**T.C.**

**KIRIKKALE ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

****

**BİLGİSAYAR MÜHENDİSLİĞİ**

**BİLGİSAYAR MÜHENDİSLİĞİNDE ÖZEL KONULAR**

**DOĞAL DİL İŞLEME**

**GÖKHAN POLAT**

**160255008**

**NİSAN 2020**

**ÖZET**

İletişim, duygu ve düşüncelerin çeşitli yollarla kişiden kişiye aktarılmasıdır. Sosyal bir varlık olan insan da varolduğu günden bugüne iletişimi kullanmıştır. İletişim ilk ortaya çıktığı dönemlerde konuşma ve işaret ile sağlanırken, ilerleyen zamanlarda yazının bulunması ile bu aktarımların döküman haline getirilmesi sağlanmıştır. Matbaa ve sanayinin gelişmesiyle dokümanlar kitap haline gelmiş ve seri üretim olarak basılmaya başlanmıştır. Google Books projesiyle 150 farklı kaynaktan yapılan araştırmaya göre günümüze kadar basılan birbirinden farklı kitapların sayısının 130 milyon civarı olduğu saptanmıştır. ([www.webtekno.com](http://www.webtekno.com/))

Yazılı metinlerin içlerinde bulunan bilgi ve birikim insanlık açısından gelecek adına önemli olmuştur. Çünkü geçmiş bilgiler gelecek adına hep yol göstericidir. İnsanların öğrenmeye olan açlığı ve dokümanların hızlı artışı büyük bir sorun ortaya çıkarmıştır. Bu kadar bilgiye nasıl hakim olunabilecek? Bu sorunu çözmek için çeşitli araştırmalar ve yöntemler geliştirilmeye çalışılmıştır. Bunların en kayda değer ve dikkat çekeni ise Alan Turing'in "Makineler düşünebilir mi?" sorusunu sorup, ilk yapay zeka temellerini atmasıdır. Bu sayede makinelere bilgiler tanıtılıp, işlenmeye başlanmış ve ileriye dönük tahminlemeler yapılmaya başlanmıştır.

Yapay zeka çeşitli dallara ayrılmakla beraber, doğal dil işlemede bir alt kategorisidir. Bilgisayar dünyasında iki farklı dil vardır. Bunlardan birisi makine dilleri olan programlama dilleri, diğeri ise doğal diller olan günlük konuşma dilleridir. Doğal dil işlemeyi de insan konuşma dillerinin makine diline işlenmesi olarak da kısaca tanımlayabilirirz.

**GİRİŞ**

**1. Dil Oluşum Evreleri**

**1.1. Dil Nedir?**

Dil, insanların birbirleriyle iletişimde olabilmeleri ve kendilerini ifade edebilmekte kullandıkları bir araçtır. Biz gülerek, heyacanımızı, mutluluğumuzu; bağırarak, kızgınlığımızı, korkumuzu; yumruklarımızı sıkarak, tehditlerimizi, şiddetimizi; gözlerimizi büyük açarak, şaşkınlığımızı, kınamalarımızı ifade edebiliriz (Eifring & Theil, n.d.). Bu kullanılan iletişim yöntemlerinden hiçbiri dil kadar önemli ve etkili değildir.

Chomsky, dilin çocukluk yıllarında duyulandan, doğal bir dile dönüşümünün, insanın genetik yapısıyla ilişkili olduğunu ifade etmektedir (Chomsky, 1986). Dil bilimcileri de genel olarak insanların dili nasıl edindiği, ürettiği ve anladığı ile ilgili araştırmalar yapmaktadırlar.

**1.2. İletişim**

En az iki kişi arasında gerçekleşen fikir, bilgi, düşünce, duygu, his, davranış vs. alışverişidir ("İletişim Nedir" n.d). İnsanlar en eski tarihten bu yana doğadaki tüm canlı ve cansız varlıklarla iletişim kurma eğiliminde olmuştur. Bu iletişimde bir kaç unsura ihtiyaç vardır. Bunlar: Gönderici (kaynak), ileti (mesaj), kanal, alıcı (hedef), dönüttür (geri bildirim) (Samet Kaya, 2018).

