bileşen” olarak belirlendikten sonra; şablonlar veritabanından “bileşen” sorusuna cevap olabilecek şablonlar seçilmiştir. Çizelge 3.9'da bu şablonlar verilmiştir.

Çizelge 3.9 Bileşen sorularında aranan şablonların listesi

Şablonlar		
sorutipi	cevaptipi	ağırlık
bileşen	orj=ana%maddesi	5
bileşen	orj=elde%edil	6
bileşen	orj=hammadde	5
bileşen	orj=içerir	5
bileşen	orj=içinde	5
bileşen	orj=kullanılarak	5
bileşen	orj=yapılır	7
bileşen	tnm=ayr#orj=elde%edil	9
bileşen	tnm=ayr#orj=oluş	7
bileşen	tnm=ayr#orj=üretilir	8
bileşen	tnm=ayr#orj=yapılır	9

Cevap cümlesinin içinde Çizelge 3.9'daki yapılardan biri yer alıyorsa; yer alan şablonun ağırlık değeri cümlelerin şablon puanı olarak belirlenmiştir.

Üç farklı şekilde puanlandırılan soruların aldıkları puanlar toplanarak o cümlelerin puanı elde edilmiştir.

Çizelge 3.10'da örnek aday cümleler, cevabı oldukları sorularındaki anahtar kelimeler, soru türleri ve aldıkları puanlar verilmiştir.

Çizelge 3.10 Aday cümlelerin aldıkları puanlar

Aday cümle	Sorudaki anahtar kelimeler	Soru türü	Eşleştirme	Sıralama	Şablon
yeni GSM telefonu M50'yi ilk kez tüketicilerle buluşturdu .	telefon ilk kez kullanıldı	zaman	3	5	0
"@", e-mail adresinin bir parçası olarak ilk kez 1971 yılında Ray Tomlinson tarafından kullanıldı.	telefon ilk kez kullanıldı	zaman	3	2	10
Probleme dayalı öğrenme (PDÖ) kavramı ilk kez 1976 yılında Kanada McMaster Üniversitesi'nde Howard Borrows tarafından kullanıldı.	telefon ilk kez kullanıldı	zaman	3	1	10
Amerika, 1867'de, Alaska'yı Rusya'dan 7.2 milyon \$ altına satın aldı.	Amerika Alaska aldı	zaman	3	2	8
Amerika Alaska'yı da, 1867'de Ruslar'dan satın aldı.	Amerika Alaska aldı	zaman	3	3	8

Amerika'nın Bağımsızlık Günü'nden banane; 4 Temmuz, oğlum Berent'in doğumgünü :) Geçtiğimiz günlerde, Berent, doğumgünü'nü Paris Purple .	Amerika bağımsızlık günü	zaman	2	0	5
Eylül 2001 Çarşamba Bilindiği üzere 11 Eylül 2001 Salı günü Amerika'da çok .	Amerika bağımsızlık günü	zaman	2	6	5
4'ün Today (Bugün) adlı programında Amerika'nın bağımsızlık .	Amerika bağımsızlık günü	zaman	2	6	0
Hürriyet, Amerika'daki Bağımsızlık günü kutlamalarına ilişkin 4 Temmuz tarihli haberinde .	Amerika bağımsızlık günü	zaman	2	3	5
1 Kasım 1927, TBMM'nin üçüncü dönemi açıldı.	TBMM açıldı	zaman	2	3	8
Teklif, bugün (24 Şubat 2000) koalisyon partileri genel başkanlarının düzenledikleri zirveden hemen sonra TBMM'de imzaya açıldı.	TBMM açıldı	zaman	2	5	8
Sonuçta 23 Nisan 1920 Tarihinde Ankara'da TBMM açıldı ve Türk .	TBMM açıldı	zaman	2	10	8
TBMM Genel Kurulu saat 15.00'te açıldı.	TBMM açıldı	zaman	2	2	0
25 Aralık'ta Gaziantep'in kurtuluş gününde açacağımız sergiye gelecek insanlardan böyle bir grup çıkarabileceğimizi ve gelecek sene beş .	Gaziantep kurtuluş	zaman	2	10	5
Şöyle ki, bugün, 25 Aralık 1999, Gaziantep İlimizin kurtuluş günüdür.	Gaziantep kurtuluş	zaman	2	5	10
Körfez Başkent'te vuruldu:2-0 Ligden çok kupayı düşünen Kocaelispor, Ankaragücü deplasmanında Kennedy'nin gollerine boyun eğdi Ligde UEFA Kupası .	Kennedy vuruldu	zaman	2	1	0
O zaman hoppala diyeceksiniz; Kennedy 1963'de vuruldu biliyorsunuz, bunların çıkması ise 1980'den sonra.	Kennedy vuruldu	zaman	2	5	8
Senatör Kennedy vuruldu (6 Haziran 1968) Eski ABD Başkanı JF.	Kennedy vuruldu	zaman	2	10	8
Sadri Alışık Sadri Alışık 5 Mart 1925 yılında İstanbul'da doğdu.	Sadri Alışık doğdu	dogum zaman	3	2	10
Sadri Alışık, 5 Mart 1925, 18 Mart 1995.	Sadri Alışık doğdu	dogum zaman	2	6	0
Sadri Alışık 5 Mart 1925 yılında İstanbul'da doğdu.	Sadri Alışık doğdu	dogum zaman	3	2	10
Sadri Alışık 1925 yılında İstanbul'da doğdu.	Sadri Alışık doğdu	dogum zaman	3	4	0
Kerem Alışık, 1963, İstanbul Sadri Alışık ve Çolpan İhan'ın oğlu olan Kerem Alışık, 1963 yılında İstanbul'da doğdu.	Sadri Alışık doğdu	dogum zaman	3	1	0
Türkiye Cumhuriyeti'nin ilk nüfusu 12.4 milyon olarak kabul edildi.	Türkiye ilk nüfus sayımı yapıldı	zaman	3	4	0
Türkiye'nin ilk nüfus sayımı ise 1927 yılında yapıldı.	Türkiye ilk nüfus sayımı yapıldı	zaman	5	5	10
28 Ekim 1927, Türkiye'de birinci genel nüfus sayımı yapıldı.	Türkiye ilk nüfus sayımı yapıldı	zaman	4	4	8
28 Ekim 1927 Türkiye'de ilk nüfus sayımı yapıldı.	Türkiye ilk nüfus sayımı yapıldı	zaman	5	10	8
28 Ekim 1927, Türkiye'de ilk genel nüfus sayımı yapıldı.	Türkiye ilk nüfus sayımı yapıldı	zaman	5	8	8
28 Ekim 1927, Türkiye Cumhuriyeti'nde ilk nüfus sayımı yapıldı.	Türkiye ilk nüfus sayımı yapıldı	zaman	5	8	8
ALPMAN Avrupa Birliği'nin kapısına dayanmış çağdaş Türkiye'nin 2000'li yıllardaki ilk modern (parmakla) nüfus sayımı için İstanbul'un .	Türkiye ilk nüfus sayımı yapıldı	zaman	4	3	8

Babalar Günü, 1914 yılında kutlanmaya başlayan Anneler Günü'nden 61 yıl sonra ilk kez Amerika'nın Spokane kentinde kutlandı.	Babalar Günü ilk kez kutlandı	zaman	5	2	10
Babalar Günü ilk kez 19 Haziran 1910'da Washington'ın .	Babalar Günü ilk kez kutlandı	zaman	4	8	8
Dünya Babalar günü, ilk kez ABD'de kutlandı (1910).	Babalar Günü ilk kez kutlandı	zaman	4	4	10
Anneler Günü resmi olarak ise ilk kez Amerika Birleşik Devletleri'nde 1872 yılında kutlandı.	Babalar Günü ilk kez kutlandı	zaman	4	2	10
geçen Fırat-Dicle veya 9 ülkenin ihtiyacını karşılayan Nil Nehri'nin su .	Nil Nehri uzunluğu	mesafe	2	6	0
İnşaat sektöründeki en son gelişme Aralık 2002'de Aswan'da Nil Nehri üzerinde ilk defa olarak hizmete giren 1km uzunluğundaki asma köprü inşası .	Nil Nehri uzunluğu	mesafe	3	2	0
Nil Nehri, Kahire'nin 25 Km .	Nil Nehri uzunluğu	mesafe	2	6	0
Nil nehri Mısırın tek hayat kaynağı olup uzunluğu 6500 km'dir.	Nil Nehri uzunluğu	mesafe	2	0	0
Bu uzak gökadalardan gelen ışık saniyede 300.000 kilometre hızla yol alır, gene de Dünya'ya ulaşması milyarlarca yıl sürer.	ışık saniyede yol alır	hız	4	5	8
Güneşten gelen ışık saniyede 300.000 km gibi müthiş bir hızla yol alarak dünyaya .	ışık saniyede yol alır	hız	3	1	8
Çünkü ışık saniyede 186.000 mil yol alır.	ışık saniyede yol alır	hız	4	6	8
Aslında 15 C 'lik tatlı suda ses, yaklaşık saniyede 1410 m yol alır.	ışık saniyede yol alır	hız	3	3	0
İnsan vücudunun yüzde 66'sı, insan beyninin yüzde 75'i sudur.	insan vücudunun sudur	oran	3	2	8
Büyüklerde de, 1.77 boyunda ve 65 kilo ağırlığında bir insan vücudunun ortalama yüzde 33 kilosu sudur! .	insan vücudunun sudur	oran	3	3	8
Erişkin bir insan vücudunun %55-75'ini su oluşturur.	insan vücudunun sudur	oran	2	6	0
Unutmayın insan vücudunun %60-70'i sudur.	insan vücudunun sudur	oran	3	6	0
kan:kan şekeri düşer,kanın vizkozitesi azalır,dokulara yeterince su ulaşamaz,insan vücudunun %71 i sudur (zebellah, 26.11.2000 16:48).	insan vücudunun sudur	oran	3	5	0
Eğer insan vücudunun en küçük kılcal damarlarının 10 bin tanesini yan yana getirirsek .	insan vücudunun sudur	oran	2	6	0
Yetişkin bir insanın vücut ağırlığının yaklaşık % 60'ı sudur.	insan vücudunun sudur	oran	2	1	0
Yetişkin insan vücudunun ortalama % 59 sudur.	insan vücudunun sudur	oran	3	4	0
12 Nisan 1961'de ilk defa uzaya insanlı bir roket fırlatıldı.	uzaya çıkan ilk insan	kim	3	5	0
Vostok-1 adlı roketle birlikte uzaya çıkan bu ilk insan Rus Yuri Gagarin idi.	uzaya çıkan ilk insan	kim	4	7	6
Halbuki Dennis Tito, (uzaya çıkan ilk yaşlı misafir astronot) 61 yaşında .	uzaya çıkan ilk insan	kim	3	7	6
SSCB'li Yuri Gagarin uzaya çıkan ilk insan oldu.	uzaya çıkan ilk insan	kim	4	10	6
yıldönümünü kutlayan Rusya'da, uzaya çıkan ilk insan Yuri Gagarin'i anma törenlerine gölge düştü.	uzaya çıkan ilk insan	kim	4	10	6

