

# SORU CEVAPLAMA ÜZERİNE BİR ARAŞTIRMA

Müge Sayıt\* ve Bahar Karaođlan\*

(\*) Ege Üniversitesi, Uluslararası Bilgisayar Bölümü, 35100, İZMİR  
[muge.fesci@ege.edu.tr](mailto:muge.fesci@ege.edu.tr), [bahar.karaoglan@ege.edu.tr](mailto:bahar.karaoglan@ege.edu.tr)

## ÖZET

“Ege Üniversitesi ne zaman kurulmuştur?”, “Bir üzüm salkımında ne kadar kalori vardır?” türünde sorulara cevap veren bilgi geri getirim sistemleri, 1970li yılların başından itibaren geliştirilmeye başlanmıştır. Bu makalede, günümüzde üzerinde bir çok çalışma gerçekleştirilen bu alanda yapılan çalışmalar, geliştirilen soru cevaplama sistemleri ve bu sistemlerin değerlendirme teknikleri incelenmektedir.

**Anahtar Kelimeler:** Bilgi Geri Getirimi, Soru Cevaplama Sistemleri, Soru Cevaplama Değerlendirme Teknikleri.

## A SURVEY ON QUESTION ANSWERING

### ABSTRACT

Information retrieval systems answering questions such as “When was Aegean University established?”, “How many calories are there in grape?”, were started to develop in the beginning of 1970s. In this paper, we investigate question answering systems and their evaluation techniques.

**Keywords:** Information Retrieval, Question Answering Systems, Question Answering Evaluation Techniques.

### 1. GİRİŞ

Soru cevaplama, bir bilgi geri getirim sistemine yöneltilen sorulara cevap döndürülmesidir. Bu yaklaşımla geliştirilmiş bilgi geri getirim sistemi ise soru cevaplama sistemi olarak adlandırılır. Soru cevaplama sistemleri, doğal dil ile sorulan sorulara, döküman uzayında yapılan araştırma sonucu uygun cevaplar döndürmeyi hedefler. 70li yılların sonunda geliştirilen STUDENT ve LUNAR problem çözme sistemleri, ilk soru cevaplama sistemleri olarak sayılabilirler [8]. Bir soru cevaplama sistemine yöneltebilecek sorular gerçek tabanlı sorular (factoid questions) ya da listeleme soruları (list questions) olabilir. Bunlar dışında TREC seminerlerinde kullanılan bir soru tipi, “diđer” sorular (other questions) vardır; bu sorular daha önce yöneltilen sorularla hedef olan cevap elde edilememişse yöneltilen ve cevapla ilgili ek bilgi içeren soru tipleridir [15]. Gerçek tabanlı sorulara örnek olarak “Türkiye’nin başkenti neresidir?”, “Altın oran kaçtır?” tipi sorular, liste sorularına örnek olarak ise “Türkiye’nin bölgelerini listeleysin” tipi sorular verilebilir.

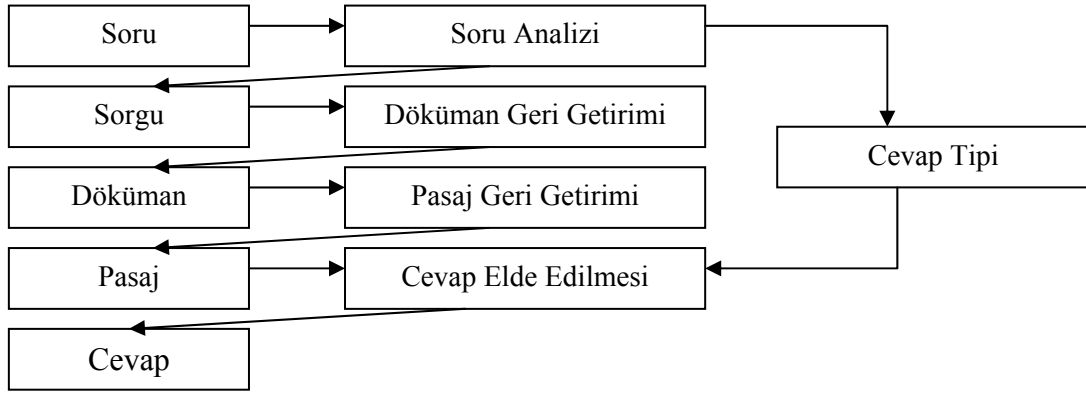
Gerçek tabanlı soru cevaplama sistemlerinin genel yapısında ilk olarak soru analiz edilir, daha sonra döküman ya da pasaj geri getirmesi gerçekleştirilir, son olarak cevap elde edilir [13]. Makalenin izleyen bölümleri şu şekildedir: 2. bölümde soru cevaplama aşamaları, 3. bölümde soru cevaplama sistemleri, 4. bölümde değerlendirme teknikleri ve 5. bölümde sonuçlar yer almaktadır.

