

METİN ANALİZİ VE SORGULAMA (MAvS)

Hüseyin Çabuk¹, Çiğdem Yüksel², Zeki Mocan³, Banu Diri⁴, M.Fatih Amasyalı⁵

Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, İstanbul

¹ husmanc@yahoo.com
² cigdem_y79@hotmail.com
³ zekmcn@hotmail.com
⁴ banu@ce.yildiz.edu.tr
⁵ mfatih@ce.yildiz.edu.tr

Özetçe

Bu çalışma, Yapay Zeka'nın bir alanı olan doğal dil işleme (NLP Natural Language Processing) teknolojisi kapsamında gerçekleştirilmiştir. Birçok dilde uygulanan doğal dil işleme tekniklerinin, dünyada en yaygın altıncı dil olan ve Asya ile Avrupa'da yoğun olarak konuşulan Türkçe'de de kullanılması araştırma alt yapısı oluşturması açısından önemlidir. Bu çalışmada, internet üzerinden girilen bir metindeki cümleleri analiz eden, kullanıcıya cümlelerle ilgili sorular sunan ve kullanıcının verdiği cevabı inceleyerek doğruluğunu kontrol eden bir sistem geliştirilmiştir.

1. Giriş

Bilgisayar teknolojilerindeki gelişmelerin ilk yıllarından itibaren insanlar, kendi aralarında kurdukları iletişimin bir benzerini bilgisayarlar ile sağlamayı istemişlerdir. İnsan tarafından yapıldığında zeka olarak adlandırılan davranışların makine tarafından da yapılması hedeflenmiştir. İnsan aklının nasıl çalıştığını gösteren bir kuram olan Yapay Zeka'dan yararlanılmıştır. Bu kuram ile çeşitli problemlerde optimal çözümü bulabilen bilgisayar yazılımları geliştirilmiştir.

Yapay Zeka'nın bir alanı olan Doğal Dil İşleme, ana işlevi doğal bir dili çözümlene, anlama, yorumlama ve üretme olan bilgisayar sistemlerinin tasarımını ve gerçekleştirilmesini konu alan bir bilim ve mühendislik alanıdır. Doğal dil işlemenin uygulamalarında; makinelerin doğal insan dillerini kullanabilmesi, yazılan metinleri anlayarak doğal dilde cevap verebilmesi, veritabanları ve robot sistemleri arasında arabirim oluşturma gibi konular üzerinde çalışılmaktadır. Dil tabanlı yapay zeka araştırmalarının amacı, ilgili semantikleri, sonuç üretecek birimin kullanabileceği hazır formlara dönüştürmektir[1].

Türkçe, yapım ve çekim eklerini içeren sondan eklemeli bir dildir. Eklerin sonda yer alması sözcüklerin eklerine ve köklerine ayrılması işlemi zorlaştırmaktadır. Kelimelere eklenen ekler, kelimenin anlamını ve cümle içindeki görevini belirlemektedir.

Bu çalışmada gerçekleştirilen adımlar aşağıda maddeler halinde sunulmuştur :

1. Kullanıcının girdiği metin içindeki kelimeler kök ve eklerine ayrılır.
2. Eklerden yararlanılarak cümleler, temel ve yan cümlelere ayrılır.

3. Temel ve yan cümleciklerin öğeleri bulunur.
4. Öğelere ayrılmış olan cümleler Prolog'a uygun formata dönüştürülüp önermeler olarak Prolog'a gönderilir.
5. Prolog'a gönderilen önermeler sorgulanarak cümlelerde bulunan öğelere göre sorular üretilir ve uygun formata dönüştürülerek kullanıcıya yansıtılır.
6. Kullanıcının, soruya verdiği cevap için de ilk 4 adım uygulanır.
7. Kullanıcının verdiği cevapları içeren Prolog önermeleri, sorulan soruya uygun şekilde sorgulanarak cevabın doğruluğu kontrol edilir.

Bu adımları gerçekleştirmek üzere çeşitli modüller gerçekleştirilmiştir. Bu modüllerden *kelimeparser*, *cümleparser* ve *öğeparser* internet üzerinde ISAPI teknolojisi ile çalışan modüllerdir. Diğer modüller ise Windows uygulaması (.exe) şeklinde çalışmaktadır.

2. Kelimeparser

Türkçe'deki bir kelime genel olarak (sözcük kökü + yapım ek(ler)i + çekim(ler)i) yapısından oluşmaktadır. Yapım ekleri, sözcük köklerine bitişerek sözcüğün hem anlamını hem de türünü değiştirirler. Çekim ekleri ise sözcüğün anlamını değiştirmemekle birlikte, türünü değiştirebilmektedir. Fakat hiçbir çekim eki isim soylu sözcüğe bitişerek fiil soylu sözcük ya da fiil soylu sözcüğe bitişerek isim soylu sözcük türetmez; yalnızca bir isimden zamir türetmek gibi, isim soylu sözcüklerin alt grupları arasında tür geçişi yapılabilir.