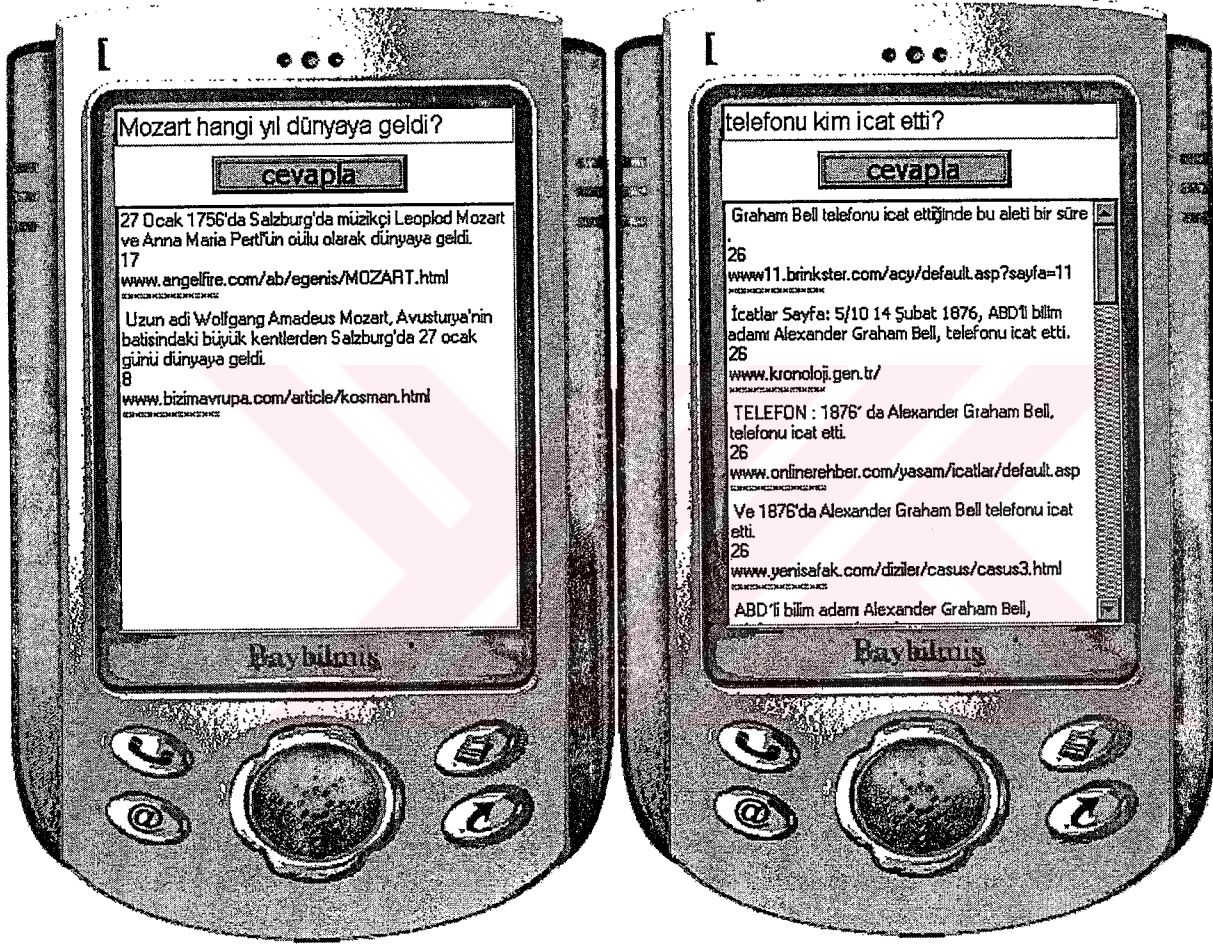
Çiller 1993'de SHP ile Hükümeti yeniden kurdu, Türkiye'nin ilk kadın Başbakanı oldu.	Türkiye ilk kadın başbakanı	kim	3	7	6
Demirel'in cumhurbaşkanı seçilmesinin ardından genel başkan seçilen Tansu Çiller, Türkiye'nin ilk kadın başbakanı unvanını kazanıyordu 1993 .	Türkiye ilk kadın başbakanı	kim	4	10	6
Adliyede kadın sesi Gürel ERGEN / Bursa Türkiye'nin 'ilk kadın mübaşiri' 23 yaşındaki Şenay Gedük, Bursa'nın Karacabey İlçesi Adliyesi .	Türkiye ilk kadın başbakanı	kim	3	2	6
Necmettin Erbakan'dan, Özal'ın ANAP'ını dönüştürüp kendi damgasını vuran Mesut Yılmaz'a; Türkiye'nin ilk kadın başbakanı Tansu Çiller .	Türkiye ilk kadın başbakanı	kim	4	10	6
'Türkiye'nin ilk kadın başbakanı' sıfatını alarak gücüne .	Türkiye ilk kadın başbakanı	kim	4	10	6
Türkiye'nin ilk kadın Başbakanı olan Çiller'in kurduğu yeni DYP-SHP Koalisyon Hükümeti 25 Haziran 1993'ten 25 Aralık 1995 seçimlerine kadar görevde .	Türkiye ilk kadın başbakanı	kim	3	7	6



3.1.7 Cevapların kullanıcıya verilmesi

Puanlanmış cümlelerden, birbirinden farklı olan ve en yüksek puanı almış ilk 5 cümle kullanıcıya cevap olarak iletilmektedir.

Şekil 3.2'de sistemin ara yüzü ve verdiği cevap örnekleri gösterilmiştir.



Şekil 3.2 Sistemin arabirimi ve verdiği cevap örnekleri

3.2 Sistemde kullanılan araçların ayrıntılı incelenmesi

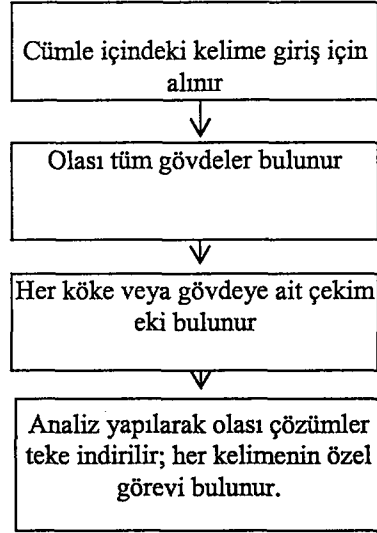
Sistemde kullanılan modüllerin (ayrıştırıcı hariç) hepsi sistemle birlikte geliştirilmiştir.

3.2.1 Kelimeleri gövde ve eklerine ayıran ayrıştırıcı

Cümlelerdeki kelimelerin gövdelerinin ve aldıkları eklerin belirlenmesinde Seyhan İşler tarafından 2000 yılında geliştirilen (İşler, 2000) ayrıştırıcı kullanılmıştır.

Kelimenin cümle içindeki durumuna bakılarak gövde ve eklerine ayrılması; doğal dille yazılmış metinler üzerinde çalışan bir sistemin önemli bir bileşenidir. Gövdeleme algoritmalarında temel olarak iki yöntem izlenmektedir. İlk yöntemde, kelime soldan sağa doğru harf harf eklenerek veri tabanında aranmaktadır. Diğer yöntem ise kelimenin sağdan sola doğru harf harf eksiltilecek incelenmesidir. Bu çalışmada, daha iyi sonuç veren ilk yöntem kullanılmıştır.

Sistemde, ilk olarak kelimenin olası tüm gövdeleri bulunmaktadır. Örneğin; “ağacı” kelimesi için “ağ”, “ağa” ve “ağaç” kelimeleri gövde olarak bulunmaktadır. Daha sonra ise çekim eklerinin varlığı araştırılmaktadır. Eğer gövde isim soylu ise kelimenin içerdiği isim çekim ekleri, şayet fiil soylu ise çekim ekleri ve filimsi ekleri araştırılır. Türkçe’deki eklerin birbirini içermesinden dolayı (örnek: -m, -miz gibi), bu işlemler sonucunda birden fazla muhtemel çözüm bulunmaktadır. Bu çözümlerden bir kısmı, kelimenin tamamını içermediğinden dolayı zaten elenir. Geriye kalan olası çözümler içerisinde doğru çözümü bulurken ise kelimenin cümle içerisindeki durumuna bakılmaktadır. Örneğin, “evi” kelimesi için “ev” gövde olarak ve “-i” eki ise ya hal eki yada iyelik eki olarak bulunur. Eğer cümlede “Ayşe’nin evi” benzeri bir tamlama varsa bunun iyelik eki olduğu anlaşılır. “Şu evi beğendim” cümlesinde ise “-i” ekinin hal eki olduğu anlaşılır. En son olarak da varsa her kelimenin cümle içerisindeki özel görevi (sayı, tarih, mesafe..) bulunur. Kelimenin gövde ve eklerine ayrılmasında izlenen adımlar Şekil 3.3’te gösterilmiştir.



Şekil 3.3 Sistemin kelimeleri gövde ve eklerine ayırmasında izlediği yollar

Çizelge 3.11’de kelimelerin gövde ve eklerine ayrılmasından çeşitli örnekler gösterilmiştir.

Çizelge 3.11 Kelimelerin çözümlenme örnekleri

Kelime	Kelimenin çözümlenmesi	Çözümlemedeki ek tanımları
Ağaçlaştıramadıklarımızdan	ağaç*laş*tır*a*ma*dık*lar*ı*mız*dan	İsim*FYE*FYE*yt*ol*SFE*ç*ys*ieç1*ayr
10 Mayıs 2002’de	10 Mayıs 2002*de	İsim isim isim*bul (Görevleri : gun ay yıl)
Kırılğan	kır*ı*1*gan	fiil*ys*FYE*İYE
Güzeldim	güzel*di*m	İsim*ef_dgz*şet1
Yorucu	yor*ucu	fiil*İYE
Ağaca	ağaç*a	İsim*yon
cesurdur	cesur*dur	İsim*ef_gez
isteyebileceklerinden	iste*y*ebil*ecek*leri*n*den	fiil*kh*yt*SFE*ieç3*ku*ayr
gitmiştiler	git*miş*di*ler	fiil*mgz*dgz*şet3
gelemeyebilirim	gel*e*me*y*ebil*i*r*im	fiil*yt*ol*kh*yt*ys*gez*şet1
gittiydim	git*di*y*di*m	fiil*dgz*ku*dgz*şet1
gidiyorken	git*i*yor*ken	fiil*ys*şz*ken
ağlayıncaya	ağla*y*ıncaya	fiil*kh*ZFE

3.2.2 Şablonlar Veritabanı

Çizelge 3.12’de görüleceği gibi şablonlar veritabanında hangi soru türünün şablonu oldukları, şablonların kendileri ve ağırlık değerleri bulunmaktadır.

Sistemde şablonlar üç türlü bulunmaktadır. Birincisi kelimelerin orijinal hallerini içeren şablonlar; ikincisi kelimelerin ayrıştırılmış hallerini içeren şablonlar; üçüncü ise kelimelerin görevlerini içeren şablonlardır.

Bir sorunun cevap şablonlarında yukarıda bahsedilmiş olan üç tür şablonda yer alabilir. Çizelge 3.12’de şablonlar veritabanından bir kesit verilmiştir.