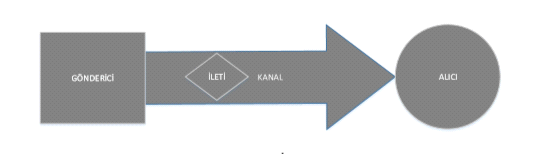
**Gönderici (Kaynak):** İletişimin oluşmasındaki en başlıca unsurdur. Göndericinin vermek istediği iletinin anlamlı bir hale dönüştürüldüğü kısımdır.

**İleti (Mesaj):** Göndericiden alıcıya gönderilmek istenen tüm bilgi, duygu, düşünce vb. verilerin bütünüdür.

**Kanal:** Gönderici ve alıcı tarafından iletilen verinin iletişim bağlantısıdır.

**Alıcı (Hedef):** Göndericinin gönderdiği mesajın iletildiği yerdir.

**Dönüt (Geri Bildirim):** Alıcının göndericiden aldığı mesaja karşı verdiği cevaptır.



**Şekil 1.** İletişim Aşamaları

**2. Doğal Dil İşleme**

Doğal dil işleme, esas görevi doğal dili çözümleme, anlama, yorumlama ve üretme olan bilgisayar sistemlerinin tasarımını ve gerçekleştirilmesini konu alan bir mühendislik dalıdır (Prof.Dr. Banı Diri, NLP). Doğal dil işleme, insan dilinin otomatik analizi ve gösterimi için teorik olarak motive edilmiş hesaplama teknikleridir (Cambria & White, 2014).

Bilgisayara bir dili algılatma ve bilgisayarın insanlarla iletişime geçebilmesi, yirminci yüzyıl ilk yarısı için bilim kurgu filmlerinden bir kesit olarak gelmişti. Alan Turing, makine ve zekayı işleme (Computing Machinery and Intelligence) isimli klasik makalesinde "Makineler düşünebilir mi?" diye sormuş ve Turing deneyiyle makinenin insan gibi konuşup konuşmayacağını sorgulamıştır. Yapay zekanın hızlı gelişimi ile web siteleri otomatik dil çevirebilmekte, cep telefonları insanların ne demek istediklerini anlayabilmekte, arama motorları otomatik düzeltmeler yapabilmektedir. Ancak doğal dili tamamen anlayabilen makinelerden uzağız. Otomatik çevirmeler hala bir insan danışmanının gözleminden geçmek durumunda kalmakta ve Turing testini hala tam geçebilmiş bir makine bulunmamaktadır (Kibble, 2013).

İnternetin gelişiminden önce veriler yapılandırılmış ve veri tabanı sistemlerinde tutulan bu veriler, SQL sorgulama dili ile istenilen verilerin sistemden getirilmesini sağlamaktaydı.

İnternetin hızlı gelişimiyle birlikte birçok veri işlenmemiş şekilde elektronik ortamda durmaktadır. Bu durum beraberinde verileri işleyebilmek, bilgiden bilgi çıkarabilmek ve anlam çıkarabilmek vb. çeşitli problemleri getirmektedir. Bu problemi çözebilmek için çeşitli sistemlere ihtiyaç duyulmaktadır. Doğal dil işlemenin metin analitik, başlık tespiti, bilgi çıkarımı gibi alanları bu sorunlara çözüm aramaktadır (Kibble, 2013).

Bazı modern uygulamalar:

\* Arama motorları (Google, Yahoo, Bing, Baidu)

\* Soru-Cevap (IBM's Whatson)

\* Doğal Dil Asistanı (Apple's Siri, Google Assistant)

\* Translation System (Google Translate)

\* Haber Özetleri (Yahoo)

\* Otomatik Deprem Raporlayıcı (LA Times)

