**DOĞAL DİL İŞLEME (NATURAL LANGUAGE PROCESSİNG )**

ÖZET

Doğal dil işlemenin hedefi insanların konuştuğu dili okuyup anlamak, deşifre etmek ve anlamlandırmaktır. Bilgisayar dünyasında iki farklı dil vardır. Bunlardan birincisi makine dilidir, işlemlerini makine dili üzerinden yürütür. İkinci dil ise doğal dildir, doğal dil insanların konuştuğu dil yani Türkçe, İngilizce Korece gibi dilleridir. İnsanların konuştuğu bu dilleri makine algılayıp kendi diline(makine dili) dönüştürdükten sonra bununla işlem yapmasına doğal dil işleme denir. Doğal dil işleme ile çalışabilmek için ilk olarak dilin morfolojisi hakkında bilgi edinme gerekmektedir. İşlemler yapılırken her zaman olduğu gibi NLP uygulanırken de bazı zorluklar karşımıza çıkmaktadır. Kullanım alanlarını incelediğimizde ise görüyoruz ki hayatımızda birçok alanda fark etmeden dahi kullanıyoruz.

ANAHTAR KELİMELER: Doğal Dil İşleme, Makine Dili, Doğal Dil, Dönüştürme,

DOĞAL DİL İŞLEME NEDİR?

İngilizce de NATURAL LANGUAGE PROCESSİNG (NLP) olarak geçmektedir. Doğal dil işleme yapay zekanın bir kategorisidir. Bilgisayar dünyasında iki farklı dil vardır. Bunlardan birincisi makine dilidir, işlemlerini makine dili üzerinden yürütür. İkinci dil ise doğal dildir, doğal dil insanların konuştuğu dil yani Türkçe, İngilizce Korece gibi dilleridir. İnsanların konuştuğu bu dilleri makine algılayıp kendi diline(makine dili) dönüştürdükten sonra bununla işlem yapmasına doğal dil işleme denir.

Doğal dil işlemenin hedefi insanların konuştuğu dili okuyup anlamak, deşifre etmek ve anlamlandırmaktır.

Metin verileri ve ses verileri üzerine işlem yapılıyor. Metin verilerini ile doğrudan işleyip analiz yapılabiliyor fakat ses verileri ile yapılamıyor, ses verileri metinsel veri haline getiriliyor, ardından doğal dil işleme modüllerine uygun hale getiriliyor ve makine tarafından analiz ediliyor.

Doğal dil işleme analizine öncelikle sözlükleri, cümleleri normalize ederek başlamamız gerekir. Normalize etmek yazım anında yapılan yanlışlar, bozukluklar varsa onları düzeltmek demektir. Basit bir örnek: “Ayşeeeee nerdsinn?” olarak yazılmış bir cümleyi normalize ederek “Ayşe neredesin?” olarak düzenlemiş olması gerekmektedir. B u işleme de normalizasyon denmektedir.

Ardından sözcük analizi ile cümle öğelerine ayrılıp yalın hale getirilmelidir. Cümle analizi ile de cümlede ki tüm kelimeler eklerine ve köklerine ayrıştırılarak sözcüklerin yalın hali elde edilir ve kullanıma hazır hale getirilir.

Doğal dil işleme ile ilgilenirken üzerinde çalışılan dilin dil morfoloji seviyelerini tespit etmek de oldukça önemlidir. Bunlar şu şekilde sıralanabilir;

* Kelime bilimi (Morphological-Lexical)
* Sözdizimi (Syntactic )
* Anlamsal (Semantic)
* Söylevdir (Sragmatic-discourse )

1. KELİME BİLİMİ: En küçük parçalar olarak ifade edilebilir. Yapım ekleri, çekim ekleri bunlara örnektir.(-cı, -ci vb.)
2. SÖZ DİZİMİ: Birimlerin birleşerek anlamlı sözcükler oluşturacak biçimde bir araya gelme, tümcelerin birbirine bağlanma, üretilme, dönüştürülme, noktalama işaretlerini kullanım yerleri vb. kurallarını inceleyen dilbilim dalıdır.
3. ANLAMSAL: Kelime gruplarının makine dilinde, kelimeleri birbiriyle karşılaştırabileceği gruplayabileceği eklemeler yaparak yeni kelime türetebileceği, üretirken kullanacağı formüller halinde ele alınabilecek bir gösterim durumudur..
4. SÖYLEV: Bir kelimenin hangi alanda ne anlama geldiğini ifade eder. Cümlede kullanılan bir kelime farklı durumlar için bambaşka anlamlar taşıyabilir.