### 2. SORU CEVAPLAMA AŞAMALARI

Soru cevaplama sistemleri, soru analizi, döküman geri getirmesi, pasaj analizi ve cevap elde edilmesi aşamalarından oluşmaktadır. Bu bölümde bu aşamalarla ilgili genel bilgilerden ve kullanılan tekniklerden söz edilecektir. Bir soru cevaplama sisteminin aşamaları Şekil 1’de gösterilmiştir.

#### 2.1. Soru Analizi

Soru cevaplama sisteminde ilk aşama olan soru analizinde, sorunun amacını anlamak ve ne tür bir soru olduğuna karar vermek hedeflenir. Bunun için bir çok sistemde, sorular belli sınıflara ya da sınıflara dahil edil-



Şekil 1 Soru Cevaplama Aşamaları

mektedir. Soru sınıfları, beklenen cevabın tipine bağlı olarak belirlenir; bu sınıflar arasında kişi, doğum ve/veya ölüm tarihi, yer, başkent vb. sayılabilir. Soru analizi aşamasında sorunun hangi sınıfa dahil olduğunu bulmak için en sık ve kolayca uygulanan yöntem, desen (pattern) kullanmaktır. Desen, her bir soru sınıfı için önceden tasarlanmış olan yapılardır. Örneğin ‘Kim’, ‘Kimi’ ya da ‘Kimin’ ile başlayan ya da bu kelimeleri içeren sorular kişi sınıfına dahildir ve bu soru sınıfının deseni “Kim .... , Kimi ..., Kimin ..., ...kim...” şeklinde tanımlanabilir.

Bazı soru cevaplama sistemlerinde, sorudaki terimlerin dilbilgisinde fiil, özel isim, nesne gibi hangi sınıfa dahil olduğunu belirledikten sonra desenler ile karşılaştırarak soru sınıflandırması yapılmaktadır.

Soru sınıflandırma tamamlandıktan sonra soru üzerinde geri getirim için gerekli olan işlemler gerçekleştirilir. Bu işlemlerin amacı soruyu, geri getirim sistemi tarafından anlaşılabilir soru yapısına dönüştürmektir. Örneğin doğal dilde soru kabul etmeyen bir sistem için sorunun değiştirilmesi bu aşamada gerçekleştirilir. Soru oluşturulduktan sonra soru cevaplama için ikinci aşama, döküman geri getirim aşamasına geçilir.

## 2.2. Döküman Geri Getirimi

Gerçek tabanlı sorular için metinlerin kısa olması cevap çıkarım aşamasının daha başarılı bir şekilde sonuçlanmasını ve uçtan uca performansın artırılmasını sağladığı bilinen bir gerçektir [14]. Döküman ya da fazla sayıda cümleden oluşan pasajlar geri getirildiğinde, bu dökümanlar arasından cevap çıkarma aşaması daha güç

gerçekleşmektedir. Bu sebeple, döküman geri getirim aşaması döküman uzayından konu ile ilgili dökümanların geri getirilmesini sağlar, daha sonraki aşamada bu dökümanlar arasından cevap çıkarma işlemini kolaylaştıracak şekilde pasajlar geri getirilir. Çok fazla döküman geri getirildiğinde ise kapsama (geri getirilen dökümanlar içerisinde cevapları bulunan soru sayısı) arttırılmış olmasına rağmen tüm sistem performansının düştüğü gözlemlenmiştir [4]. Bunun sebebi, fazla sayıda döküman geri getirmenin yüksek oranda gürültü taşımasıdır. Bu yüzden yüksek oranda kapsamanın yanı sıra yüksek hassaslık derecesine (precision) sahip olduğu takdirde üst sıralarda cevap içeren dökümanlar geri döndürülecektir.