Basit bir derin öğrenme çerçevesinin, adlandırılmış varlık tanıma, anlamsal rol etiketleme ve sözcük etiketleme gibi birçok doğal dil işleme görevinde, en modern yaklaşımlardan daha iyi performans gösterdiği ortaya konulmuştur (Collobert, ve diğerleri, 2011). Doğal dil işleme alanında istatistiksel yöntemler, kural tabanlı yöntemlere göre daha başarılı olmuştur. Bu alanda sözcük etiketlemesi ile beraber bilgisayarın cümlede anlatılmak istenileni anlamasında kolaylık sağlanmıştır. Ancak her kelime tek bir anlam içermediği için, anlamda ifade etmek istenilene ulaşılamamaktadır. Buna örnek vermek gerekirse, Türkçe'de sayı olarak kullanılan "yüz" kelimesinin, aynı zamanda insan suratı anlamını taşıması verilebilir. Bu sorunu ortadan kaldırmak için etiketleme yapılırken kelimenin önündeki ve arkasındaki kelimelere de bakılarak doğru etiketleme yapılmaya çalışılmalıdır. Etiketleme işlemlerinde, iki yönde ki bilgileri kullanarak olasılık üreten, bunu tekrarlayan sinir ağı türü olan, iki yönlü uzun-kısa vadeli bellek (BLSTM) ağlarının, sıralı verileri etiketlemek için doğal dil işleme çalışmalarında çok etkili olduğu gösterilmektedir (Wang, Qian, K. Soong, He, & Zhao, 2015). BLSTM mimarisi, dili modelleme (Sundermeyer, Schlute & Ney, 2012) (Sundermeyer, Ney & Schluter, 2015), dili anlama (Yao, Zweig, Hwang, Shi & Yu, 2013), makine çevirisi (Sundermeyer, Alkhouli, Wuebker & Ney, 2014) ve sözcük etiketlemesi gibi doğal dil işleme alanında ki uygulamalar için üstün performans elde edilmesine yardımcı olmaktadır (Wang, Qian, K. Soong, He & Zhao, 2015).

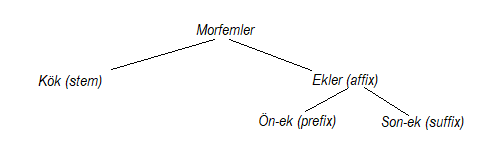
**2.1. Doğal Dil İşleme Adımları**

Genel olarak doğal dil işleme sözdizimi (Syntax), anabilim (Semantic) ve faydabilim (Pragmatic) ana başlıkları altında analiz edilir. Cümleye ilk olarak sözdizimi analizi uygulanarak başlanır. Bu anabilim açısından ya da sözlüksel açıdan bir sıra ve yapı sunar. Bunu kelimenin metin içerisinde söyleniş veya konumunun cümleye kattığı anlamı irdeleyen faydabilim analizi takip eder. Son kısım genellikle kelimenin cümle içerisindeki ilişkisel anlamını inceleyen söylem çözümlemesi (Discourse Analysis) ile ilgilidir. Kelimenin cümle içinde ne anlama geldiği, cümlenin metin içerisinde ne anlama geldiği gibi kısımlar bazı karışıklıklara neden olur. Fakat doğal dil işlemeyi kısımlara ayırmak, yazımsal açıdan anlaşılırlığını ve çözümlenmesini kolaylaştırır (Indurkhya & Damerau, 2010).

*Sözdizimi (Syntax):* Ardı ardına gelen kelimelerin cümle oluşturmak için oluşturduğu yapının çözümlenmesidir. Bu aşamada analizi biten sözcükler, dilin daha büyük elamanları olan metinleri oluşturmak için birleştirilirler.

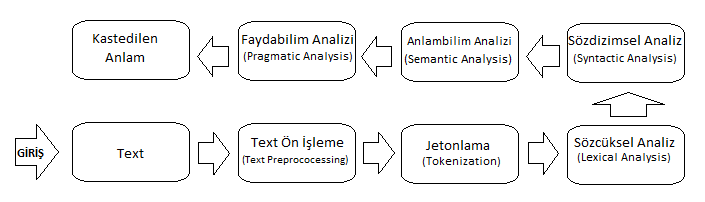
*Anabilim (Semantic):* Dilin gerçek dünyada kurulan iletişim sayesinde anlaşılması ve bunun sonucunda eyleme geçirilmesi bu aşamada gerçekleşir. Dilde birbiriyle ilişki kelimelerin ardarda gelmesiyle oluşturulan cümlelerin, bilgisayar mantığına çok uymasada matematiksel yaklaşımlarla çözülmeye çalışılır.

*Faydabilim (Pragmatic):* Kelimeler tek başlarına, dilin kuralları çerçevesinde ele alınır. İncelemeler sırasında kelimenin ekleri, kökleri ve bunlara ilişkin kurallar incelenerek çözümlenme işlemi yapılmış olur. Aynı zamanda "morferm" denilen kelimelerin en küçük anlamlı birimlere ayrıştırılmasına, **Şekil 2.** örnek gösterilmiştir.