Çizelge 3.12 Şablonlar veritabanından bir kesit

Şablonlar		
sorutipi	Cevaptipi	ağırlık
eser	tnm=ö[i]#orj=eseridir	8
eser	tnm=ö[i]#orj=tarafından	8
eser	tnm=ö[i]#orj=yapıtlarından	8
eser	tnm=ö[i]#orj=yaptı	7
eser	tnm=ö[i]*iet2#orj=başyapıtı	8
eser	tnm=ö[i]*iet2#orj=eseri	8
eser	tnm=ö[i]*iet2%ö[i]*iet3	6
eser	tnm=ö[i]*tm#orj=başyapıtı	10
eser	tnm=ö[i]*tm#orj=eseri	10
eser	tnm=ö[i]*tm%ö[i]*iet3	8
eser	tnm=ö[i]*yon#orj=yaptır	9
hız	grv=hız	9
hız	orj=saatte#grv=uz	7
hız	orj=saniyede#grv=uz	8
neden	orj=bu%yüzden	5
neden	orj=çünkü	6
neden	orj=dolayısıyla	7
neden	orj=nedeni	6
neden	orj=olduğundan	6
neden	orj=sebebi	6
neden	orj=sebepe%olur	7
zaman	grv=gagun	5

Şablonlar		
sorutipi	Cevaptipi	ağırlık
zaman	grv=gagun#tnm=*de	6
zaman	grv=tyil#tnm=*de	5
zaman	grv=yil#orj=günü	10
zaman	grv=yil#orj=tarihinde	10
zaman	grv=yil#orj=tarihli	9
zaman	grv=yil#orj=yılında	10
zaman	grv=yil%	8
icat	tnm=ö[i]*bli#orj=icat%eden	8
icat	tnm=ö[i]*tm#orj=geliştirdiği	8
icat	tnm=ö[i]*tm#orj=geliştirdiği#tnm=ö[i]	8
icat	tnm=ö[i]*tm#orj=icadı	10
icat	tnm=ö[i]*tm#tnm=i[i]*iet3#orj=icat	10
il	grv=özelisim%yer	9
il	grv=özelisim%yer#orj=sınırları%iç	10
il	grv=özelisim%yer%özelisim%yer	8
il	grv=yer#orj=sınırları%içinde%bulunan	10
il	grv=yer#orj=sınırları%içinde%yer%alan	10
il	tnm=ö[i]*tm#grv=yer#orj=ilçesine%bağ	8

Çizelge 3.12'de yer alan şablonlardaki 'orj=' ifadesi kelimelerin orijinal hallerini, 'grv=' ifadesi kelimelerin görevlerini, 'tnm=' ifadesi kelimelerin gövde eklerinin tanımlarını belirtmektedir. Şablonlar veritabanındaki '#' işareti 've', '%' işareti ise 'ardından' anlamındadır.

Örnek olarak 'tnm=ö[i]*tm#orj=icadı' ifadesi cümlede 'XXX'in icadı' şeklinde bir yapıyı ifade etmektedir.

3.2.3 Kullanıcı sorularının toplanacağı ortak veritabanı

Sistem kullanıma geçtiğinde kullanıcılardan geri besleme alabilmek amacıyla kullanıcıların sorduğu her soru Internet üzerinde bulunan ortak bir veritabanına kaydedilmesi düşünülmektedir. Bu sayede sistemin performansı ölçülebilecek ve en çok sorulan soru türleri

için ek şablonlar eklenebilecektir.

3.3 Sistemin Gerçekleştirilmesi

Sistem Delphi 6.0 uygulama geliştirme ortamında yazılmıştır. Veritabanı yöneticisi olarak MS Access kullanılmıştır. Sistem oldukça yoğun işlemler gerçekleştirdiğinden Internet üzerinde bir siteye çalışan sunucu tabanlı bir mimari tasarlanmamıştır. Bunun başlıca iki nedeni vardır. Birincisi, aynı anda kendisine bağlanan yüzlerce kişiye bu servisi verebilecek bir sunucunun maliyetinin çok yüksek olması, ikincisi ise cevap süresinin daha makul bir süreye indirilmesidir. Internet üzerinde çalıştırılması yerine kullanıcının bilgisayarında çalışacak bir EXE olarak tasarlanması uygun görülmüştür. Ancak kullanıcıların sordukları soruların sisteme geri besleme olarak dönebilmesi içinde programa kullanıcının sorduğu her soruyu Internet'te ortak bir veritabanına kayıt eden modülün eklenmesi planlanmıştır.

3.4 Şablonların Oluşturulma Metodu

Daha önce gerçekleştirilmiş olan birçok soru cevaplama sisteminde olduğu gibi cevap şablonları ve ağırlık değerleri elle oluşturulmuştur. Bu işlem oldukça zaman isteyen bir modüldür. Sistemin performansını arttırmak için gerçekten çok sayıda şablona ihtiyaç duyulmaktadır. Şu an mevcut olan sistem için 279 adet şablon oluşturulmuştur. Şablonların elde edilmesinde “nedir soruları” için ve “nedir haricindeki sorular” için farklı yöntemler uygulanmıştır. 3.4.1 ve 3.4.2 bölümlerinde şablonların belirlenmesinden ayrıntılı olarak bahsedilmiştir..

3.4.1 Nedir sorularının şablonlarının oluşturulması

“Nedir soruları” bir kelimenin tanımını sorduğu için Internet'teki bir sözlükten yararlanılmıştır. Yaklaşık 4000 Türkçe kelimeden oluşan bir listedeki kelimeler Türk Dil Kurumu'nun Internet sitesindeki sözlüğe (<http://www.tdk.gov.tr/tdksozluk/sozara.htm>) gönderilip sonuç sayfalarındaki kelime anlamları bir dosyaya kaydedilmiştir.

Bir sözlükteki en çok tekrar edilen kelimeler Türkçe'deki tanım ifadelerini içereceğinden; kelime anlamlarını içeren dosyadaki kelimelerden en çok tekrarlanan 50 kelime(Türkçe'de en sık tekrarlanan kelimeler çıkarıldıktan sonra) “nedir sorularının” şablonları olarak belirlenmiştir. Çizelge 3.13'te nedir sorularının şablonlarının kendileri verilmiştir.

Çizelge 3.13 Nedir soru türü şablonları

orj=:	orj=gösteren	orj=olarak%bilin
orj=ad	orj=herhangi	orj=olarak%tanım
orj=adı%verilir	orj=için%	orj=olmak
orj=alan	orj=için%kullan	orj=olmayan
orj=alınan	orj=ifade%eder	orj=oluşan
orj=amacıyla	orj=ilgili	orj=oluşturan
orj=anlamında	orj=ismidir	orj=sağlamak
orj=araç	orj=kalan	orj=sağlayan
orj=aracı	orj=karşılık	orj=tanım
orj=bakımından	orj=kimse	orj=tanımlamak%için%kullan
orj=biçimde	orj=kısaltılmış%halidir	orj=tarafından
orj=biçiminde	orj=kısaltma	orj=türlü
orj=bulunan	orj=kısaltması	orj=veren
orj=denir	orj=kişi	orj=verilen
orj=dir%.	orj=konulan	orj=verilen%addır
orj=duruma	orj=kullanılan	orj=yani
orj=eden	orj=kullanılır	orj=yapan
orj=edilen	orj=madde	orj=yapıl
orj=etmek	orj=olan	orj=yarayan
orj=genellikle	orj=olarak	orj=yaşayan

3.4.2 Nedir haricindeki soruların şablonlarının oluşturulması

Her soru türü için soru cevap çiftleri belirlenmiş ve Internet'te soru ve cevabın birlikte geçtiği cümleler bulunmuştur. Bu cümleler incelenerek, her soru türü için sıkça kullanılan kelimeler ve yapılar elle belirlenmiştir.

Çizelge 3.14'te zaman soran sorular için belirlenen şablonlar ve bu şablonların elde edildiği örnek cümleler verilmiştir.

Çizelge 3.14'te X ler soruyu, Y ler ise cevabı temsil etmektedir.

Çizelge 3.14 Zaman şablonlarının elde edilmesinde kullanılan cümlelerden örnekler

X (Y -	ERZURUM KONGRESİ (23 Temmuz -
Y X	23 TEMMUZ ERZURUM KONGRESİ
X Y de	Erzurum Kongresi 23 Temmuz 1919'da
Y günü ... X	4 Eylül 1919 günü toplanan Sivas Kongresi
Y de ... X	4 Eylül 1919'da toplanan Sivas Kongresi'nin
X, Y günü	Sivas Kongresi, 4 Eylül 1919 günü
Y tarihinde ... X	17 Ağustos 1999 tarihinde saat 03.02 te meydana gelen
Y X (Y)	13 Kasım. Napolyon Viyana'ya girdi (1805).
Y tarihli X	3 Kasım 1839 tarihli Gülhane Hattı Hümayunu'nun
Y: X	19 Temmuz: Paris'te metro hizmete girdi.
Y yılında ... X	25 Haziran 1950 yılında çıkan Kore Savaşı

4. SİSTEMİN UYGULAMA SONUÇLARI

Soru cevaplama sistemlerinin değerlendirilmesi oldukça güç bir işlemdir. Birbirinden farklı sorularla test edilen iki sistemin birbirleriyle karşılaştırılması pek bilimsel bir yöntem değildir. Bu yüzden geliştirilen sistemin diğer sistemlerle karşılaştırılabilmesi için aynı sorularla olmasa bile aynı tür sorularla test edilebilmesi için öncelikle TREC-9 ve TREC-10 soruları yarışmanın sitesinden toplanmış ve daha sonra Türkçe'ye çevrilmiştir. Türkçe web sitelerinde cevabı olması mümkün gözükmeyen bazı özel sorular elendikten sonra 4 kişiye bu soruları başka türlü nasıl sorabilecekleri sorulmuştur. Kişilerden gelen sorular birleştirilerek 524 adet test sorusu elde edilmiştir. Cevapları sadece Google'ın sonuç sayfasında arayan sisteme, 524 soru verilmiş ve cevapları aldıkları puanlarla birlikte bir veritabanına kaydedilmiştir. Sistemin verdiği cevaplardan oluşan veritabanını incelemek için Şekil 4.1'de görülen arabirim tasarlanmıştır.

The screenshot displays a web-based interface for evaluating a question-answering system. At the top, there are input fields for 'Kriter =', 'Soru No = 20030608_17:28:45_15', 'Soru Türü = zaman', and 'Soru = Amerika Bağımsızlık günü'. Below this, a section titled 'CEVAPLAR' contains five rows, each representing a different answer option. Each row has a text area containing a paragraph of text and a set of radio buttons for selection. The radio buttons are labeled '0', '1', '2', '3', '4', and '5'. The interface also includes a 'Soru Puanı Kaydet' button at the bottom right. The bottom of the screen shows a Windows taskbar with various application icons and the system clock.

Şekil 4.1 Sistemin değerlendirilmesi için tasarlanan arabirim

Sistemin cevap verdiği sorularlar 5 farklı puanlama tekniğine (1. eşleme 2. eşleme+sıralama 3. eşleme+şablon 4. sıralama+şablon 5. eşleme+sıralama+şablon) göre sıralanıp doğru cevabın kaçınıcı sırada olduğu bir tabloya kaydedilmiştir.

Şekil 4.1’de “Amerika’nın bağımsızlık günü ne zamandır?” sorusuna sistemin vermiş olduğu cevaplar 5 farklı şekilde puanlanmıştır. Eşleme ve eşleme+şablon puanlarında doğru cevap ilk sırada yer alırken, eşleme-sıralama puanlamasında doğru cevap 3.sırada ve diğer 2 puan türünde doğru cevap 2. sırada yer almıştır. Puanlama türlerinden hangisinin en iyi MRR puanına sahip olduğunu bulmak için cevabını doğru olarak buldukları her sorudan (1/ kaçınıcı sırada doğru buldukları) kadar puan almışlardır. Doğru cevap veremedikleri yada doğru cevabın ilk 5’te yer almadığı durumlarda sorudan puan alamamaktadırlar.

Çizelge 4.1’de sisteme sorular 524 adet sorunun soru türlerine göre dağılımı ve 5 farklı puanlama tekniğinin kaçar soruya doğru cevap(ilk 5 içinde) verdiği gösterilmiştir. Görüldüğü gibi en fazla soruya doğru cevap veren puanlama türü 230 soruya doğru cevap(ilk 5 içinde) vererek %43.90’lık bir doğruluk payına sahip olmuştur.

