



TÜRKÇE DERS ANLATAN YAPAY ZEKÂYA GİDEN YOLDA DOĞAL DİL İŞLEME

Sinan KUL¹

¹ Bilgisayar Kullanımı, Açık Öğretim Fakültesi, Atatürk Üniversitesi, Erzurum Türkiye

Özet

Alan yazın incelendiğinde, öğrencilerin dil, fen ve matematik gibi alanlardaki yeterliliklerinin artırılması amacıyla, sınıflarda öğretmeni asiste etmek üzere robotların kullanıldığı az sayıda çalışmanın bulunduğu görülmektedir. Öğretmenin yerine geçebilecek bir yapay zekâ projesinin gerçekleştirilebilmesi için de makinelerin, insanlar tarafından konuşulan dili anlaması gerekmektedir. Böylece kaliteli eğitmen sorununa çözüm alternatifi olabilecek ve kişiselleştirilmiş eğitim modellerinin geliştirilebilmesine katkı sunulabilecektir.

Bu çalışmada Türkçe özelinde, doğal dil işleme çalışmaları bağlamında gelinen mevcut durum analiz edildikten sonra bu projenin nasıl alt başlıklara ayrılacağına dair öneriler sunulmuş ve tasarımında izlenecek yol haritası belirlenmeye çalışılmıştır. Türkçe doğal dil işleme çalışmalarının hangi düzeylerde yapıldığı anlatılırken, konular, yapay zekâ projesi içindeki konumu dikkate alınarak filtrelenmiştir. Bu bağlamda diller arası çeviri çalışmaları, bu çalışmanın kapsamı içine dâhil edilmemiştir.

Anahtar Kelimeler: Yapay Zekâ, Doğal Dil İşleme, Robot Öğretmen

NATURAL LANGUAGE PROCESSING ON THE WAY TO TURKISH LECTURER ARTIFICIAL INTELLIGENCE

Abstract

When literature is examined, it is seen that there are few studies in which robots are used to assist teachers in classrooms in order to increase the competence of students in fields such as language, science and mathematics. In order to realize an artificial intelligence project that can replace the teacher, the machines need to understand the language spoken by people. Thus, it can be an alternative solution to the problem of the lack of quality instructors and contribute to the development of personalized training models. Thus, an alternative solution to the problem of the lack of quality trainers can be developed, and contributed to the development of personalized training models.

In this study, after analyzing the current situation in the context of natural language processing studies in Turkish, suggestions on how this project can be subdivided has been submitted and it has been tried to determine the road map to be followed in the design of the project. As Turkish natural language processing at which levels is explaining, the subjects has been filtered by considering its position in the artificial intelligence project. In this context, interlingual translation studies are not included in the scope of this study.

Keywords: Artificial Intellegent, Natural Language Programming, Robot Lecturer

GİRİŞ

Yüzyüze eğitimde öğrenciden, öğretmenden, öğrenme ortamından ve materyalinden kaynaklı çok sayıda sorun dolayısıyla kaliteli bir eğitim verilememektedir. Bu sorunların başında öğrencilerin maddi imkânsızlıkları paralelinde ailelerinin ve/veya kendilerinin eğitime ilgisizlikleri gelmektedir. Eğitim ortamının fiziki yetersizliği, araç gereçlerin eksikliği, sınıf ortamlarının uygun olmayışı, kalabalık sınıflarda eğitim verilmesi ve kaliteli öğretmen azlığı gibi sorunlar, kaliteli eğitimin standart

olarak verilmesinin önündeki engeller arasında sayılabilir (İşman, 2008). Bu sorunları aşmak için sıklıkla teknolojiye faydalanılmaktadır.