. Burada bir dilin anlaşılması için gerekli zamanı yine bu dili üretmek için de kullanılmalıdır. Örneğin bilgisayarın ürettiği bir Türkçe kelime olarak düşünülebilir.

**Doğal Dil İşlemede Karşılaşılan Zorluklar**

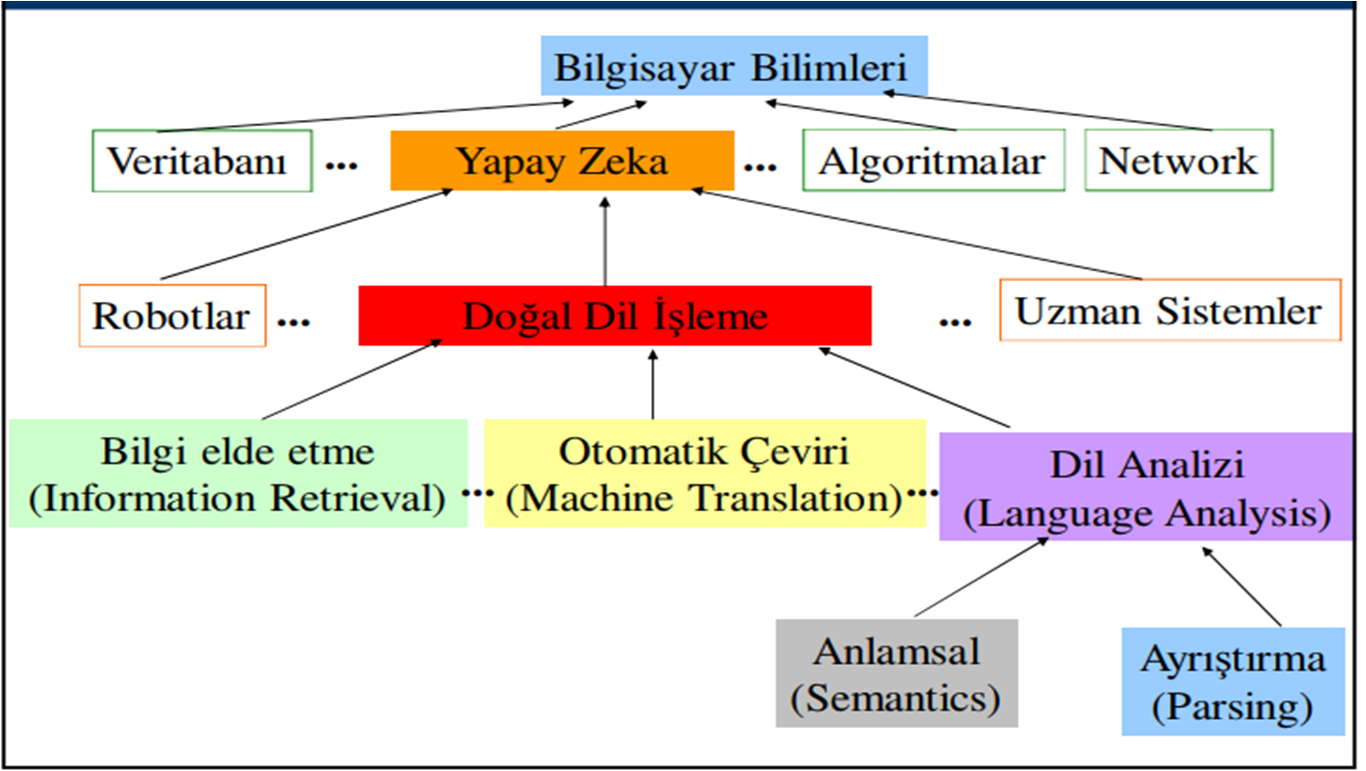
* Belirsizlik-Muğlaklık Durumu (Ambiguity): Kelime ve ya cümlede birden fazla anlam ifade edilme durumudur. Birden fazla anlam ifade ettiğinde hangisininden bahsettiğini anlamada zorluk yaşanması durumudur.
* Sözcük Seviyesi(Lexical): Diller ailesinde bitişik, sondan eklemeli (Agglutinative) olarak bir dil ailesi bulunmaktadır. Yani bir kelimenin sondan eklemeli olarak ,birden fazla ek almasıyla bir çok anlama gelebilir. Bu sebepten dolayı bu kelimelerin anlamını bulmak zorlaşmaktadır. Örneğin Türkçe dili de sondan eklemeli bir dildir.
* Sözdizimsel(Syntactic): Cümlenin içinde kelimelerin yer değiştirilmesiyle oluşturulan cümleler bulunmaktadır. Bu kelimelerin cümlede bulundukları yere göre anlamı değişebilmektedir. Kelimenin vurgusu değişebilir. Bu durumun bilgisayar tarafından anlaşılması gerekmektedir.