Çizelge 4.1 Soru ve doğru cevap sayılarının soru türlerine göre dağılımı

Soru türü	Soru Sayısı	Eşleme	Eşleme+ Sıralama	Eşleme+ Şablon	Sıralama+ Şablon	Eşleme+ Sıralama+ Şablon
Ağırlık	1	1	1	1	1	1
Başkent	6	6	6	6	6	6
bileşen	8	2	1	1	1	2
Belirtili nesne	6	4	4	4	4	4
derece	9	4	4	3	3	3
Dil	4	2	2	2	2	2
Doğum yılı	2	1	0	1	1	1
Doğum zaman	7	3	2	3	3	4
Edilgen yüklem	1	1	1	1	1	1
Evet/Hayır	4	3	3	3	3	3
Eser şarkı	1	1	1	1	1	1
Etki	1	1	1	1	1	1
Fiyat	1	1	1	1	1	1
hangi	13	5	4	5	4	4
hangisi	32	14	12	14	12	12
Hız	5	3	3	3	3	3
icat	10	6	6	6	6	6
il	19	7	9	7	9	8
Keşif	2	1	1	1	1	1
kim	49	21	18	22	16	18
kıta	2	1	1	1	1	1
km	3	2	2	2	2	2
kullanım	10	2	1	2	1	1
kuş	1	1	0	1	0	0
litre	1	1	1	1	1	1
mesafe	6	4	4	4	4	4
metremesafe	9	5	4	5	4	4
nasıl	10	6	6	6	6	6
neden	6	2	1	3	1	1
nedir	117	33	33	40	39	40
nelerdir	13	2	2	2	2	2
nerededir	20	12	12	13	13	13
nesne	11	2	2	2	2	2
nüfus	10	7	6	6	7	5
olumzaman	3	3	3	3	3	3
oran	11	8	7	8	7	7
para	4	2	1	2	2	2
renk	7	1	1	1	1	1
sayı	26	15	15	14	14	14
şehir	3	1	1	1	1	1
sembol	1	0	1	1	1	1
süre	7	1	1	1	1	1
süreylil	1	1	1	1	1	1
yıl	11	3	3	3	3	3
zaman	50	22	21	21	22	22
toplam	524	224	210	230	218	220
		42.70%	40.00%	43.90%	41.60%	41.98%

Çizelge 4.2’de 5 farklı puanlama tekniğinin soru türlerine bağımlı olarak aldıkları MRR

puanları verilmiştir. En yüksek MRR puanını 0.91 'le 6 adet başkent sorusunun 6'sına birden ilk sıralarda doğru cevap veren sıralama+şablon puanlama türü almıştır.

Çizelge 4.2 Soru türlerine göre MRR puanlarının dağılımı

Soru türü	Soru Sayısı	Eşleme	Eşleme+ Sıralama	Eşleme+ Şablon	Sıralama+ Şablon	Eşleme+ Sıralama+ Şablon
ağırlık	1	0.5	0.5	0.5	0.5	0.5
başkent	6	0.866666667	0.833333333	0.875	0.916666667	0.833333333
bileşen	8	0.15625	0.125	0.125	0.0625	0.09375
blinesne	6	0.375	0.583333333	0.388888889	0.583333333	0.583333333
derece	9	0.388888889	0.388888889	0.25	0.333333333	0.333333333
dil	4	0.3125	0.5	0.3125	0.5	0.5
Doğum yılı	2	0.5	0	0.5	0.5	0.5
Doğum zaman	7	0.357142857	0.171428571	0.428571429	0.314285714	0.457142857
Edilgen yüklem	1	0.333333333	0.5	0.333333333	0.5	0.5
Evet/hayır	4	0.625	0.5625	0.625	0.5625	0.5625
Eser şarkı	1	0.5	0.5	0.5	1	1
Etki	1	0.5	1	0.5	1	1
Fiyat	1	0.5	0.5	0.5	1	1
hangı	13	0.173076923	0.166666667	0.173076923	0.166666667	0.166666667
hangisi	32	0.240104167	0.234375	0.238541667	0.278645833	0.234375
Hız	5	0.366666667	0.466666667	0.5	0.6	0.6
icat	10	0.6	0.45	0.316666667	0.29	0.3
il	19	0.333333333	0.372807018	0.328947368	0.295614035	0.352631579
keşif	2	0.5	0.5	0.5	0.5	0.5
kim	49	0.274489796	0.247278912	0.310204082	0.216666667	0.265986395
kıta	2	0.5	0.5	0.5	0.5	0.5
km	3	0.666666667	0.666666667	0.666666667	0.666666667	0.666666667
kullanım	10	0.133333333	0.033333333	0.12	0.025	0.025
kuş	1	0.2	0	0.2	0	0
litre	1	1	1	1	0.333333333	1
mesafe	6	0.472222222	0.361111111	0.347222222	0.444444444	0.444444444
metremesafe	9	0.3	0.3	0.3	0.3	0.3
nasıl	10	0.283333333	0.286666667	0.256666667	0.27	0.345
neden	6	0.25	0.083333333	0.305555556	0.083333333	0.083333333
nedir	117	0.178917379	0.197578348	0.212393162	0.235327635	0.242735043
nelerdir	13	0.115384615	0.115384615	0.115384615	0.115384615	0.115384615
nerededir	20	0.268333333	0.276666667	0.435	0.405833333	0.433333333
nesne	11	0.181818182	0.090909091	0.181818182	0.075757576	0.090909091
nüfus	10	0.483333333	0.375	0.458333333	0.475	0.35
olumzaman	3	0.777777778	0.777777778	0.777777778	0.777777778	0.777777778
oran	11	0.545454545	0.590909091	0.545454545	0.454545455	0.5
para	4	0.375	0.25	0.375	0.333333333	0.333333333
renk	7	0.028571429	0.028571429	0.028571429	0.028571429	0.028571429
Sayı	26	0.509615385	0.477564103	0.490384615	0.448717949	0.467948718
şehir	3	0.333333333	0.333333333	0.333333333	0.333333333	0.333333333
sembol	1	0	1	0.333333333	1	1
Süre	7	0.142857143	0.071428571	0.142857143	0.071428571	0.071428571
Süreyl	1	1	1	0.5	0.5	0.5
Yıl	11	0.151515152	0.2	0.227272727	0.212121212	0.212121212
Zaman	50	0.362333333	0.361666667	0.373333333	0.385	0.395

Çizelge 4.3'te puanlama türlerine göre ortalama MRR puanları verilmiştir. En fazla soruya doğru cevap veren puanlama türü eşleme+şablon iken en yüksek MRR puanını 0.313 ile eşleme+sıralama+şablon puanlama türü almıştır.

Çizelge 4.3 Puanlama türlerine göre MRR puanları

	Soru Sayısı	Eşleme	Eşleme+Sıralama	Eşleme+Şablon	Sıralama+Şablon	Eşleme+Sıralama+Şablon
MRR	524	0,298	0,289	0,308	0,301	0,313

TREC-9 yarışmasına katılan sistemlerin ortalama MRR puanı 0.234 olduğu ve sistemin sadece Google'ın sonuç sayfasında cevapları aradığı ve yarışmaya katılan diğer sistemlerin sahip oldukları doğal dil işleme araçları kullanılmadığı düşünüldüğünde sistemin ulaştığı 0.313 'lük MRR puanı iyi bir sonuç olarak değerlendirilebilir. Daha yüksek MRR puanı elde etmek için yapılabilecekler öneriler bölümünde belirtilmiştir. Çizelge 4.4'te TREC yarışmasının birinci ve ikincilerinin MRR puanları verilmiştir.

Çizelge 4.4 TREC yarışmasının en iyileri(MRR puanları)

	TREC-8	TREC-9	TREC-10
En iyi	0.66	0.58	0.68
En iyi ikinci	0.56	0.32	0.59

Çizelge 4.5'te sisteme sorulan sorulardan ve sistemin verdiği cevaplardan örnekler verilmiştir. Puanlama türü eşleme+sıralama+şablon'dur.

Çizelge 4.5 Sisteme sorulan sorulara ve sistemin cevaplarına örnekler

soru		Cevaplar
TBMM ne zaman açıldı?	1	İşte hemen sonra da Ankara'da 23 Nisan 1920 tarihinde TBMM açıldı.
	2	Sonuçta 23 Nisan 1920 Tarihinde Ankara'da TBMM açıldı ve Türk .
	3	23 Nisan 1920'de geniş bir katılımıla TBMM açıldı.
	4	Teklif, bugün (24 Şubat 2000) koalisyon partileri genel başkanlarının düzenledikleri zirveden hemen sonra TBMM'de imzaya açıldı.
	5	1 Kasım 1927, TBMM'nin üçüncü dönemi açıldı.
Gaziantep'in kurtuluş günü ne zamandır?	1	Şöyle ki, bugün, 25 Aralık 1999, Gaziantep İlimizin kurtuluş günüdür.
	2	25 Aralık'ta Gaziantep'in kurtuluş gününde açacağımız sergiye gelecek insanlardan böyle bir grup çıkarabileceğimizi ve gelecek sene beş .
	3	-
	4	-
	5	-
Tarkan ne zaman doğdu?	1	Tarkan Sülalp 1968 yılında İstanbul'da doğdu.
	2	TARKAN GÖZÜBÜYÜK Tarkan Gözübüyük Bass Gitar & Vokal 1970 yılında Erzurum'da doğdu.
	3	Tarkan Gözübüyük: Bass, Vocal 1970 yılında Erzurum'da doğdu.
	4	-
	5	-
Türkiye İnternet ile ne zaman tanıştı?	1	12 Nisan 1993'de Ankara-Washington 64K'lık bağlantısı ile Türkiye İnternet'le tanıştı.
	2	Türkiye İnternet ile 12 nisan 1993'de tanıştı ve İnternetimiz 6.
	3	12 Nisan 1993'te Ankara-Washington arasında kiralık hatla kurulan bağlantı ile Türkiye internetle tanıştı.
	4	Türkiye İnternet'le 12 nisan 1993'de ODTÜ'den gerçekleştirilen Ankara-Washington bağlantısı ile tanıştı.
	5	-
Boğaz Köprüsü'nün yüksekliği ne kadardır?	1	Boğaz Köprüsü'nün yüksekliğinin yaklaşık 3 katına eşdeğerdir .
	2	64 metre yüksekliğindeki Boğaz Köprüsü'yle, Saipem 7000 .
	3	Boğaz Köprüsü'nün denizden yüksekliği ise 60 metre.
	4	Mesela Boğaziçi Köprüsü'nün deniz seviyesinden yüksekliği sadece 64 .
	5	boyu 165 metre olan Birinci Boğaz Köprüsü'nün .
Ay'ın Dünya'dan uzaklığı ne kadardır?	1	Pembe-mavi benekli bu uydunun Neptün'e uzaklığı, yaklaşık olarak Dünya-Ay arası uzaklığı kadardır (354.200 Km.
	2	Ay'ın Yer yüzeyine uzaklığı 350000 km ile 400000 km .
	3	Ay'ın Dünya'ya olan ortalama uzaklığı 384000 km dir.
	4	Yer çevresinde yaklaşık olarak her 48 saatte bir dönmekte ve bu hareketi sırasında kimi zaman Yer'e Ay uzaklığı kadar yaklaşmakta, kimi zaman da Ay .
	5	Bu uzaklık, Ay-Dünya arasındaki uzaklığın (6) katıdır.
toplumda şizofreninin görülme oranı nedir?	1	belirgin umutlar bulunan şizofreninin toplumda görülme sıklığı %1 .
	2	DR.MERCAN:Sayın okuyucu, Genel toplumda şizofreni