**Şekil 2.** Morfemlerin Yapısı

Bu üç adım başlangıç için yeterli fakat veri üzerinde doğal dil işleme yapacaksak farklı adımlara da ihtiyaç vardır. Aşağıdaki şekilde doğal dil işleme adımları verilen bir metinden, anlam çözümlemeye kadar verilmiştir. Bu adımlar ihtiyaç doğrultusunda artırılabilir.



**Şekil 3.** Doğal Dil İşleme Aşamaları

*Metin Ön İşleme:* Ses ya da yazı olarak bulunan veriyi işlenecek hale getirilmesini sağlayan aşamadır. Bu adımda metinde akışı bozan hataların giderilmesi gibi pürüzleri halledip, bundan sonraki adımlara uygun hale gelmesi sağlanmaktadır.

Her metin dil kurallarına uygun olmaz hatta içerisinde hedef dile uygun olmayan kodlama farklılığından (UTF8, ISO-8859-1) dolayı eksik ya da yanlış karakter bulunabilir. Bunların temizlenmesi ya da düzeltilmesi gerekir. Tüm diller boşluklarla sınırlandırılmış kelimeler sunmaz.

Örneğin İngilizce kelimeler boşluklarla ayrılabilirken, Çince, Japoncagibi diller boşluklarla ayrılmaz ve kelime bölütleme işleminde farklı bir yol izlenmesine ihtiyaç duyarlar. Farklı diller farklı işlemlere tabi tutulabilir (Samet Kaya, 2018).

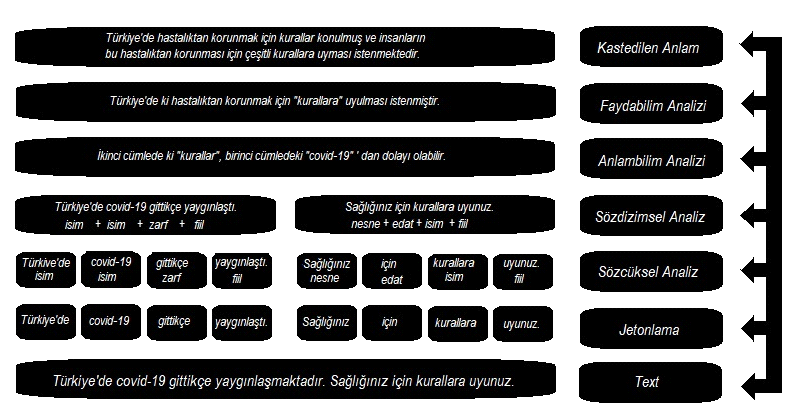
*Jetonlama:* Bu işlemde, işlenmeye uygun hale getirilmiş veriyi bir sonraki aşamaya hazır hale getirebilmek için küçük parçalara ayrılması işlemidir. Bu parçalar kelimeler, noktalamalar, harf kümeleri vb. olabilir.

*Sözcüksel Analiz:* Hazır hale gelmiş jetonların dilde ne anlama geldiği ya da yapacağımız işlem için ne anlama geleceği işlemi bu aşamada gerçekleştirilir. Ayriyeten oluşturulmuş jetonlar tam istenildiği gibi olmayabilir, bu aşamada veriyi ihtiyaç doğrultusunda eklere ayırıp daha fazla bilgi çıkarımı yapılabilir. Bu işleme örnek olarak, kelimelerin o dilde isim, sıfat veya fiil olarak sınıflandırılması verilebilir.

*Sözdizimsel Analiz:*  Her dilin anlamlı bir sözdizimi ve kuralları olmalıdır. Veri kümeleri her zaman dil kurallarına uygun gelmese de cümlenin anlamını çözmek sözdizimsel kurallar kullanılır.

*Anlambilimsel Analiz:* Bu adımdan önce analiz edilmiş kelimelerin ve jetonların anlaşılmaya çalışıldığı bölümdür. Kelimelerin bulunduğu cümlede ne anlama geldiği bu bölümde incelenir fakat, unutulmamalıdır ki bir kelime birden fazla anlama gelmektedir. Kelimelerin olası anlamları bu bölümde olsa dahi ileri adımlarda tam olarak neyin anlatılmaya çalışıldığı bulunmaya çalışılır.

