

DOGAL DİL İŞLEME NEDİR ?

Doğal Dil İşleme, NLP (Natural Language Processing) olarak bilinen Yapay Zeka ve Dil Biliminin bir alt kategorisidir. Türkçe, İngilizce, Almanca, Fransızca gibi doğal dillerin (insana özgü tüm diller) işlenmesi ve kullanılması amacı ile araştırma yapan bilim dalıdır. Doğal dil işleme bilgisayarların günlük konuşma dilimizi en iyi şekilde işleyebileceği ve analiz edebilmesi için nasıl programlanacağını gösteren yöntemleri içeren bir çalışma alanıdır. Yani kısaca bilgisayar dili ve doğal dil arasında köprü oluşturmak şeklinde tanımlanabilir. Doğal dil işleme farkında olmasak bile günlük hayatımızı birçok alanına dahil olmuştur Örneğin internette ve sosyal medyada birçok yere Entegre edilmiş olan otomatik makine çevirisi e-posta kutusunu dolduran gereksiz veya zararlı e-postaları filtreleyip gereksiz e-posta(junk) kutusuna Taşıyan metin sınıflandırma arama motorlarından “Bunu mu demek istediğiniz” gibi sorularla günlük hayatımızda kullandığımız doğal dil işleme uygulamalarından sadece birkaçıdır. Günlük konuşma dilinin canlı olması zaman içinde değişime uğraması veya kullanılan platform içeriye göre değişiyor olması ve her dilin yapısının birbirinden farklı olması sebebiyle doğal dil işleme alanına özgü pek çok farklı çalışma alanı ortaya çıkmıştır. Metin özetleme, soru cevaplama, konuşma, tanıma, ses tanıma, döküman sınıflandırma sesi yazıya çevirme Duygu analizi bunlardan sadece bir kısmıdır

Doğal dil işleme girdilere göre ikiye ayrılır. Yazılı metinlere(text processing) göre ve ses (speech processing) üzerine yapılanlardır. Doğal dilin genellikle iki kaynağı bulunmaktadır. Birinin konuşması durumunda ona sesli yanıt verilmesi veya yazılı bir metnin analiz edilmesi olarak düşünülebilir. Doğal dil işleme çalışmalarını şu şekilde olmaktadır; ses üzerinden çok fazla çalışma yapılamamakla birlikte daha çok sesli anlatımların yazılı hale getirilip ve daha sonra yazılı hale getirilen metinler işleme dahil edilmeleriyle oluşmaktadır. Doğal dil işlemenin çalışma seviyelerine bakıldığında dört ana madde görülmektedir. Kelime bilimi(Morphological-Lexical), söz dizimsel (syntactic), anlamsal(semantic) ve

söylevdir(pragmatic-discourse). Doğal dil işleme aşamaları bu dört ana madde üzerinden anlatılmaya çalışılacaktır.

Kelime bilimi(lexical), kelimelerin anlamlarının anlaşılmasıyla ilgilenir. Bir kelimenin kökünün hangi ekler alarak hangi anlama dönüştüğünü inceler ve bu sayede kelimenin ne olduğu anlaşılır.

Söz dizimsel (syntactic), cümledeki kelimelerin dizilim şeklidir. Bir cümlede kelimelerin nasıl dizildiğiyle ilgilenir.

Anlamsal(semantic), kullanılan cümlelerin anlamlarını inceler. Çünkü doğal dilin bilgisayar tarafından doğru işlenebilmesi için ilk, bu cümlelerin doğru bir şekilde anlaşılmasıdır.

Söylev(pragmatics-discourse),bir konuşma sırasında kullanılan kelimeler ve anlamlarıyla ilgilenir. İlk önce bilgisayar tarafından konuşma anlaşılmalı ve nelerden bahsedildiği anlaşılan cümlelere doğru kelimeler kullanılarak cevap verilmesi durumudur.