		görölme oranı %1 iken anne veya babanın şizofrenik olması durumunda şizofreninin ortaya çıkma .
	3	Psikiyatri polikliniklerinde görölme oranı Batıda %1-3 .
	4	konversiyon belirtilerinin görölme oranı %82 olarak .
	5	-
Suyun kaldırma kuvvetini kim bulmuştur?	1	Archimedes daha sonra Dünyaya kendisini tanıtan buluşunu yapmıştır yani sıvıların dengesi kanunu bulmuştur(suyun kaldırma kuvvetini).
	2	Suyun kaldırma kuvvetini ilk defa kim bulmuştur?
	3	-
	4	-
	5	-
Atmosferdeki oksijen oranı yüzde kaçtır?	1	Atmosferdeki oksijen oranı yüzde 21 değil de yüzde 22 olsa atmosfer yanıcı hale .
	2	Hele oran yüzde 25'lere çıksaydı Dünyamız dev yangınlarla .
	3	-
	4	-
	5	-
Kafein hangi içeceklerde bulunur?	1	Kafein (kahve, çay, kola gibi içeceklerde bulunur) plasentadan geçer ve bebek kan dolaşımına girer.
	2	Kafein içeren içecekler (çay, kahve ve kola) azaltılmalı .
	3	Kafein, kahve, çay ve çikolata gibi bazı yiyecek ve sıcak içeceklerde de al olarak bulunur.
	4	Kafein içeren kahve, çay gibi içecekler mümkün olduğu .
	5	Kolalı içeceklerde karbondioksit ve kafein bulunur.
Türkiye'nin en büyük yüzölçümüne sahip ili hangisidir?	1	Bu alanı ile Türkiye'nin en büyük yüzölçümüne sahip olan ilidir.
	2	Konya'nın Türkiye'nin en büyük yüzölçümüne sahip ili olduğunu, bu nedenle arazi probleminin bulunmadığını bildiren Özkafa, pist için aday iller .
	3	KONYA TARİHİ VE COĞRAFI DURUM İç Anadolu Bölgesi'nde yer alan ve 38.183 km ² ile Türkiye'nin en büyük yüzölçümüne sahip ili olan Konya, Ankara .
	4	Konya, Türkiye'nin en büyük yüzölçümüne sahip ili.
	5	174 kasaba ve 610 köyden oluşan, Türkiye'nin en büyük yüzölçümüne sahip Konya ilimiz .
sigaradaki bağımlılık yapan madde ?	1	Sigaradaki bağımlılık yapan madde ise nikotindir.
	2	14-Bağımlılık yapıcı madde kullanımı (sigara, alkol ya da uyuşturucu-uyarıcı madde kullanımı).
	3	Nikotin eroin ve kokain kadar bağımlılık yapan bir maddedir.
	4	Nikotin eroin kadar bağımlılık yapan bir maddedir .
	5	Nikotin ise, sadece bağımlılık yapan bir madde değil, aynı .

Arjantin dünya kupasını ne zaman kazandı?	1	önce 1974'te kazanmıştı; 1990 + 1974= yine 3964, Arjantin dünya kupasını en son 1986'da kazanmıştı, ondan önce de 1978'de kazandı 1986 + 1978 .
	2	Arjantin dünya kupasını en son 1986 .
	3	-
	4	-
	5	-
kontakt lens niçin kullanılır?	1	Bu tip kontakt lensler göz şekline tamamen uyabilen .
	2	-
	3	-
	4	-
	5	-
puro neden yapılır?	1	Havana'da yapılan puro, özel bir seçimden sonra iç ve kılıf yapraklar olarak ayrılan tütün yapraklarının elle sarılmasıyla yapılır.
	2	-
	3	-
	4	-
	5	-
hamsterların ortalama ömrü ne kadardır?	1	Bunun yanında hamsterin ortalama ömrü 3 sene kadardır.
	2	-
	3	-
	4	-
	5	-
Galatasaray kaç kere şampiyon olmuştur?	1	Son 6 senede 5 kere şampiyon olan Galatasaray, her sene .
	2	-
	3	-
	4	-
	5	-
İsmet İnönü'nün ölüm tarihi?	1	Baykal-İsmet İnönü anıldı-25 Aralık 2000 İsmet İnönü anıldı.
	2	İsmet İnönü Atatürk'ün ölümü üzerine 11 Kasım 1938'de cumhurbaşkanlığına seçildi.
	3	İSMET İNÖNÜ GÖREV SÜRESİ: 11 KASIM 1938 - 22 MAYIS 1950 Atatürk'ün silah arkadaşı İsmet İnönü savaş meydanlarındaki kahramanlıklarına, Lozan .
	4	Milli Şef İsmet İnönü öldü (1973).
	5	İsmet İnönü Cumhurbaşkanlarımız İsmet İNÖNÜ 11 Kasım 1938 - 22 Mayıs 1950.
ışık hızı nedir?	1	ışık pek hızlı yayıldığını için bu metotla ışık hızı ölçülemedi.
	2	ışığın boşlukta bir saniyede aldığı yola ışık hızı denir.
	3	ve milimetre dalga gibi gittikçe artan frekans veya gittikçe azalan dalgaboyuna doğru bir eğilim göstermiştir (dalga boyu = ışık hızı / frekans).
	4	(burada c: ışık hızı) olarak ifade edilir.
	5	Bilimsel araştırmalarda: Çok hassas bilimsel ölçümlerde (ışık hızı ölçümü),, .
dünyanın en yaşlı insanı	1	Doğu'dan gelmiş ve İstanbul'da yaşamak için işportacılık

kimdir?		yapmak zorunda olan bir sürü genç-yaşlı insan, bozuk Türkçeleriyle bağırıp duruyordu .
	2	10:40 Guinness Rekorlar Kitabı'na "dünyanın en yaşlı insanı" olarak geçen Sarah Knauss, 119 yaşında öldü.
	3	Elizabeth'in dünyanın en yaşlı insanı olduğuna inanılıyor.
	4	1925'te 'dünyanın en yaşlı insanı' olarak İtalya'ya,1930'da yılında Amerika'ya, 1931 yılında İngiltere'ye götürülür.
	5	Kitabı'nda, dünyanın en yaşlı insanı olarak 1997 yılında 122 yaşında ölen Fransız Jeanne Louise Calment, yaşayan en yaşlı insan olarak ise 115 .
uzaya çıkan ilk insan kimdir?	1	yıldönümünü kutlayan Rusya'da, uzaya çıkan ilk insan Yuri Gagarin'i anma .
	2	Gagarin uzaya çıkan ilk insan olarak tarihe geçti.
	3	Yuri Gagarin uzaya çıkan ilk insan olarak dünyaya dönünce, Kruçev tarafından kabul edilmişti.
	4	Dün Zulkarneyn'in boynuzlarını tetkik edenler bugün onun uzaya ilk çıkan insan olduğu tezlerini yazarak taze buluşlarıyla iftihar ediyor olabiliyorlar.
	5	Fotoğrafta Sergey Korolyev uzaya çıkan ilk insan Yuri Gagarin'le birlikte görülüyor.
Tarkan Tevetoğlu kimdir?	1	bilmedin Fark etmez artık, fark etmez artık Kalmadı değerim senin Söz: Alpay Aydın Müzik: Tarkan Tevetoğlu Düzenleme: Ozan Çolakoğlu, MINI POLL.
	2	Tarkan Tevetoğlu, anlaşma çerçevesinde Pepsi'nin reklam filmi ve etkinliklerinde rol alacak.
	3	Söz: Alpay Aydın Müzik: Tarkan Tevetoğlu Düzenleme: Ozan Çolakoğlu Translation: Serkan Altay.
	4	173 Beyazıt Öztürk ve Tarkan Tevetoğlu 26 Tem 2001.
	5	Türk pop müziğinin mega- starı Tarkan Tevetoğlu'nun, "tarkan.
Namık Kemal kaç yaşında öldü?	1	Şair, siyaset ve düşünce adamı Namık Kemal 48 yaşında öldü.
	2	Mutasarrıfı Abdülatif Bey ve 6 yaşında Namık Kemal, piyano, ayna .
	3	Haziran 1865) Namık Kemal 25 yaşında bir gençti.
	4	kampanyasına Adana'dan 10 milyar lira bağışla katkıda bulunan Namık Kemal Biçer 39 yaşında bir organizatör.
	5	Eğitime Ankara Namık Kemal İlkokulu'nda başlayan Çandar, 11 yaşında iken Kayseri .
ortalama kalp atış sayısı nedir?	1	Dakikadaki kalp atış sayısı 220, solunum sayısı 450 dir .
	2	VRX-1000, 1958 Lazer ve kalp atış hızını.
	3	Aylara göre deprem sayısı ortalama 150 -200 .
	4	çeken Karpuz, sözlerini şöyle sürdürdü: "Normalde dakikada

		ortalama 60 olan kalp atış hızı, cinsel .
	5	kişinin kalbi dinlenme sırasında ortalama 60-70 kez atarken, bu atış sayısı antrenmanlı kişilerde 50'nin altındadır.
insan vücudundaki su oranı nedir?	1	Yeni doğmuş bir bebeğin vücudundaki su oranı %75 kadardır ancak büyüme ile birlikte bu oran %60'a düşmektedir.
	2	Hücre içi su hücreler içinde bulunur ve toplam vücut suyunun 2/3'nü oluşturur.
	3	Yeni doğanın vücudundaki su oranı %80 iken, yetişkinlikte bu oran .
	4	İnsan vücudu ısıyı, kan dolaşımı oran ve miktarını değiştirerek, deri ve ter bezleri ile su kaybederek ve vücut sıcaklığı 37 °C?
	5	birlikte bu oran, erkeklerde vücut ağırlığının % 45 .
en büyük gezegen hangisidir?	1	Çünkü civarda bir güneş ve çevresinde birbirlerine oldukça yakın üç büyük gezegen görülüyordu.
	2	Bu gezegen büyük bir idealizmin, kendini, tüm benliğini bir alana adayan ve bunu yaparken de adeta kurban haline dönüşen bir süreç yaşatır.
	3	En büyük gezegen olan Jupiter' in hacmi dünyanınkinin 1300 katıdır.
	4	327 AnaSayfa Puan Ver Kırık Bağlantı Bildir Detaylar Kategori: Duvar kağıtları/Astronomi/Astronomi Jüpiter Tanım: En büyük gezegen Jüpiter.
	5	Astrolojinin tanımı Güne , 8 büyük gezegen ve Ay' m oniki takımyıldızın do?
Türkiye'nin en derin gölü hangisidir?	1	Gölde Deve turu Türkiye'nin en derin gölü olan Salda .
	2	Gölde Deve turu Türkiye'nin en derin gölü olan Salda kıyılarında süslenmiş develer gezi ve fotoğraf çekirmek isteyen turistlerce kiralaniyor.
	3	Buradan Türkiye'nin en derin gölü olan Salda gölüne uğramak için yola çıkılacak.
	4	Salda Gölü (Türkiye'nin en derin ve temiz gölü).
	5	Doğal güzellikleri ile de öne çıkan Burdur; İnsuyu Mağarası, Türkiye'nin en derin ve temiz gölü olan Salda Gölü, uluslararası öneme sahip Burdur .
Türkler neleri severler?	1	Türkler'i severler bir de tabii İngiliz'in karsısında olan herkesle Sultan Abdülhamit ilgilenmiş.
	2	Konuyla ilgili olarak yabancı eserlerde şu övgü dolu ifadeler rastlarız: "Hiçbir istisnası olmamak şartıyla bütün Türkler hayır severler; ne din .
	3	Onların çoğunluğu Türkler'i tanır, severler.
	4	Türkler yazıyı severler ve bu sanatta ustadırlar.
	5	Türklerle Arnavutlar birbirlerini çok severler.