* Kısmi Bilgi Yorumlanması(Interpretation of Partial Info): Bir tercüme ,çeviri yapılması durumunda bir cümlede bulunan örneğin bir zamirin hangi kelimeyi karşıladığı bulunması durumudur. İngilizce zamirlerden “he,she,it” durumu gibi.
* Bağlam(Context): Bilgisayara girilen bir cümlenin hangi alandan veya konudan bahsedildiği bilgisayar tarafında anlaşılmalıdır. Örnek olarak ‘Türkiye de ege bölgesinde dağlar denize diktir. ’cümlesi bilgisayara girildiğinde bunun coğrafi bir bilgi olduğu anlaşılmalı ve içerisindeki kelimelerin coğrafi terimler olabileceği varsayabilmelidir. Bir şeyi ifade eden çok şey, çok şeyi ifade eden bir şey vardır cümlesi bu durumu açıklayan çok güzel bir ifadedir. Bir kişi bir cümle ile karşısındaki kişiye bir çok şey ifade edebilir ya da bir kelimeyle her şeyi açıklayabilir. İşte bu durumda bilgisayarın bu cümleyi nasıl anladığı önemli bir sorun olabilir ya da bir kelimeyi ifade eden bir çok kelime olabilir, bilgisayarın en doğru karşılığı bulması da burada yine önemli bir sorundur.