Döküman geri getiri için kullanılan yöntemler, geri getirilen dökümanların verilen sorgu için cevap içermesi gerektiği nedeniyle klasik bilgi geri getirim algoritmalarından farklıdır. Bu yüzden, bilgi geri getirim aşamasında sorgu, cevap içeren dökümanlar döndürmeye yönelik bir şekilde oluşturulduktan sonra klasik bir bilgi getirim yöntemi ya da direk olarak soru cevaplama yönelik bir bilgi geri getirim sistemi kullanılabilir. Soru cevaplama yönelik bilgi geri getirim sistemleri, dökümanlar içinde eş anlamlı sözcüklerin belirlenmesi ve birbirine bağlanması (linking), sözdizimsel inceleme (syntactic parsing), desen eşleştirme (pattern matching) gibi analizler yapılmasını gerektirmektedir.

Döküman geri getirim aşamasında geri getirilen dökümanlar arasında sorunun cevabını içeren bir döküman bulunmuyorsa, sistemin diğer aşamalarının performansı ne kadar iyi olursa olsun, soru doğru bir şekilde cevaplandırılmayacaktır. Dolayısıyla bu

aşamanın bir soru cevaplama sistemindeki en önemli aşama olduğu söylenebilir. Döküman geri getirmesi muhtemel pasajların sıralanması için bir sonraki aşamaya, cümle geri getirmine geçilir.

### 2.3. Pasaj Analizi

Pasaj, paragraf benzeri metin parçalarıdır. Aynı sayıda cümle sayısı içerir ya da her bir cümle bir pasaj olarak değerlendirilebilir. Sistemdeki ve cevap çıkarım stratejisindeki sorgulama yöntemi daha önemli olduğu için pasajın ne kadar uzunlukta olacağı konusunda çok fazla araştırma yapılmamıştır. Pasaj geri getirmesi aşamasında, bir önceki aşamada döndürülen dökümanlar arasından pasajlar seçilerek cevap içerme ihtimalinin yüksekliğine göre sıralanmaktadır.

Pasaj geri getirmesi için kullanılan bazı algoritmalar şunlardır:

MITRE:

Herbir cümlenin ayrı bir pasaj olarak sayıldığı bu algoritmada, sorgu ve cevap adayları arasındaki ortak kelimeler sayılır [12].

“Word overlap” adı verilen bu yöntemde, cevap adayları olarak cümleler seçilmektedir. Ağırlıklandırılmalı “word overlap” yönteminde ise sorgudaki kelimelere farklı ağırlıklar verilir [6]. Cevap adayları ile ortak kelimeler bulunduğu bu ağırlık göz önüne alınarak hesaplama ve daha sonra sıralama (ranking) işlemi gerçekleştirilir.

Bm25:

Okapi bm25 algoritması pasaj geri getirme algoritmaları arasında en sık kullanılan algoritmadır. Bu algoritma, döküman geri getirmesinde de kullanılan Okapi bm25 benzerlik ölçümüne dayanır [11]. Döküman uzayında N adet belge bulunan bir koleksiyon için bu benzerlik ölçütü formül 1’de verilmiştir.

$$bm25(q, d) = \sum_{t \in q} \log\left(\frac{N - f_t + 0,5}{f_t + 0,5}\right) \times \left(\frac{(k_1 + 1)f_{d,t}}{K + f_{d,t}}\right) \dots\dots\dots(1)$$

Formülde yer alan t, q sorgusundaki terimleri temsil etmektedir.  $f_t$ , t terimini içeren döküman sayısı,  $f_{d,t}$  ise t teriminin d

dökümanındaki (ya da pasajındaki) frekansını vermektedir. K değeri formül 2 ile hesaplanır.

$$k_1((1 - b) + b \times L_d / AL) \dots\dots\dots(2)$$

Formül 2’de yer alan  $k_1$  ve b parametreleri 1,2 ve 0,75 olarak alınır.  $L_d$  d dökümanının (pasajının) uzunluğu ve AL tüm dökümanların (pasajların) ortalama uzunluğudur.

MultiText:

MultiText pasaj geri getirmesinde yüksek idf (inverse document frequency – ters döküman frekansı) değerine sahip çok sayıda terim içeren kısa pasajlar üst sıralarda geri döndürülür. Algoritmadaki her bir pasaj penceresi sorgu terimiyle başlar ve yine bir sorgu terimi ile biter. Bu algoritmanın pasajlar arasında yaptığı puanlamada, pasaj içerisindeki sorgu terim sayısı kadar pencere boyutu yani sorgu ile ortak kelimeler arasındaki mesafe de önemlidir.