Süpernova nedir?	1	Pulsarlar ve Süpernova Kalıntıları (TÜBİTAK).
	2	Süpernova Artıkları (SNR: Supernovae Remnants).
	3	Pulsarlar ve Süpernova Kalıntılarının Evrimi (TÜBİTAK).
	4	Nova veya Süpernova olarak adlandırılan bu yıldızlardaki demir miktarı belli bir oranı geçince, artık yıldız bunu taşıyamaz ve patlar.
	5	Ulusal Astronomi Toplantısı Poster Özeti Gönderme Yakın Sarmal Gökadalarındaki Süpernova Kalıntılarının Optik bölgede gözlenmesi.
Deli Dana hastalığı nedir?	1	ŞAP NEDİR Deli Dana hastalığı, ya da teknik söylemi ile sığırların süngersi beyin hastalığı (Bovine Spongiform Encephalopathy), ilk kez 1986 yılında .
	2	Deli Dana hastalığı, ya da teknik söylemi ile sığırların süngersi beyin hastalığı (Bovine Spongiform Encephalopathy), ilk kez 1986 yılında .
	3	İngilizcesi ile mad cow disease olarak adlandırılan Deli Dana hastalığı (Bovine spongiform encephalopathy - BSE - olarak da tanımlanmakta .
	4	Bunlardan biri olan Deli Dana hastalığı,.
	5	Ama öbür yanda Deli Dana hastalığı kedi ve farelerde gözükabilirken hamsterlarda ve köpeklerde zarar verici bir etken olarak gözükmemiştir.
enzimlerin vücuttaki görevi nedir?	1	Bir tür protein olan enzimlerin görevi vücuttaki reaksiyonları başlatmak ve hızlandırmaktır.
	2	Bunun için birçok enzim görev alır.
	3	Krom; vücuttaki enzim ve hormonlar için .
	4	kompleks vitaminleri vücuttaki enzim sistemlerindeki görevleri .
	5	Sadece bir molekül olan enzimlerin bu mükemmel görevi kusursuz bir .
Kriptografi nedir?	1	mevcuttur: · Gerekli olan sertifikaların sayısı(bu durum CA sunucu yazılımı ile satın alınacak lisansda belirtilmelidir) · Kriptografi ve özellikle .
	2	Bruce Schneier: Applied Cryptography (Uygulamalı Kriptografi), 2.
	3	Kriptografi, okunur bir formattaki bilginin başkalarının okuyamayacağı bir formata dönüştürülmesi bilimidir.
	4	(HTU) - Kriptografi KRİPTOGRAFİ.
	5	Kriptografi mesajların gizli tutulması sanatı veya bilimidir.
Öğrenmenin en kolay yolu nedir?	1	Öğrenmenin daha kolay olmasını sağlamaktadır.
	2	Öğrenmenin en kolay yolu deneme yapmak.
	3	Yabancı dil öğrenmenin kolay yolu, 1 2 3 4 5 6 7 Hazırlayan: Zeynep Özyalın.

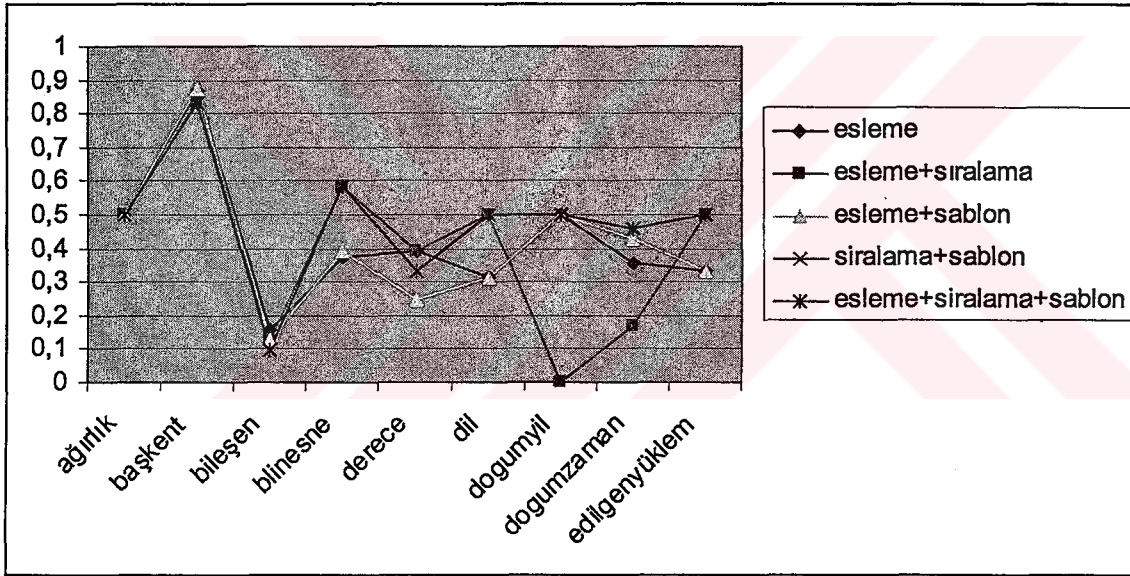
	4	Telefon borcunuzu öğrenmenin en kolay yolu size bir tuş kadar yakın 163 ayrıca sitemizden de borcunuzu öğrenebilirsiniz.
	5	Nerede olduğunu bilmiyorsanız öğrenmenin en kolay yolu KFM'yi açıp "Düzenle/Applications" yolunu izleyip, adres satırındaki yolu yazmaktır.
Plastiğin hammaddesi nedir?	1	HY: Plastiğin hammaddesini Petkim'den alıyoruz.
	2	HY: Plastiğin hammaddesi petrol.
	3	-
	4	-
	5	-
Bumbar hangi ilimizin yemeğidir?	1	Siirt'in millî yemeği Bumbar'ın, çeşit çeşit .
	2	Siirt'in millî yemeği Bumbar'ın, çeşit .
	3	-
	4	-
	5	-
Meryem Ana Kilisesi nerededir?	1	MÜJDE ARSLAN Melikahmet'te doğup büyüyen, org sesini ilk kez Amed'deki Meryem Ana Kilisesi'nde duyan Borusan Flarmoni Orkestrası Şefi Gürer Aykal, 44 .
	2	B) Diyarbakır - Cahit Sıtkı Tarancı Müzesi - Eski Diyare Bekir Bölümü - Meryem Ana Kilisesi, B) Elazığ -Harput Kalesi -Hazar Gölü -Keban Barajı.
	3	05 EYLÜL 2002 AÇIKGÖZ, Hermine (MERZİFON, 1336) uğurlama töreni 9 Eylül 2002'de Kumkapı Meryem Ana Kilisesi'nde yapıldı.
	4	Büyükada'daki Meryem Ana Kilisesi'nin misafirhanesinde yangın çıktı 3 katlı tarihi Meryem Ana Kilisesi'nin çatı katında bulunan.
	5	metre mesafede, altında Markao Mikolaidis'e ait Aleko'nun Yeri Deniz Park isimli restoran bulunan ve üst iki katı da Meryem Ana Kilisesi'nin lojmanı .
Eyfel Kulesi nerede?	1	Aşıklar kenti Paris'in simgesi Eyfel Kulesi, seks fantezilerinin de vazgeçilmez mekanı haline geldi.
	2	Aksi takdirde Eyfel Kulesi, Piza Kulesi, Köln Katedrali ya da Frankfurt'taki Avrupa Merkez Bankası gibi yüksek yapılardan kimi Avrupalı'ların .
	3	da bir kitabında Çamlıca'daki radyo verici antenlerinin, Eyfel Kulesiyle Galata .
	4	Paris'teki Eyfel Kulesi açıldı (1889).
	5	Mumcu, Fransa'nın sembolü haline gelen Eyfel Kulesi ile ABD'deki en .
Nijerya'da ne kadar insan yaşıyor?	1	kalabalık nüfusuna sahip olan Nijerya'da yaklaşık 110 milyon kişi yaşıyor.
	2	-
	3	-
	4	-
	5	-

Nijerya'nın nüfusu kaçtır?	1	Bosna'yı takip eden Nijerya'nın nüfus artışı ise yüzde 2.8 olarak gerçekleşti.
	2	Nijerya'nın nüfusu yüzde 132.3 oranında artış göstererek 278.8 milyon olacak.
	3	Artışlarla birlikte Bosna'nın nüfusu 4 milyona, Nijerya'nın nüfusu 122 milyona, Türkiye'nin nüfusu ise 64 milyona ulaştı.
	4	Nijerya, 120 milyonluk nüfusu olan devasa bir ülke.
	5	Çünkü Nijerya, Afrika'da 120 milyon nüfusu olan önemli bir pazar potansiyeli olan ülkedir.
Termometre ne için kullanılır?	1	örnekleri soğuturken havanın neminden etkilenmemesi için kullanılır, m) Su .
	2	Cumulus tipi bulutlardaki demet şeklindeki görünüşleri vurgulamak için kullanılır.
	3	Metal termometreler, civalı ve alkollü termometrelerin ölçemediği sıcaklıklar için kullanılır.
	4	Kumpas ve metre, her türlü ölçü alma ve markalama için kullanılır, hırdavat cıhazlarında bulunabilir .
	5	EPG 60 10 kollektöre kadar olan sistemler için kullanılır.
termostat ne için kullanılır?	1	sıcaklık ayarını yapacak termostatin hissedicisi kullanılır hacmin içinde .
	2	belirli deşarj tüpleri ve elektroforlar için kullanılır.
	3	kontrol kapakcıklarında çeşitli termostatlar kullanılır.
	4	ısılarının ayarlanması için termostat görevi görür .
	5	soğutma sisteminde ve ısı kontrol klapelerinde çeşitli termostatlar kullanılır.
motorun parçaları nelerdir?	1	Bu motorun parçaları hâkim ve savcılar kadar, bizleriz de.
	2	Hava Soğutmalı Motor: Fan veya motorun dönem parçaları üzerine yerleştirilen pervane yardımıyla hareket ettirilen havanın , motorun ısınan .
	3	Bunun sonucu motorun bazı parçaları yağsız kalıp, kilitlenmelere, önemli motor arızalarına sebep olabilir.
	4	Fakat turbo basıncından dolayı, motorun bazı parçaları normalden daha dayanıklı olmalıdır.
	5	Bir motorun bazı parçaları şunlardır: marş motoru, piston, segman, piston kolu, silindir kapağı, supap kapağı, eme manifoldu, egzost manifoldu .
İspanya'nın başkenti neresidir?	1	• 17-21 Ekim: İspanya'nın başkenti Madrid'deki 24.
	2	motortec" Otomotiv Yan Sanayi, Ekipmanları ve Aksesuarları Ticaret Fuarı 28 Mayıs-1 Haziran 2003 tarihleri arasında İspanya'nın başkenti Madrid'de .
	3	Ekim-Kasım 1991: İsrail'in, Filistin dahil Arap komşularıyla görüşme masasına oturduğu barış süreci, İspanya'nın başkenti