[1]

**DOĞAL DİL İŞLEME KULLANIM ALANLARI**

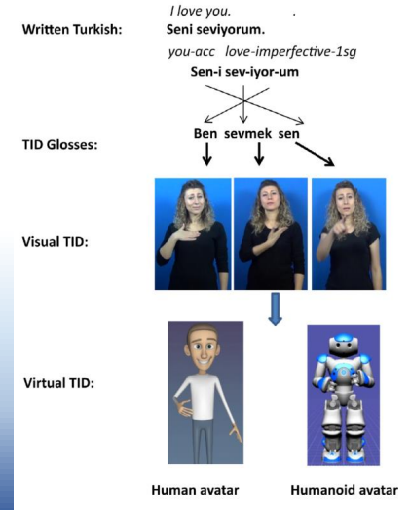
Doğal Dil İşleme günümüz teknolojilerinde birçok alanda tercih edilip ihtiyaç duyulmaktadır. Kullanım alanlarından birkaçını sıralamak mümkündür;

* Soru cevaplama(Question answering): Doğal dil işleme konusunun en önemli ve ilk problemlerinden biridir. Verilen bir metin üzerinden ele alındığında, bu metnin analiz edilerek anlamlı soru ve cevap çıkarılması ya da verilen sorular üzerinden o soruların cevaplarının bir metin üzerinden bulunması istenilir. Bu sorunun çözümü aynı zamanda arama motorlarının da bir problem olan ‘konuşarak (sesli)arama yapabilme sorunu’ çözülmesine de sebep olacaktır.
* Otomatik tercüme(machine translation): Bir makine tarafından bir metnin tercüme edilmesi, bir dilden başka bir dile çevrilmesi işlemidir.
* Bilginin Getirilmesi(Information retrieval): Aranılan bir bilginin metin kaynakları arasında incelenmesi, tespit edilmesi, aranıp bulunması ve getirilmesi işlemidir.
* Kelime işleme (Word Processing (Grammer checking, spell checking ,spell correction,Find/Replace)): Metin üzerinde yapılan yazımsal hataları kontrol etmek için dil bilgisi kontrol işlemi kullanılır ya da bir kelime için heceleme kontrol ve düzeltme işlemleri kullanılır.
* **Makine diline ceviri (Machine translation)** : Sözcük ve kelime birleşimi olan cümlelerden anlamlı bir çeviri elde etme işlemidir.
* **Spam ile mücadele (Fighting spam)** : Mail verilerinden anlamlı ifadeler çıkararak filtreleme işlemi yapılmasıdır.



**TÜRKÇE DOĞAL DİL İŞLEME ÜZERİNE YAPILAN ÇALIŞMA ÖRNEKLERİ**

1. TÜRKÇEDEN TÜRKÇEYE İŞARET DİLİNE ÇEVİRİM



Türkçe olarak ifade edilen konuşmalar bu çalışma sayesinde simüle edilerek işitme engelli insanlar için hizmete sunulmuştur. Her kelime/ cümle tek tek analiz edilerek işaret dilinde ki anlamına çevrilmektedir.

Referans:

Cihat Eryiğit, Hatice Köse, Meltem Kelepir and Gülşen Eryiğit. 2016. Building machine-readable knowledge representations for Turkish sign language generation. Knowledge-Based Systems, 108:179–194

Cihat Eryiğit, Hatice Köse, Meltem Kelepir and Gülşen Eryiğit. 2016. Building

1. ÖZGEÇMİŞLERDEN VE İŞ İLANLARINDAN DOĞAL DİL İŞLEME YÖNTEMLERİ İLE BİLGİ ÇIKARIMI

İş arayan kitlenin özgeçmişini gereken özellikler doğrultusunda doldurup gönderdikten sonra insan kaynakları kısmında bir tarama gerçekleşecek. Üzerinde çalışılan bilgi çıkarımı çalışmaları sayesinde istenen aralıkta ki kişiler verdiği bilgiler üzerinden tespit edilecek, bu yapılan çalışma tamda burada işe girmiş olacak. Yapılan bu çalışmada iş ilanlarından ve özgeçmişlerden doğal dil işleme yöntemleri ile bilgi çıkarımı gerçekleştirilip, elde edinilen bilgilerin eşleştirilmesi yapılarak en uygun adayın ve en uygun iş ilanının adaya/işverene sunulmasına yönelik bir sistemin geliştirilmesi amaçlanmıştır.

Referans:

<https://www.cmpe.boun.edu.tr/>

<https://www.cmpe.boun.edu.tr/>

1. TÜRK KORPUSUNDAN HİPONYM-HİPERNİM İLİŞKİSİNİN ÖRÜNTÜ VE SEMANTİK BENZERLİK TEMELLİ OTOMATİK ÇIKARIMI

Kaynaklardan anlamlı ve bilimsel manaların çıkartılması doğal dil işleme ile ele alınmakta ve oldukça ön plana çıkan konulardan bir tanesidir. Yapılan bu araştırmada Türk korpusundan hiponym-hipernim çiftlerinin kendiliğinden hazılanmış olarak çıkarılması amaçlanmıştır. Hiponym-hipernim çiftlerinin ekstraksiyonu için desen ve semantik benzerlik temelli yöntemler birlikte kullanılır. Desenler ilk hiponym-hipernim çiftlerinden çıkarılır ve desenler kullanılarak çeşitli hipernimler için hiponimler çıkarılır. Yanlış aday hiponymleri belge sıklığı ve anlamsal benzerlik temelli eleme yöntemleri kullanılarak kaldırılır. 14 hipernim için yapılan deneylerden sonra ortalama% 77 doğruluk elde edilmiştir. Bu deneyler sayesinde elde edilen başarı açığa çıkmaktadır.

Referans:

G. Sahin, B. Diri and T. Yıldız, "Pattern and semantic similarity based automatic extraction of hyponym-hypernym relation from Turkish corpus," 2015 23nd Signal Processing and Communications Applications Conference (SIU), Malatya, 2015, pp. 674-677.