MultiText algoritmasının standart idf ya da terim ağırlıkları için idf kullanan iki versiyonu bulunmaktadır [3,12].

IBM:

IBM pasaj geri getirmesinde bir pasaj için benzerlik ölçümleri ile ilgili hesaplar yapılır. Eşleşen kelime ölçümü (matching word measure) hem sorguda hem de pasajda görülen kelimelerin idf değerlerinin toplamıdır. Kavramlar dizini benzeme ölçümü (thesaurus match measure), pasajda sorgu ile eş anlamlı olan kelimelerin idf değerlerinin toplamıdır. Eşleşmeyen kelime ölçümü (mis-match word measure) sorguda bulunan ama pasaj içerisinde bulunmayan kelimelerin idf değerleri toplamıdır. Dağılım ölçümü (dispersion measure), pasajda bulunan sorgu terimleri arasındaki kelimeleri; öbek kelime ölçümü (cluster word measure) pasajda bulunan sorgu ile ortak ve bitişik (adjacent) kelimeleri yani kelime guruplarını sayar. Tüm bu ölçümler hesaplandıktan sonra bir seri oluşturacak şekilde birleştirilir ve pasajlar bu serilere bakılarak sıralanır.

### 2.4. Cevap Elde Edilmesi

Pasaj geri getirmesi aşaması gerçekleştirildikten sonra geri bildirilen ve cevabı içermesi

muhtemel pasajlar arasından cevap seçilmesi aşaması gerçekleştirilir. Bu aşamada sistemlerin her zaman tek bir cevap döndürmesi beklenmez, sıralı cümle listesi de cevap olarak döndürülebilir.

Cevabı içermesi muhtemel pasajlar arasında birinin diğerine göre üstünlüğü konusunda ilk kriter, pasaj analizi sonuçlarıdır, üst sıralarda geri bildirilen pasajların doğru cevabı içermesi daha büyük olasılığa sahiptir. Aday cevaplar sıralamasında bir diğer kriter ise, aday cevabın soru ile bağlantılı olduğuna karar verilirken hangi yolun kullanıldığıdır. Eğer bir cevap adayının belirlenmesinde soruyla bağlantı kurulurken uygun bir desen, inceleme ağacı (parse tree) ya da bağımlılık grafi (dependency graph) bulunmuşsa bu cevap adayı, soru ile sadece kelime benzerliği bulunan cevap adaylarından daha üst sırada yer alır. Hangi cevabın soru ile ilişkili olduğunda, aday cevabın frekansı da cevap çıkarmada bir kriter olarak değerlendirilir. Aday cevabın frekansı, yapılan geri getirimler arasında o cevabın soru ile kaç kez bağlantılı bulunduğu ve bu şekilde cevap çıkarma tekrar tabanlı cevap çıkarma olarak adlandırılır [3]. Bu yöntem, sürekli değişen ve gelişen döküman uzayına rağmen aynı soruya aynı cevap döndürülüyorsa, bu cevabın yüksek ihtimalle doğru olduğu prensibine dayanır.

Eğer sistemden tek bir cevap döndürmesi bekleniyorsa, sıralamada en üst sıradaki cevap döndürülür. Pasaj analizi sonucunda elde edilen aday cevapların frekansı düşük ya da yeterli sayıda değilse, veya hiçbir aday cevap döndürülemediyse, cevap çıkarma bileşeni sistemi soru analizi bileşenine döndürür. Bu durumda, sorgudaki bazı terimler silinebilir ya da ilişkili yeni terimler eklenebilir ve sorgulama tekrar yapılır [8]. Böylelikle farklı döküman ve pasaj seti geri getirilebilir ve aday cevabı arama bu yeni döküman uzayı içerisinde yapılır.

Eğer sorunun birden fazla cevabı varsa, ki bu durum ile özellikle listeleme benzeri sorularda karşılaşılabılır, sistemin doğru cevabı geri döndürme olasılığı artmaktadır. Örneğin “Pınar firmasının ürünleri nelerdir?” gibi bir soru için birden fazla cevap bulunmaktadır. [6] TREC-8 soru cevaplama sistemi üzerinde yaptığı deneylerde cevap

fazlalığının sistem performansını arttırdığını ölçmüştür.