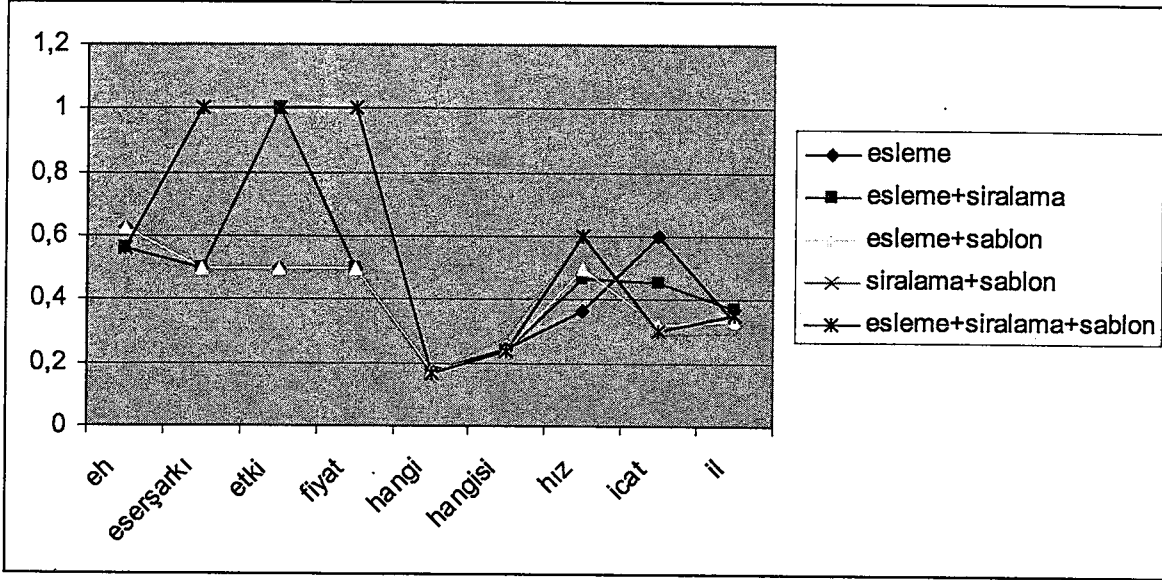
		Madrid'de başladı.
	4	da yapılan İslam Konferansı Örgütü Dışişleri Bakanları Terörizm Özel Toplantısı'na katılan ve ardından da İspanya'nın başkenti Madrid'de .
	5	Böylece Ronaldolar karı-koca İspanya'nın başkenti Madrid'de top koşturmaya başlayacaklar.
pusulayı kimler buldu?	1	Ertesi sabah sekizde uyanan Temel yanındaki masada şu pusulayı buldu.
	2	şekli kaşığa benzerdi, sonra insanlar güneyi gösteren pusulayı buldular.
	3	Manyetik pusulayı Çinliler buldu.
	4	Hadi olmuyorsa köyünden birini bul biz bunu yapamayız.
	5	türkchat pusula ayna ayva dost izmir istanbul sevgililer turkiye memleket.
depremi önceden bilmenin yolu var mıdır?	1	İşıkara, depremi önceden bilmenin mümkün olmadığını; ancak .
	2	2- Kazalara yol açabilecek davranışlardan kaçınma.
	3	anlatımla haberi, genelde topluma önceden bil- medi i .
	4	mevsim koşulları ve deprem önce- sinde bölgeye .
	5	da matrak: 'Son Adapazarı depremini önceden bildik!' Deprem olduktan sonra bilmenin çok önemi var .
İlahi Komedyayı kim yazmıştır?	1	batı dünyası onun insan sevgisinden etkilenmiş, İlahi Komedyayı?
	2	İlahi Komedyası / Dante Aligheri.
	3	Ayrıca Dante'nin (1261-1321) ünlü eseri İlahi Komedyası'da yer alan evren görüşü de .
	4	Papa,İtalyan fikir adamı Dante'nin İlahi Komedyası'nı okusun" dedi.
	5	İlahi Komedyası" adlı eseriyle Hristiyanlık öğretisinin .
bir insan günde kaç kez nefes alır?	1	Yetişkin bir insan günde ortalama 23.000 kez nefes alır.
	2	yetişkin bir insan günde ortalama olarak 23 bin kez nefes alır.
	3	Yetişkin bir insan günde ortalama olarak 23 bin kez nefes .
	4	6.) Bir insan 1'den 1 milyara kadar sayabilir ama bu sayım işlemi 12 sene sürer! 7.) Yetişkin biri günde ortalama 25.000 kez nefes alır.
	5	-
ilk kopyalanan hayvan nedir?	1	Jean Paul'ün kopyalamaya karşı gelişinin ardından kopyalanan ilk hayvan olan Dolly'nin babası İskoç bilim adamı Ian Wilmut'tan Papa'ya sert .
	2	Bu gelişme kopyalanan ilk hayvan olarak tarihe geçen Dolly adlı koyuna uygulanan klonlama işleminin devamı sayılabilir.

	3	26.12.2002) Merkezi Kanada'nın Quebec eyaletinde bulunan Raelian tarikatı, kopyalanan ilk insanın doğduğunu iddia .
	4	1997'de, başarıyla kopyalanan ilk memeli hayvan olan koyun Dolly .
	5	tarihte kopyalanan ilk canlı.

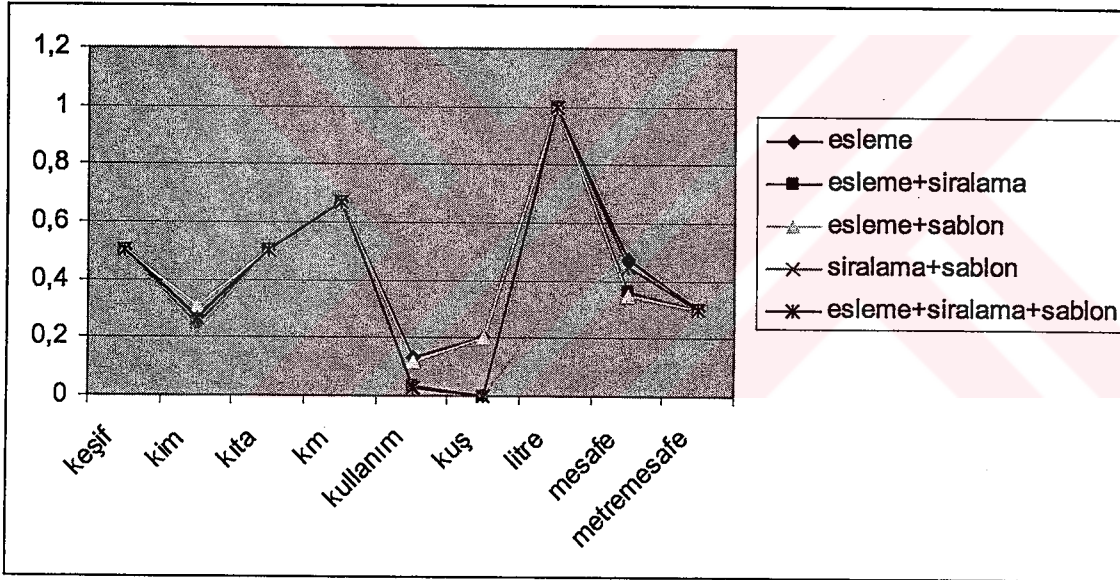
Şekil 4.2'te ağırlık, başkent, bileşen, belirtili nesne, derece, dil, doğum yılı, doğum zamanı ve edilgen yüklem soru türleri için; şekil 4.3'te evet/hayır, eser şarkı, etki, fiyat, hangi, hangisi, hız, icat ve il soru türleri için; şekil 4.4'te keşif, kim, kıta, kilometre, kullanım, kuş, litre, mesafe ve metre mesafe soru türleri için; şekil 4.5'da nasıl, neden, nedir, nelerdir, nerededir, nesne, nüfus, ölüm zamanı ve oran soru türleri için; şekil 4.6'de para, renk, sayı, şehir, sembol, süre, süre yıl, yıl ve zaman soru türleri için sistemin 5 farklı puanlama türüne göre MRR puanları gösterilmiştir.



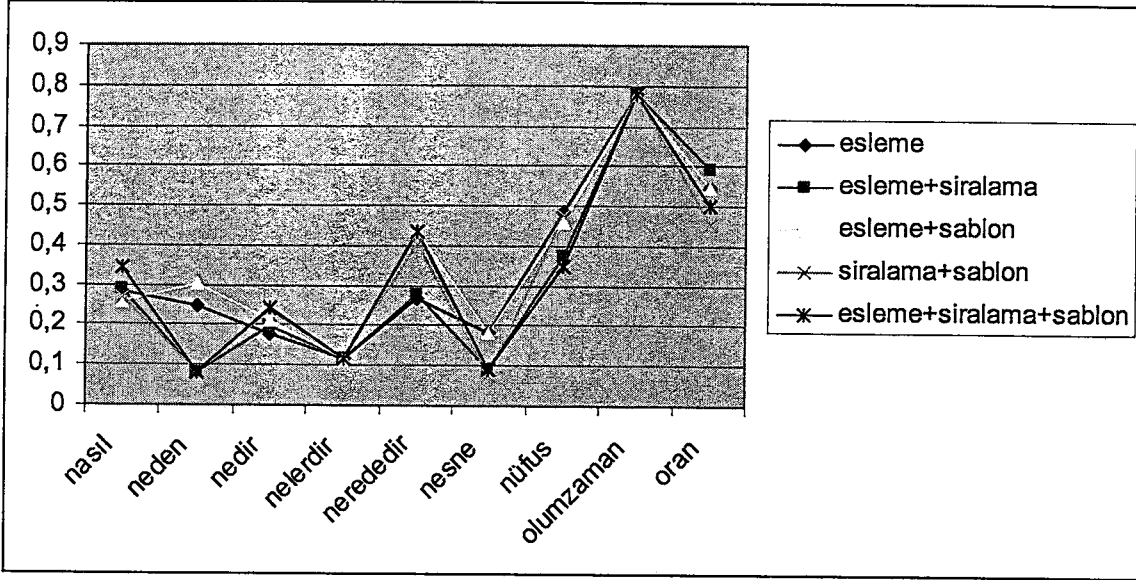
Şekil 4.2 Soru türlerinin MRR puanlarının gösterimi-1



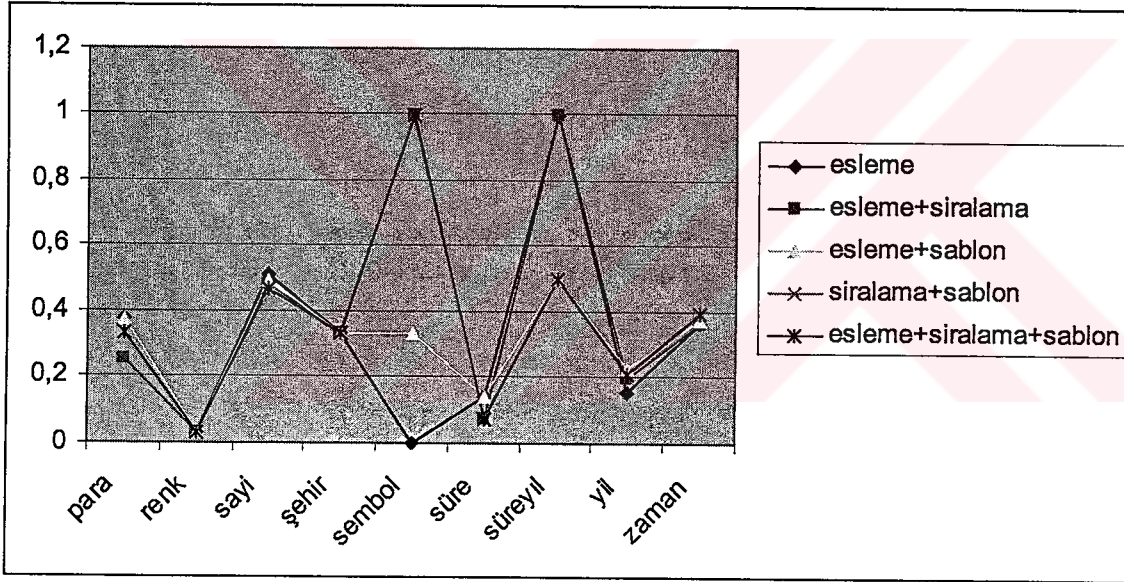
Şekil 4.3 Soru türlerinin MRR puanlarının gösterimi-2



Şekil 4.4 Soru türlerinin MRR puanlarının gösterimi-3



Şekil 4.5 Soru türlerinin MRR puanlarının gösterimi-4



Şekil 4.6 Soru türlerinin MRR puanlarının gösterimi-5

Şekil 4.2, 4.3, 4.4, 4.5 ve 4.6’da gözüken MRR puanlarına bakılarak MRR puanı yüksek olan soru türlerinin daha alçak olanlara göre; cevap formatlarının daha belirgin (klişe) olduğu ve/veya cevaplarının Internet’te daha çok yer aldığı söylenebilir.