G. Sahin, B. Diri and T. Yıldız, "Pattern and semantic similarity based automatic extraction of hyponym-hypernym relation from Turkish corpus," 2015 23nd Signal Processing and Communications Applications Conference (SIU), Malatya, 2015, pp. 674-677.

1. DOĞAL DİL İŞLEME KULLANILARAK BİLGİSAYAR AĞ TERİMLERİNİN WORDNET ONTOLOJİSİNDE UYARLANMASI

Bilgisayar ile doğal dilde iletişimin sağlanması için bilgisayarın doğal dil kurallarını öğrenmesi gerekmektedir. Türkçe bir kelime, ekleriyle birlikte yeni bir kelime oluşturma özelliğine sahiptir. Doğal dil işleme yaygın olarak bilinen yapay zeka ve dilbilimin alt kategorisidir, doğal dillerin kurallı yapısını çözümleyerek anlaşılması veya yeniden anlam katması amacını taşır. Türkçe için yapılan doğal dil işleme çalışmaları, bilgisayar terimleri için otomatikleştirme adına yetersizdir. Bu araştırma bilgisayar terimlerinin ontoloji üzerine doğru yerleştirilmesi amacıyla yapılmıştır. Bu tez çalışmada eklerinden ayrılmış ağ terimlerini içeren bir Wordnet ontolojisi oluşturularak, ontolojide bulunan iki terim arasındaki bağlantı hesaplanmıştır. Terimlerin adlandırılması için en etkili yol ise doğal dil işlemedir.  Verilen bir yazı paragrafında bilgisayar ağ terimlerinin ontolojik bir sözlükte aranmış ve sözlükte bulunmayan kelimeler otomatik olarak ontolojiye eklenmiştir.

Referans:

<http://dergipark.org.tr/en/pub/tbed/issue/30206/300389>

Öncelikle bilgisayara doğal dil işleme kuralları, ardından çalışacağı dil olan Türkçenin kelime oluşturma yapısı öğretilmelidir. Bu yapıya uygun olarak kelime/cümle analizi için yetersiz gelmesinden kaynaklı yukarıda ifade edilen çalışma yapılmıştır.

**DOĞAL DİL İŞLEME ÜZERİNE YAPILAN GÜNCEL ÇALIŞMALAR**

1. DOĞAL DİL İŞLEMEYİ KULLANARAK E-MAİL SINIFLANDIRMASI

Teknoloji ve elektronik iletişimdeki hızlı artış sayesinde, e-posta ciddi bir iletişim aracı haline gelmiştir. İş yazışmaları, hatırlatmalar, akademik bildirimler, web sayfası üyelikleri, e-posta gibi birçok uygulamada birincil iletişim yöntemi olarak kullanılır. Spam e-postaları yok sayarsak, her gün yüzlerce e-posta alınır. Alınan e-postaların önemini belirlemek için, her e-postanın konusu veya içeriği kontrol edilmelidir. Bu çalışmada, alınan e-postaları sınıflandırmak için denetimsiz bir sistem önerdik. Alınan e-postaların koordinatları, Word2Vec algoritması adı verilen doğal dil işleme yöntemiyle belirlenir. Benzerliklere göre, işlenmiş veriler denetimsiz bir eğitim modeliyle k-ortalama algoritması ile gruplandırılmaktadır. Bu çalışmada eğitimde 10517 e-posta kullanılmıştır. Sistemin başarısı 200 e-posta test grubunda test edilir. Test aşaması M3 modelinde (pencere boyutu 3, min. Kelime sıklığı 10, Gram atlama) en yüksek başarıyı (% 91) pekiştirdi. [2]

Günümüzde de çok sık kullanılan e-postaları sınıflandırmak adına yapılan bir çalışmadır. Kişiye göre değişen kategoriler oluşturup kullanım kolaylığı sağlaması amacıyla yapılan bir çalışmadır.