Sorulan soruya kısa bir cevap döndürülmesi bekleniyorsa, soru cevaplama aşamasında karşılaşılabilecek en önemli sorun döndürülen pasajlar ya da cümleler içinde aynı cevap tipinde birden fazla (kişi tipinde cevap döndürmesi istenen soru için aday cevap cümle içerisinde 2 ya da daha fazla kişi adı bulunması gibi) varlık bulunmasıdır. Bu durumda sistemin doğal dil işleme yeteneği yoksa cevap olarak bu varlıklardan biri rastgele olarak seçilir.

### 3. SORU CEVAPLAMA SİSTEMLERİ

#### 3.1. AnswerBus

AnswerBus açık alanlı (open-domain) cümle seviyesinde bilgi döndüren web üzerinde çalışan bir soru cevaplama sistemidir. Sistem, İngilizce, Almanca, Fransızca, İspanyolca, İtalyanca ve Portekizce dillerinde sorulmuş sorulara İngilizce karşılık verebilmektedir. Yöneltilen sorulara 5 arama motoru kullanılarak (Google, Yahoo, WiseNut, AltaVista ve Yahoo News) Web sayfaları üzerinden doğru cevaplar aranmakta ve sonuçlar birkaç saniye içerisinde döndürülmektedir [17].

Kullanıcı sorularının doğal dilde yöneltilmesini sağlayan AnswerBus, ilk olarak dil tanıma modülünü çalıştırır; eğer soru İngilizce dışında bir dil ise BabelFish tercüme programı kullanılarak İngilizce'ye çevrilir. Bu adımdan sonra seçilen 2 ya da 3 arama motoru seçilir ve sorgu yeniden biçimlendirilir. Seçilen arama motorlarına yöneltilen sorgu ile bilgi geri getiri yapılır. Geri getirilen dökümanlar arasından cevap içermesi muhtemel cümleler seçilir. Son olarak çıkarılan cevaplar cümleler halinde kullanıcıya geri döndürülür.

AnswerBus arama motorlarını seçerken sorgunun tipine göre bir seçim gerçekleştirmektedir. Örneğin güncel konularla ilgili bir soru yöneltildiğinde, Yahoo News, Google'dan daha iyi sonuç verecektir. Bu şekilde araştırma yapmak üzere 5 tane içerisinden 2 ya da 3 arama motoru seçimine karar verilmektedir. Seçim yapıldıktan sonra sorgunun yeniden biçimlendirilmesinin nedeni, her arama motorunun doğal dil kabul etmemesi ya da daha iyi sonuçlar elde edilecek biçime

getirilmesidir. Bunun için sorgudan bazı kelimeler çıkarılabilir ya da eklenebilir; bazı kelimelerin eklerinde deęişim yapılabilir.

Arama motorlarından bilgi geri getirmeye yapıldıktan sonra gerçekleştirilen cevap çıkarma aşamasında ise, geri getirilen cümlelerin cevap içerip içermediğini anlamak için benzeyen kelimeler ve benzemeyen kelimeler ölçümleri kullanılır. Formül 3 geri getirilen cümleleri filtreleyen formüldür.

$$q \geq \lfloor \sqrt{Q-1} \rfloor + 1$$

.....(3)

Formülde q cümle ile sorgu arasındaki ortak kelime sayısı, Q ise sorgudaki kelime sayısıdır. Örneğin bir sorgu 3 kelime içeriyorsa bunların en az iki tanesi cümlede bulunmalıdır. Bu formülü sağlayan cümleler içerdiği ortak kelime sayısına göre puan alırken, formülü sağlayamayan cümlelerin puanı 0 olarak değerlendirilir.

### 3.2. YorkQA

YorkQA sistemi ilk olarak cümlelerin anlamlarını ifade eden yapılardan oluşan bir veri tabanı oluşturur (döküman indeksleme). Sonraki adımda soru ile ilgili yapılar kümesi seçilir. Soru ile ilgili yapı seçiminde önerilen cevap ile soru arasındaki sözcüksel benzerliğe önem verilmektedir. Son olarak seçilen yapılar tekrar sorgu ile karşılaştırılır [2].

YorkQA sisteminde kullanılan algoritmanın adımları şunlardır:

~ Bir bilgi geri getirme sistemi kullanılarak dökümanlar indekslenir.

~ Sırayla yöneltile tüm sorular okunur.

~ Her soru sınıflandırılmak üzere incelenir. Sınıflandırma cevabın içerdiği özelliklere yani varlık sınıfına göre gerçekleştirilmektedir (Soru analizi aşaması).