Çizelge 4.6’da sistemin puanlama türlerinin MRR puanları ve doğru cevap oranları karşılaştırılmıştır.

Çizelge 4.6 MRR ve doğru cevap oranlarının karşılaştırılması

	MRR	Doğru cevap oranı
eşleme	0.298	42,70%
eşleme+sıralama	0.289	40,00%
eşleme+şablon	0.308	43,90%
sıralama+şablon	0.301	41,60%
eşleme+sıralama+şablon	0.313	41,98%

Çizelge 4.6 incelendiğinde sistemin; sadece cevap cümlesindeki soru kelimelerinin sayısına göre hesaplanan eşleme puanı ile çalıştırıldığında MRR puanının oldukça düşük ancak buna karşın doğru cevap oranının yüksek olduğu görülmektedir. Bunun sonucu olarak kullanıcı emeğinin artması durumunda(düşük MRR) doğru cevaba erişme ihtimalinin arttığı söylenebilir.

Puanlamaya soru kelimelerinin birbirine yakınlığı katıldığında MRR ve doğru cevaplama oranının oldukça düşük olduğu görülmektedir. Bunun Türkçe'de cümlelerdeki kelime sıralamasının önemsiz olmasından kaynaklandığı düşünülmektedir.

Eşleme puanına şablon puanı katıldığında MRR puanının ve doğru cevap oranının oldukça yüksek değerlere ulaştığı görülmektedir. Bu şablonların sisteme katılmasının ne kadar önemli olduğunun bir göstergesidir.

Sistemde mevcut olan yaklaşık 250 şablonun sayısının artırılmasının sistemin performansına yapacağı olumlu etki puanlamaya eşlemenin katıldığı her puanlama türü için sonuçların yüksek olmasından anlaşılmaktadır. İçinde eşleme puanı içermeyen puanlama türünde bile sistemin performansı oldukça yüksektir.

Sistemin performansına en büyük etkiyi eşleme ve şablon puanlarının yaptığı görülmektedir.

Sistemin performansı seçilen puanlama metotlarının hepsinin birden kullanıldığı puanlama türünde en yüksek değere ulaşmıştır. Buna dayanılarak sistemi oluşturan parçaların birbirini tamamlayıcı nitelikte oldukları söylenebilir.

5. SONUÇ ve ÖZET

Bilgiye erişimde Internet'in, Internet'teki bilgiye erişimde ise arama motorlarının önemi tartışılmazdır. Ancak arama motorlarından sadece bilginin adresi öğrenilebilmektedir. Ayrıca bu cevaba erişmek için yine bir site içinde arama yapmamız gerekmektedir. Buna ek olarak aradığımız bilgiyi belirtmek için kendi dilimizi(doğal dilimizi) değil aradığımız şeyi ifade eden anahtar kelimeler seçilmelidir. İşte tüm bu işlemleri(anahtar kelime seçme, arama motorundan gelen sonuç sayfasını inceleme, sonuç sayfasındaki linklerdeki sayfalara gitme ve ihtiyaç duyulan bilgiyi arayıp bulma) bizim yerimize yapacak sistemlere olan ihtiyaç artmıştır. Bu işlemin "niçin", "nasıl" ve "nelerdir" soruları haricindeki sorularda (nerede, ne zaman, kim, nedir vb.) gerçekleştirilmesi mümkün olmuştur.

İşte bu noktadan hareketle, bu çalışmada İngilizce için birçok örneği bulunan ancak Türkçe için bir ilk olan Internet üzerinden soru cevaplama sistemi gerçekleştirilmiştir.

Sistem kullanıcının günlük diliyle sorduğu sorusunu, önce arama motoru sorgularına dönüştürüp daha sonra Google'a göndermekte ve cevabı sonuç sayfası üzerinde yada sonuç sayfasındaki bağlantılardaki sayfalar üzerinde aramaktadır. Cevap aranırken cevap olabilecek cümleler, sayfalardan seçilip sorunun türüne göre çeşitli yöntemlerle puan verilip en yüksek puanı alan cevap yada cevaplar kullanıcıya iletilmektedir.

Sistem değerlendirilirken bu konuda yapılmış olan diğer sistemlerle karşılaştırılabilmesi için ortak soru bankasından (TREC) çeşitli sorular alınmış ve Türkçe'ye çevrilip sisteme sorulmuştur(524 adet soru). Sistemin(sadece sonuç sayfası üzerinde arama yapan hali) soruların %43.9'una(230 adet); sonuç sayfalarındaki adreslerde bulunan sayfalar üzerinde arama yapan hali ise %60'ına ilk 5 cevap içinde doğru cevap verdiği görülmüştür. Ayrıca sistemin MRR puanı 0.313 olarak ölçülmüştür. TREC-9 yarışmasına katılan sistemlerin ortalama MRR puanı 0.234 olduğu ve sistemin sadece Google'ın sonuç sayfasında cevapları aradığı ve yarışmaya katılan diğer sistemlerin sahip oldukları doğal dil işleme araçları kullanılmadığı ve Türkçe Internet içeriğinin İngilizce'ye göre daha az olduğu düşünüldüğünde sistemin ulaştığı MRR puanı iyi bir sonuç olarak değerlendirilebilir.

Sistemin başarısızlık nedenleri olarak cevabın farklı kelimelerle ifade edilmiş olması, cevabın çeşitli ekler almış olması gösterilebilir. Bu nedenleri ortadan kaldırmak için altıncı bölümde öneriler verilmiştir.

6. GELECEKTEKİ ÇALIŞMALAR İÇİN ÖNERİLER

Soru cevaplama sistemlerinin kullanıcılarına belli soru türleri için sağladığı avantajlar düşünüldüğünde üzerinde çalışılması gereken bir alan olarak karşımıza çıkmaktadır. Bu çalışmada geliştirilen sistemin elde ettiği performansın daha da yükseltilebilmesi için sistem geliştirilirken elde edilen deneyimler ışığında şunlar söylenebilir:

- Sistemde kullanılan şablonlar veritabanı daha da geliştirildiğinde sistemin performansının artacağı kuşkusuzdur.
- Sisteme yeni kelime etiketlerinin(görevlerinin) eklenmesi, şablonların ve filtrelerin dolayısıyla tüm sistemin performansını arttıracaktır.
- Aday cevapların puanlanmasında ve kullanıcı sorusunun arama motoru sorgularına dönüştürülmesinde, eş anlamlı kelimelerin yada yapıların kullanılması sistemin başarısına en büyük katkıyı sağlayacaktır.
- Google yerine Türkçe için tasarlanmış olan arama motorları içerik olarak daha da zenginleştiği takdirde başarıya olumlu bir katkısı olacaktır. Örneğin Google'da 'yüzük' ifadesi aratıldığında içinde 'yüzüklerin' olan sayfalar döndürülmemektedir. Bu durum Türkçe için tasarlanmış arama motorlarında söz konusu değildir. Örneğin www.arama.com'da bulunan arama motoru içinde 'yüzüklerin', 'yüzüklerdir', 'yüzükleri' kelimelerini içeren sayfaları da döndürmektedir. Sistemin bu tür arama motorlarında kullanılabilmesi için sadece arama sorgusundaki kelimelerin çekim eklerinden arındırılmış hallerini göndermek yeterli olacaktır.

KAYNAKLAR

- Abney S., Collins M., ve Singhal A., (2000), "Answer Extraction", <http://www.vinartus.net/spa/99f.pdf>.
- Akmajian A., Heny F. (1975), "An Introduction to the Principle of Transformational Syntax", MIT Press, Cambridge
- Antword E. L., (1990), "PC-KIMMO: a two level processor for morphological analysis".
- Brill E., Dumais S. ve Banko M., (2002), "An Analysis of the AskMSR Question-Answering System", [http:// research.microsoft.com/~brill/Pubs/EMNLP2002.pdf](http://research.microsoft.com/~brill/Pubs/EMNLP2002.pdf)
- Buchholz S., "Open-Domain Question Answering on the World Wide Web", 2002, <http://ilk.kub.nl/~antalb/abvi/week3/>.
- Charniak E., (1999), "A Maximum-Entropy-Inspired Parser".
- Daelemans R., (2002), "Learning Text Mining".
- Erkman F. ve Özil S.(1998), "Türkçe'de Niteleme", Simurg yayınları, İstanbul.
- Erkman F., (2000), "Dile genel bir bakış", Multilingual Yayınları.
- Fromkin, Victoria, Rodman ve Robert (1993), "An Introduction to Language", 17.
- Henzinger M. R., Heydon A., Mitzenmacher M., Najork M., (1999), "Measuring index quality using random walks on the Web"
- Hovy E., Hermjakob U. ve Ravichandran D., (2001), "The Use of External Knowledge in Factoid QA", www.isi.edu/~cyl/papers/TREC10-webclopedia.pdf
- İşler S., Amasyalı M. F., Tatlı E., (2000), "Bir Türkçe Metindeki Kelimelerin Cümle İçindeki Durumlarına Bakılarak Eklerine Ayrılması ", Bilişim Zirvesi'01, İstanbul
- Kwok C., Etzioni O. ve Weld D. S., (2000), "Scaling Question Answering to the Web", <http://www10.org/cdrom/papers/120/>
- Plamondon, L. ve ,Kosseim, L., (2002), "QUANTUM: A Function-Based Question Answering System".
- Pinto D, Branstein M., Coleman R., Croft W. B. ve King M., (2000), "QuASM: A System for Question Answering Using Semi-Structured Data", [http:// ciir.cs.umass.edu/pubfiles/ir-244.pdf](http://ciir.cs.umass.edu/pubfiles/ir-244.pdf)
- Robertson S.E. ve Walker S., (1998), "Okapi/Keenbow at TREC-8", In Proceedings of The Eighth Text Retrival Conferanse, 151-162.
- Zheng, Z. 2002, "Developing a Web-based Question Answering System", <http://www2002.org/CDROM/poster/203/>.

ÖZGEÇMİŞ

Doğum tarihi	17.1.1978	
Doğum yeri	İstanbul	
Lise	1991-1994	Fatih Gelenbevi Lisesi
Lisans	1996-2001	Yıldız Üniversitesi Mühendislik Fak. Bilgisayar Mühendisliği Bölümü

Çalıştığı kurum

2001-Devam ediyor YTÜ Bilgisayar Mühendisliği Bölümü Araştırma Görevlisi

Yayınlar**Yayınlanmış Bildiriler:**

Tatlı E. İ., Amasyalı M. F., Karslıgil M. E., (2000), "Recognition of Human Face Expression", ISCIS XV, İstanbul

İşler S., Amasyalı M. F., Tatlı E., (2000), "Bir Türkçe Metindeki Kelimelerin Cümle İçindeki Durumlarına Bakılarak Eklerine Ayrılması ", Bilişim Zirvesi'01, İstanbul

Diri B. ve Amasyalı M. F., (2003), "Automatic Authorship Attribution in Turkish Language", ICANN/ICONIP 2003, İstanbul

Amasyalı M.F., Erdem N., Haberdar H. ve Koyuncu F., (2003), "Neuro-Chain: A Handwritten Character Recognition System", TAINN 2003, Çanakkale

Albayrak S. ve Amasyalı M. F., (2003), "Fuzzy C-Means Clustering on Medical Diagnostic Systems", TAINN 2003, Çanakkale

Yüksel Ç., Çabuk H., Mocan Z. ve Amasyalı M. F., (2003), "Metin Analizi ve Sorgulama", SIU 2003, İstanbul