*Faydabilim Analizi:* Anlambilimsel analizi yapılmış kelimlerin bu bölümde gerçekte neyi anlatmak istediğini bu adımda bulmaya çalışılır. Tüm metin içerisinde ne anlama geldiği, neye atıf yaptığı gibi ilişkiler faydabilim analizinde yer alır. Bu işlem oldukça karmaşıktır ve hala tam anlamıyla çözülebilmiş değildir (Samet Kaya, 2018).



**Şekil 4.** Doğal Dil İşleme Örneği

**Şekil 4.**' de bu anlatıklarımıza örnek verilmiştir. Örnek cümlemiz üzerinde, doğal dil işlemenin nasıl yapıldığı basit bir şekilde anlatılmaya çalışılmıştır. İlk adımda cümle verilip doğal dil işleme uygulanması istenilmiştir. İkinci adımda cümlemiz jetonlarına ayrılmış olmasına rağmen duruma veya isteğe göre farklı bölme işlemleri de yaptırılabilir. Sözcüksel analiz kısmında tamamen kullanılan dilin gramersel olarak öğelerine ayrılması sağlanmıştır. Sözdizimsel kısımda gramerleri ile birlikte dizilişi gösterilmiştir. Anlambilimde ise eş atıf çözümlemesi yapılarak bir önceki cümledeki atıfta bulunduğu kelime saptanmıştır. Faydabilimde ise kurulan cümlenin faydalı bir şekilde kullanılması anlamı gelmektedir. Son adımımız olan kastedilen anlamda ise cümlenin tam anlamıyla çözümlenip, makinenin tıpkı bir insan gibi düşünüp cümleyi yorumlaması ifade edilmiştir.

**KAYNAKLAR**

*Alan Turing. (1995) .Computing machinery and intelligence.Computers & Thought, 11-35.*

*Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. IEEE Computational Intelligence Magazine, 9(2), 48-57.*

*Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research. IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, 9(2), 48-57.*

*Chomsky, N. (1986). Knowledge of Language: Its Nature, Origin, and Use. New York.*

*Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research, 12, 2493-2537.*

*Eifring, H., & Theil, R. (n.d.). The 2005 manuscript version of Halvor Eifring & Rolf Theil: Linguistics for Students of Asian and African. Retrieved from.*

*Indurkhya, N., & Damerau, F. J. (2010). Handbook of Natural Language Processing. Processing (Vol. 2).* [*https://doi.org/10.1038/nbt1267*](https://doi.org/10.1038/nbt1267)*.*

*Kibble, R. (2013). Introduction to natural language processing. University of London.*

*Prof.Dr. Banu Diri. (2014). Doğal Dil İşleme(DDİ)-Natural Language Processing (NLP).*

*Samet Kaya,(2018). Doğal Dil işleme Teknikleriyle Yazar-Kitap Tanıma.*

*Sundermeyer, M., Schlüter, R., & Ney, H. (2012). LSTM Neural Networks for Language Modeling. Interspeech. USA.*

*Sundermeyer, M., Alkhouli, T., Wuebker, J., & Ney, H. (2014). Translation Modeling with Bidirectional Recurrent Neural Networks. EMNLP Conference, (s. 1425). Doha, Qatar.*

*Sundermeyer, M., Ney, H., & Schluter, R. (2015). From Feedforward to Recurrent LSTM Neural Networks for Language Modeling. IEEE/ACM Transactions on Audio, Speech and Language Processing, 23(3), 517-529.*

*Yao, K., Zweig, G., Hwang, M.-Y., Shi, Y., & Yu, D. (2013). Recurrent Neural Networks for Language Understanding. Interspeech, (s. 2524–2528).*

*Wang, P., Qian, Y., K. Soong, F., He, L., & Zhao, H. (2015, 10 21). Part-of-Speech Tagging with Bidirectional Long Short-Term Memory Recurrent Neural Network. arXiv:1510.06168v1 [cs.CL]. adresinden alındı.*

[*https://www.webtekno.com/tum-dunya-tarihinde-gunumuze-kadar-toplam-kac-farkli-kitap-basildigini-biliyor-musunuz-h46605.html*](https://www.webtekno.com/tum-dunya-tarihinde-gunumuze-kadar-toplam-kac-farkli-kitap-basildigini-biliyor-musunuz-h46605.html)

*kaynağı:*[*http://mentalfloss.com/article/85305/how-many-books-have-ever-been-published*](http://mentalfloss.com/article/85305/how-many-books-have-ever-been-published).