~ Bilgi geri getirme sistemine uygun bir şekilde sorgular yeniden biçimlendirilir ve gönderilir (döküman geri getirme aşaması).

~ Geri alınan dökümanlar cümle bölüştürme (sentence splitter), etiketleme (tagger), ve varlık tanımlama (named entity recognizer) (tarih, yer, insan ismi gibi varlıklardan biri mi) işlemlerinden geçer.

~ Cümlelerin beklenen cevap tipinde varlık sınıfı içerip içermediği kontrol edilir.

~ Cevap tipinde varlık sınıfı içeren cümleler, soru ile anlamsal uzaklığına göre sıralanır. Bunun için WordNet kullanılır ancak birebir benzerlik dışında kavramsal benzerlikler de dikkate alınır.

~ Soru ile sıralanan cümleler arasındaki benzerliği bulmak için bir ayrıştırıcı (parser) kullanılır.

~ Eğer ayrıştırıcı, sıralanmış listede üst sıralarda uygun varlık tipi bulunmadığına karar verirse, sorgu 50 byte metin uzunluğunda yakın anlamlı kelimelerle genişletilir ve yeniden arama yapılır.

### 3.3. QUALIFIER

“Everest Dağının yükseliği ne kadardır?”, “Türkiye’nin en uzun nehri hangisidir?” gibi sorulara kullanıcılar kısa cevaplar beklemektedir. Ancak cevap olarak uzun cümleler döndüren bazı soru cevaplama sistemleri kullanıcıların bu beklentisine tam olarak cevap verememektedir. QUALIFIER sistemi sistemlerin bu eksikliğini gidermek amacıyla geliştirilen sistemlerden birisidir. Bu tip sistemler doğal dil işleme ve bilgi çıkarımı yöntemlerini kullanır [16].

QUALIFIER olay tabanlı bir soru cevaplama sistemidir. Sistemde, her şey varlık ya da olay olarak sınıflandırılmıştır. Sorular da bu şekilde varlığa yönelik ya da olaya yönelik olarak değerlendirilmiştir. Varlığa yönelik sorular tanım soruları olarak adlandırılır. Örneğin “Ege Üniversitesinin sembolü nasıldır?” sorusu varlığa yönelik bir sorudur. Olaylar ise yer, zaman, nesne, obje, özellik, tanım, hareket gibi birden fazla sayıda elemandan oluşur.

QUALIFIER sistemi soru analizi, ek kaynak kullanarak sorgu genişletme, döküman geri getirmeye ve cevap çıkarma aşamalarını gerçekleştirir. Soru analizinin amacı soru içerisine gömülü bilinen soru cevaplama elemanlarını ve beklenen cevap tipini çıkarmaktır. İnce taneli isimlendirme varlıkları kullanan QUALIFIER soru sınıflandırıcı ile hedef cevap tipini belirler [16].

Cevap çıkarma aşamasında aday cevapların içerdiği varlıklar belirlenir, daha sonra uygun cevap tipi varlığı içeren cümleler belirlenir.

İnce-taneli (fine grained) isim varlıkları Tablo 1’de listelenmiştir. Sorgu yeniden yapılandırma aşamasında QUALIFIER hem Web hem de WordNet’e dayalı olarak sorguya ek bilgi eklemektedir.

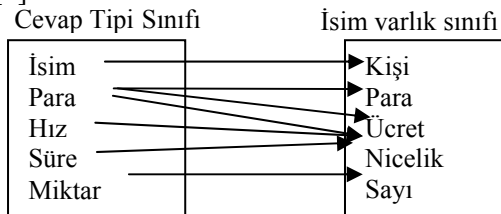
Tablo 1 İnce-taneli isimlendirme varlıkları[16]

Kişi	Organizasyon, insan
Zaman	Gün, yıl, ay
Yer	Şehir, ülke, ada, nehir, dağ vb.
Sayısal	Yaş, alan, derece, boyut, hız vb.
Nesne	Hayvan, renk, oyun, dil, din vb.