Geliştirilen *kelimeparser* fonksiyonu ile girilen metin kelimelere ayrılır, her kelimenin ilk karakterinden başlayarak ve her seferinde bir karakter ekleyerek bu kelimenin içindeki kökler bulunur. Kökleri bulurken Türkçe kelimelerin köklerinin ve türlerinin bulunduğu bir veritabanını kullanır[2]. Daha sonra kelimenin aldığı eklerle ve cümle içerisindeki durumuna göre olabilecek ihtimallerden birini seçer. İhtimallerin birinin seçilmesi mümkün değilse kullanıcıya hangisinin kullanılması gerektiğini sorar. Sonuçta metin içindeki her kelimeyi (kök + çekim ekleri) şeklinde bir yapıya dönüştürür.

3. Cümleparser

Geliştirilen *kelimeparser* modülünden ek ve köklerine ayrılmış olarak gelen cümleler, öğelerine ayrılma işlemi için

eğer gerekliyse parçalanırlar. Parçalanma işlemi birden fazla yargı içeren cümlelere uygulanır. Cümle temel ve yan cümlelere ayrılarak bu cümleler arasındaki bağlantı türleri belirlenir.

Cümleler yapıları bakımından; basit, birleşik (girişik birleşik- iç içe birleşik-şartlı birleşik), sıralı, bağlı (ki li bağlı- diğer bağlaçlar) cümleler olarak ayrılırlar.

4. Öğeparser

Bir duyguyu, düşüncüyü, isteği, haberi, durumu, olayı vb. ifade etmek için kurulan ve kendi içinde bir anlam ve yargı bütünlüğü olan sözcüğe veya söz dizisine cümle denir. Her bir cümle bir yüklem ve varsa ona bağlı diğer öğelerden oluşur. Cümlede yargıyı bildiren öğe yüklemidir. Öğe cümleyi oluşturan bölümlerin her biridir. Her öğe görev ve anlam yönünden bir tek öğeye eşlik eder ve bu öğe de birinci derece önem taşıyan yüklemse, öğeler çeşitli şekillerde birbirlerinden ayrılabilirler.

Cümlelerin öğeleri;

Yüklem: İş, kılış, oluş, hareket, durum bildiren, haber veren cümleyi bir yargıya bağlayan çekimli öğedir.

Özne: Yüklemde bildirilen işi, oluşu, hareketi, durumu, kılışı yerine getiren hakkında bilgi ve haber veren öğedir.

Nesne: Yüklemde bildirilen ve öznenin yaptığı işten doğrudan etkilenen öğedir.

Dolaylı tümleç: “-e, -de, -den” eklerini alarak cümlelerin, dolayısıyla yüklem anlamını fiili, çıkma, bulunma ve yönelme bakımlarından ilgili olduğu yer yönünden tamamlayan öğedir.

Zarf tümleci: Yüklem anlamını zaman, durum, yön, miktar, tarz, vasıta, şart, sebep, birliktelik yönlerinden tamamlayan kelime ya da kelime gruplarıdır.

Yazılan *öğeparser* modülü ile *kelimeparser* ve *cümleparser*’ dan geçmiş olan, tek eylemli basit cümlelerin öğeleri bulunur.

4.2 Basit Bir Cümleyi Öğelerine Ayıran Modülün Çalışma Adımları

Tablo 1’ de basit bir cümlelerin öğelerine ayrılması için izlenen yöntemlerin bir özeti verilmiştir.

5. Çalışmanın Diğer Adımları

Bu çalışmada gerçekleştirilen *kelimeparser*, *cümleparser* ve *öğeparser* dan başka cümlelerin Prolog önermesine çevrilmesi, Soru Üretme, Yanıt İşleme ve Doğruluk Kontrolü modülleri geliştirilmiştir.

5.1 Prolog Önermesine Çevirme

Öğelerine ayrılmış basit cümleleri, Prolog’a uygun önermeler haline dönüştürür ve bu önermeleri Prolog’a gönderir.

5.2 Soru Üretme

Prolog önermelerini herhangi bir öğeye göre sorgular ve eğer o öğe önermeden bir değer döndürürse, bu sorgulamayı doğal dile dönüştürerek kullanıcıya soru olarak iletir.

5.3 Yanıt İşleme

Kullanıcının soruya verdiği yanıtı, aynı işlemlerle Prolog önermeleri haline dönüştürür.