Doğal dil işlemenin bir diğer kısmı üretim(Generation) kısmıdır. Burada bir dilin anlaşılması için gerekli zamanı yine bu dili üretmek için de kullanılmalıdır. Örneğin bilgisayarın ürettiği bir türkçe kelime olarak düşünülebilir.



Doğal dil anlama bir cümlelerin anlamını belirtmek için Metin veya konuşmanın sözdizimsel ve anlamsal analizini kullanan doğal dil işlemenin bir alt kümesidir. Söz dizimi bir cümlelerin dilbilgisel yapısını ifade ederken anlam bilim söylemek isteneni ifade eder.

Doğal dil oluşturma doğal dil işlemenin başka bir alt kümesidir doğal dil anlama bilgisayarda okuduğunu anlamaya odaklanırken doğal dil oluşturmak bilgisayarların yazabilmesini sağlar DDO insanın dilinden Metin yanıtı üretme süreci olarak da

tanımlanabilir bu Metin metinde konuşma uygulamalarına da dönüştürülebilir.

Doğal Dil İşlemede Karşılaşılan Zorluklar

1-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.

2-Sözdizimsel(Syntactic) Cümlelerin içinde kelimelerin yer değiştirilmesiyle oluşturulan cümleler bulunmaktadır. Bu kelimelerin cümlede bulundukları yere göre anlamı değişebilmektedir. Kelimenin vurgusu değişebilir. Bu durumun bilgisayar tarafından anlaşılması gerekmektedir.

3-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.

4-Bağlam(Context) Bilgisayara girilen bir cümlelerin 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. Birşeyi 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

Uygulama Alanları

Doğal dil işleme(DDİ),günümüz teknolojilerinde bir çok alanda kullanılmaktadır bunlardan bazılarını şu şekilde sıralamak mümkündür

1. Soru cevaplama
- 2.Otomatik tercüme
3. Bilginin Getirilmesi
4. Kelime işleme
5. Doğal Dil İşleme-Veri Tabanı Arayüzü(Veri tabanı üzerinden sorgulama işlemi yapıldığında kullanılan bir işlemdir. Bir üretim departmanında soru olarak gelen ‘Bu yıl ki satış oranı nedir?’ sorusunu veri tabanında uygulamak için bir sorgu cümlesine çeviren bir sistemdir)

Doğal Dil İşlemede Kullanılan Modeller

Geçtiğimiz son 10 yılda tüm platformlarda popülarlığını hızla arttıran python programlama dili DDİ alanında en fazla kullanılan dil olmasının yanı sıra veri bilimi büyük veri analitiği deri öğrenme ve sayacağımız birçok alanda en fazla kullanılan dil olmayı başarmıştır python programlama dilini öğrenmenin kolay olmasının yanı sıra çok sayıda ve çeşitli kütüphanelerin olması da kullanımını arttıran faktörler arasında yer almaktadır.

DDİ alanında en fazla kullanılan araçlardan biri olan Natural Language Toolkit(NLTK) python ile doğal dil işleme kitabında(Bird vd.,2009) birkaç yıla yayılan bir proje sonucudur bilgisayarlı Dil Bilim öğretimi için de kullanılan NLTK python'da yazılmış bir dizi yazılım molekülü ve bir dizi derlem ve diğer kaynakları içinde barındırır. NLTK ve diğer yazılım paketlerinin yüklenmesi ile ilgili tüm bilgiler Ayrıca kitabın internet sitesinde mevcuttur

Son zamanlarda python kullanılan DDİ kütüphaneleri artmaya devam etmektedir;

coreNLP (Manning vd.,2014)

polyglot (Al-Rfou)

textBlob (Loria,2020)

Bu alanda en fazla kullanılan arasında yer almaktadır derin öğrenme kütüphanelerinin de çoğunun python ile yazılması ya da python dil desteği sunması DDİ ve derin öğrenme alanında python kullanımını arttırmıştır Bu nedenle python programlama diline En azından temel seviyede öğrenmek bu alanda çalışmak için başlangıçta yeterli olacaktır.