### 3.4. Diğer Sistemler

Ramakrishnan et. al [10] tarafından gerçekleştirilen soru cevaplama sistemi, yöneltilen soruları ve cevaplarını analiz ederek cevap pasajlarını elde etmeyi ve sıralamayı öğrenen bir sistemdir. Bu sistemde kullanılan ana teknik, yapısal sorgu oluşturulurken soruda yer alan parçalara dikkat etmektir. Bu parçalardan bir tanesi, cevapta olması muhtemel soru kısımlarıdır. Diğer bir parça cevap tipinin belirlenmesine yardımcı soru kısımlarıdır. Sisteme, [1] ve [5] te bulunan soru cevap çiftleri öğretmiştir. Farklı külliyat ve farklı dillere de adapte edilebilen bu sisteme öğretilecek soru cevap çiftleri internet üzerinden de bulunabilmektedir.

Paşca [8] et. al tarafından gerçekleştirilen soru cevaplama sisteminin özellikleri geniş bir cevap tipi sınıflandırmasına sahip olması, pasaj geri getirmenin gereken durumlarda tekrarlanabilmesi, sözcüksel anlamların dikkate alınması, makine öğrenme tekniklerinin kullanılması ve cevap yakalamadır. Cevap tipi sınıflandırmada Şekil 2’de görüldüğü gibi cevap tipi birden fazla isim varlığına denk gelebilir. Doğal dilde sorulmuş bir soru için çıkarılan ayrıştırma ağacı (parse tree), soru kelimeleri arasındaki bağımlılıkları da göstermektedir [8].



Şekil 2 Cevap ve isim varlık tipleri eşleştirmesi

Bu sistemler dışında Falcon, Webclopedia, Mulder gibi sistemler iyi bilinen soru cevaplama sistemleri arasında yer almaktadır.

### 4. DEĞERLENDİRME

Soru cevaplama sistemlerini değerlendirmek için çeşitli teknikler kullanılmaktadır. Bunlardan biri TREC seminerlerinde kullanılan ve soru cevap ikilisine uzmanlar tarafından 4 ayrı değerden bir tanesinin verilmesiyle yapılan değerlendirme yöntemidir. Bu değerler, doğru, tam cevap değil, desteksiz ve yanlıştır. Tam cevap değil kararı, döndürülen cevap gereğinden fazla kelime içeriyorsa verilir. Desteksiz kararı ise döndürülen cevap ise gerçek cevapla ilgili ancak net değilse verilir [7].

Bunun dışında kullanılan temel değerlendirme teknikleri arasında tek cevapta doğruluk FHS (First Hit Success) (soruya döndürülen cevaplar arasından ilki alınır, doğru ise 1, yanlış ise 0 değeri verilir) ve MRR (Mean Reciprocal Rank-Ortalama İki Taraflı Sıralama) sayılabilir. Buradaki sıra sistem tarafından geri döndürülen cevaplar arasında doğru cevabın bulunduğu ilk sıradır. Örneğin bir sistem, yöneltilen soruya verilecek ilk doğru cevabı 4. sırada getirmişse, FARR (First Answer Reciprocal Rank) değeri 1/4 olarak hesaplanır. Eğer döndürülen cevaplar arasında doğru cevap yer almıyorsa FARR değeri 0 olarak belirlenir. MRR değeri hesaplamak için Formül 4 kullanılır [8]. Formüldeki 1/s<sub>i</sub> değeri, herbir cevap için FARR değeridir.

$$MRR = \frac{1}{n} \left( \sum_{i=1}^n \frac{1}{s_i} \right) \dots\dots\dots$$

(4)

FARWR (First Answer Reciprocal Word Rank) değerlendirme tekniğinde, döndürülen cevapta doğru cevabın başladığı pozisyona bağlı olarak hesaplanır. Örneğin “İzmir hangi bölgededir?” sorusuna Türkiye’de Ege Bölgesi’nde şeklinde döndürülen cevapta Ege 2. kelime olduğu için FARWR değeri 1/2’dir. TRWR (Total Reciprocal Word Rank) döndürülen tüm doğru cevaplar için FARWR değerlerinin toplanmasıyla hesaplanır. PRECISION değerlendirme tekniğinde döndürülen doğru cevapların karakter uzunluğu, sistemde var olan tüm cevapların toplam karakter uzunluğuna bölünerek elde edilir [9].

## 5. SONUÇ

Soru cevaplama sistemleri ile ilgili yapılan bu araştırmada sistemlerin genel özellikleri ve kullandıkları yöntemler incelenmiştir. Soru cevaplama sistemlerinin etkinliği ile ilgili olarak her yıl TREC seminerlerinde değerlendirme yapılmakta ve sistemlerin performansları incelenmektedir.