Tablo 1: Öğeleme Kuralları

Öğenin Özelliği	Öge Çeşidi
(zf)	Özne
(i)	Özne
(s) içeriyor ve sonrası sıfat değilse	Özne
Sonunda (çe) varsa	Özne
Sonunda (tm)*ki(ki) varsa	Özne
Sonunda (ie) varsa	Özne
Sonunda (i) varsa ve sonrası (ie) içermiyorsa	Özne
Sonunda ki(ki) varsa	Sıfat (öge parçasıdır)
(bln)	Belirtili Nesne
(yon) ve sonrasında kadar varsa	Zarf Tümleci
(yon) ve sonrasında göre varsa	Zarf Tümleci
Sonrasında için varsa	Zarf Tümleci
Sonrasında karşı varsa	Zarf Tümleci
(yon) ve sonrası doğru ise	Zarf /Dolaylı Tümleç
(ayr) içeriyorsa ve sonrasında	Zarf Tümleci
dolayı ya da ötürü geliyorsa	Zarf Tümleci
(ayr) ve sonrasında beri geliyorsa	Zarf Tümleci
Sonunda (zf) varsa	Zarf Tümleci
Sonunda (ayr, bul, yon) varsa	Dolaylı Tümleç
(s)	Yüklem
Zaman, şahıs, ekfiil, fiilimsi eki içeriyorsa	Yüklem
Sonrasında ve, veya, ya da varsa	Öge Parçasıdır
Sonrasında ile geliyorsa	Zarf Tümleci
Sonunda (i) varsa ve sonrası (ie) içeriyorsa	Öge Parçasıdır
(tm) içeriyorsa	(ie) içeren ilk kelime dahil öge parçasıdır

5.4 Doğruluk Kontrolü

Soru üretme modülünün, soruyu oluştururken kullandığı sorgulamayı, yanıt önermeleri için uygular ve aynı cevabı alıp-almadığını kontrol ederek yanıtın doğruluğuna karar verir. Eğer yanıt yanlışsa doğru yanıtı kullanıcıya sunar.

6. Deneysel Sonuçlar

Aşağıda örnek bir metin cümlesi verilerek, bu cümlelerin ek, kök ve öğelerine ayrılmasına ve daha sonra öğelerine göre soruların sorulmasına örnek verilmiştir.

Giriş Ekranı:

Ali Amerikaya uçakla gitti. Uçakta yaşlı bir bayanla tanıştı. Uçaktan gelirken ayağını kırdı.

Metnin ek ve köklerine ayrılması ve her bileşik cümlelerin basit cümle yapısına dönüştürülmesi:

Ali(i[i]) amerikaya (i[i]) * y(kh) * a(yon) uçak(i[i]) * la(birl) git (f[f]) * di(dgz). uçakla (i[i]) * de (bul) yaşlı (sf[sf]) bir (sf[sf]) bayan (i[i]) * la (birl) tanış (i[i]) * dı(ef_dgz). uçak(i[i]) * dan (ayr) gel(f[f]) * i(ys) * r(gez) * ken(ken) ayağ (i[i]) * ı (iet3) * n(kh) * ı(bli) kır(i[i]) * dı(ef_dgz).

Girilen metnin öğelerine ayrılması:

Ali(i[i]) → ö / Amerika(i[i]) * y(kh) * a(yon) → dt / uçak(i[i]) * la(birl) git(f[f]) * di(dgz). → yük / uçak(i[i]) * da(bul) → dt / yaşlı(sf[sf]) bir(sf[sf]) bayan(i[i]) * la(birl) tanış(i[i]) * dı(ef_dgz). → yük / uçak(i[i]) * dan(ayr) → dt / gel(f[f]) * i(ys) * r(gez) * ken(ken). → yük / ayağ(i[i]) * ı(iet3) * n(kh) * ı(bli) → bli / kır(i[i]) * dı(ef_dgz). → yük /

Programın girilen metinle ilgili öğelere göre gruplandırılmış soruları:

Nesne sorusu: Uçaktan gelirken neyi kırdı ?

Zarf tümleci sorusu: Ne zaman ayağını kırdı ?

Dolaylı tümleç sorusu: Ali nereye uçakla gitti ?

Nerede yaşlı bir bayanla tanıştı ?

Özne sorusu: Kim Amerika'ya uçakla gitti ?

Kim uçakta yaşlı bir bayanla karşılaştı ?

Kim uçaktan gelirken ayağını kırdı ?

Geliştirilen sistem yukarıdaki sorulara doğru cevap verebilmektedir.

7. Sonuç

Doğal dil işleminin amacı; analiz eden, anlayan ve doğal insan dillerini oluşturan bilgisayar sistemlerini tasarlamak ve meydana getirmektir.

Bu çalışmada, girilen metin öncelikle eklerine ayrılmış, daha sonra yan cümleler içeren karmaşık cümleler, basit cümleciklere ayrılarak aralarındaki bağlantılar tespit edilmiştir. Son olarak basit cümleler öğelerine ayrılarak Prolog önermelerine çevrilmiş ve Prolog veritabanına eklenmiştir. Kullanıcının sorulara verdiği yanıtlarda benzer şekilde öğelerine ayrılarak, doğruluğu kontrol edilmiş ve kullanıcının yanlış yanıt vermesi durumunda, doğru yanıt kullanıcıya sunulmuştur.

8. Kaynakça

- [1] Tatlı, E., Amasya, M.F., "Cümle Çözümleme ve Sorgulama", Bitirme Projesi, *Yıldız Teknik Üniversitesi*, 2001
- [2] Diri, B., "Türkçe Sözlük Veritabanı", *Yıldız Teknik Üniversitesi*, 1999.
- [3] Allen, J., Natural Language Understanding, *The Benjamin/Cummings Publishing Company, Inc.*, 1987