1. DOĞAL DİL İŞLEMEDE DERİN ÖĞRENME UYGULAMALARI EDEBİYATI ARAŞTIRMASI

Derin öğrenme, yapay zeka ve makine öğrenme alanlarının önemli ve güncel bir konusudur. Özellikle son yıllarda farklı derin öğrenme yöntemleri öneren ve bu yöntemleri farklı problemlere uygulayan çalışmaların sayısı artmaktadır. Bu yöntemler aynı zamanda doğal dil işlemenin çeşitli denizaltılarında yaygın olarak kullanılmaktadır ve halen kullanılmaktadır. Bu anket çalışmasında önce derin öğrenme tekniklerinin sınıflandırılması sunulmuştur ve daha sonra doğal dil işleme problemleri için derin öğrenme yaklaşımları hakkında önemli çalışmalar tartışılmıştır. Derin öğrenme ve doğal dil işleme problemlerine derin öğrenme çözümlerine yönelik uygulamalı uygulamalarla yapılan teorik çalışmaların ve yaygınlıkların artması beklenmektedir. [3]

Doğal dil işlemede sıkça karşılaştığımız bir alan olan derin öğrenmeyi de kullanarak yapılan araştırmada NLP esnasında karşılaşılan problemleri ele alarak çözümü üzerine yoğunlaşan bir çalışmadır.

1. TÜRKÇE HEDEF TABANLI DUYGU ANALİZİ İÇİN ALT GÖREVLERİN İNCELENMESİ-HEDEF TERİM, HEDEF KATEGORİ VE DUYGU SINIFI BELİRLEME

Hedef tabanlı duygu analizi (*Aspect based sentiment analysis*) kısaca, bir metnin içinde yer alan farklı duyguların ilgili oldukları hedef varlıklar ile birlikte tespit edilmesi olarak tanımlanabilir. Güncel tanımlamalar, hedef tabanlı duygu analizini, üç temel alanla (hedef terim, hedef kategori ve duygu sınıfı) temsil edilen duygu tanımlama gruplarını belirlemeyi amaçlayan aşamalı bir görev olarak betimlemektedir. Yürütülen çalışmalar, ABSA 2016 yarışmasındaki görevler (1- Hedef kategori belirleme, 2- Hedef terim belirleme, 3- Hedef kategori ve hedef terimin aynı anda belirlenmesi ve 4- Duygu sınıfı belirleme) takip edilerek tasarlanmış ve yine burada sağlanan Türkçe restoran yorumları veri kümesi üzerinde değerlendirilmişlerdir. Hedef kategori, hedef terim ve ikisinin aynı anda belirlenmesi görevleri için, kelime vektörleri (word vectors) ve doğal dil işleme çıktıları (sözcük ve cümle analizi bilgileri) kullanan koşullu rastgele alanlara (CRF – conditional random fields) dayalı bir dizilim etiketleme algoritması tasarlanmış ve her üç görevi de tek aşamada çözebildiği gösterilmiştir.  Elde edilen sonuçlar ile bu ilk üç görev için literatürdeki en yüksek başarımların elde edildiği görülmüştür: Hedef kategori belirlemede %66,7 F1-skoru, hedef terim belirleme %53,2 F1-skoru, hedef kategori ve hedef terimin aynı anda belirlenmesinde %46,7 F1-skoru. Bunun yanı sıra, duygu sınıfı belirleme için cümle analizi sonucunda hedef terime yakın kelimelerden özellik seçimine dayalı bir lineer sınıflandırma yöntemi sunulmuş ve literatürde sınırlı sistemler tarafından raporlanan en başarılı sonuç (%76,1 F1-skoru) elde edilmiştir.[4]

REFERANSLAR:

[1]: YBS ansiklopedi, cilt 2, sayı 4, basım yılı 2015

[2]: S. Sel and D. Hanbay, "E-Mail Classification Using Natural Language Processing," 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019, pp. 1-4.

[3]: Dergi Park, [Yıl 2018, Cilt 2, Sayı 2,](https://dergipark.org.tr/en/pub/uybisbbd/issue/41787)Sayfa 76 – 86

[4]: [Year 2018, Volume 11 , Issue 1,](https://dergipark.org.tr/en/pub/gazibtd/issue/34945)Pages 43 – 56, Dergi Park <https://dergipark.org.tr/en/pub/gazibtd/issue/34945/325865>