Soru cevaplama sistemleri, doğal dil işleme ve öğrenme tekniklerini kullanarak özellikle son yıllarda verilen önemle birlikte büyük bir gelişme göstermektedir. Bu sistemleri desteklemek üzere geliştirilen sorgu - cevap desenleri ve oluşturulan modeller yakın zamanda daha başarılı soru cevaplama sistemlerinin geliştirilebilmesine olanak sağlayacaktır.

Soru cevaplama sistemlerinin performansı analitiksel yöntemlerle cevap döndürmeye dayalı olduğu sürece belli bir sınırdadır. Doğal dil işleme konusundaki gelişmelerle birlikte soruların mantıksal olarak değerlendirilmesi ve buna uygun mantıksal cevaplar döndürmesi durumunda, gelecekte geliştirilecek soru cevaplama sistemleri başarılı birer yapay zeka uygulaması olarak görülebilir.

## 6. KAYNAKLAR

- [1]. Agichtein, E.S., Gravano, L., “Learning search engine specific query transformations for question answering”, *Proceedings of the 10th World Wide Web Conference (WWW10)*, 169-178, 2001.
- [2]. Alfonseca, E., Boni, M., Valencia, J., Manandhar, M., “A ProtoType Question Answering System Using Syntactic and Semantic Information for Answer Retrieval”, *TREC 10*, 2001.
- [3]. Clarke, C., Cormack, G., ve Tudhope, E., “Relevance ranking for one to three term queries”, *Information Processing and Management*, 291–311, 2000.
- [4]. Gaizauskas, R., Greenwood, M.H., Hepple, M., Roberts, I., Saggion, H. ve Sargaison, M., “The university of sheffield’s trec 2003 Q&A experiments”, *Proceedings of the Twelfth Text REtrieval Conference (TREC 2003)*, 2003.
- [5]. Lam, S.K., Pennock, D.M., Cosley, D., Lawrence, S., “1 billion pages = 1 million dollars? Mining the web to play who wants to be a millionaire?”, *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, 337-345, 2003.
- [6]. Light, M., Mann, G., Rilo, E., Breck, E., “Analyses for elucidating current question answering technology”, *Journal of Natural Language Engineering, Special Issue on Question Answering*, 2001.
- [7]. Lin, J., “Evaluation of Resources for Question Answering”, *SIGIR*, 392-299, 2005.
- [8]. Pasca, M.A., ve Harabagiu, S.M., “High performance question / answering”, *Research and Development in Information Retrieval*, 2001, 366-374.
- [9]. Radev, D.R., Qi, H., Wu, H., Fan, W., “Evaluating Web based Question Answering System”, *Demo section, LREC 2002*, Haziran 2002, Las Palmas, Spain.
- [10]. Ramakrishnan, G., Chakrabarti, S., Paranjpe, D., ve Bhattacharya, P., “Is question answering an acquired skill?”, *Proceedings of the 13th international conference on World Wide Web*, 111–120, ACM Press, 2004.
- [11]. Robertson, S.E., Walker, S., Hancock-Beaulieu, M.M., Gull, A., Lau, M., “Okapi at TREC”, *Text Retrieval Conference*, 21-30, 1992.
- [12]. Tellex, S., Katz, B., Lin, J., Fernandes, A., Marton, G., “Quantitative Evaluation of Passage Retrieval Algorithms for Question Answering”, *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informtaion retrieval table of contents*, 41-47, 2003.
- [13]. Terra, E., Collins-Thompson, K., Callan J., ve Clarke, C.L.A., “The Effect of Document Retrieval Quality on Factoid Question Answering Performance”, *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2004.
- [14]. Usunier, N., Amini, M.R., Gallinari, P., “Boosting Weak Ranking Functions to Enhance Passage Retrieval for Question Answering”, *SIGIR 2004 workshop on Information Retrieval for Question Answering*, 1-6, 2004.
- [15]. Voorhees, E.M., “Overview of the TREC 2004 Question Answering Track”,

- TREC 2004, Kasım 16-19, 2004, Gaithersburg, Maryland.
- [16]. Yang, H., Chua, T., Wang, S., Koh, C., “National Structured use of external knowledge for event-based open domain question answering”, SIGIR, 33-40, 2003.
- [17]. Zheng, Z., “AnswerBus question answering system”, In Human Language Technology Conference, 2002, Amerika.