Yapılan bazı araştırmalara göre eğitim sektöründe 2035 yılı itibariyle işlerin %9'unda yapay zekânın istihdam edileceği ön görülmektedir (Benitti, 2012). Robotlar, hali hazırda öğrencilerin özellikle dil, fen ve matematik gibi alanlardaki yeterliliklerinin artırmak amacıyla, öğretmeni asiste etmek üzere kullanılmaktadır (Mubin vd., 2013). Örneğin fen bilimleri dersi için geliştirilen insansı bir robot olan “Saya”, kaldırma prensipleri hakkında materyal sunmak ve robotiklere giriş yapmak için Japonya'daki sınıflarda uygulanmıştır (Hashimoto, 2011). Kanda ve arkadaşları (2004) “Robovie” adını verdikleri robotun, çocuklarla İngilizce konuştuğu bir Japon ilkokulunda yaptıkları bir saha araştırmasında, robotla sıklıkla etkileşime giren çocukların İngilizce becerilerinde gelişmeler olduğunu bulmuşlardır. Kullanılan robotun hafızasında sadece 300 hazır cümle bulunduğu ve konuşulan 50 kelimeyi anlayabildiği ifade edilmiştir.

Öğrencilere İngilizce öğretmek için tasarlanan diğer bir robot “EngKey” ise Filipinler'de ve Güney Kore'deki okullarda kullanılmıştır (Yun vd., 2011). Movellan ve arkadaşları (2009) ise erken eğitim merkezinde 2 hafta boyunca işletilen bir robotun, küçük çocukların hedeflenmiş sözcükler konusundaki bilgilerini geliştirdiğini bulmuştur. Çocukların kelime becerilerini artırmak için düşük maliyetli olarak geliştirilen “Rubi” adlı robotun kullanıldığı bu çalışmada 18-24 aylık bebeklerde %27'lik bir gelişimin kaydedildiği ifade edilmektedir.

Ancak Gaudiello (2016) tarafından yapılan bir araştırmaya göre özellikle küçük yaşta çocukların yüz yüze eğitimlerinin robotlar tarafından verilmesinin ebeveynler tarafından sağlıklı görülmediği sonucuna ulaşılmıştır. Aynı çalışmada insanların bir “iCub” robotuna ve robotun karar almadaki etkisine olan güvenini araştırmış ve teknik sorularla ilgili cevaplarının sosyal sorularla ilgili olanlarına kıyasla daha az etkili olduklarını bulmuştur. Gaudiello ayrıca, sınıflara robot yerleştirilmeden önce onlara uygun bir güven ve kabul düzeyinin belirlenmesinin önemine vurgu yapmıştır.

Öğretmenin tümüyle yerine geçebilme potansiyelindeki ilk çalışma, Hanson Robotics tarafından geliştirilip 2010 yılında piyasaya sürülen “Bina48” robotudur. 2017 yılında Kaliforniya'daki Notre Dame de Namur Üniversitesi'nde bir felsefe kursuna dâhil olup bu kursu başarıyla tamamlayan “Bina48”, 2018 yılına gelindiğinde ABD'de üniversite öğrencilerine etkileşimli olarak, etik felsefesi dersini verebilmiştir.

Sınıf ortamındaki eğitim sisteminin sorunlarına bir çözüm de uzaktan eğitim uygulamalarıdır. Öğrenci ile öğretmenin fiziksel olarak ayrı ortamlarda bulunduğu ve bilgiyi aktarmada yaygın olarak tv ve internet teknolojilerinin kullanıldığı, yaşam boyu öğrenme ve zaman/mekan bağımsız öğrenme pratikleri barındıran uzaktan eğitim, örgün öğretime katılmayan öğrencilerin de eğitim alabilmelerine imkân tanımaktadır. Öğrenciler, kendi öğrenme hızlarına göre eğitimi planlayarak zamanlarını etkin kullanabilmektedir. Uzaktan eğitim, eğitim giderlerini de azaltmasına rağmen yüzyüze eğitimin yerine tamamiyle geçememektedir. Zira uzaktan eğitimde öğrenciyle göz teması kurulamadığı için öğrenme gücünü yaşayan öğrenciye anında destek olunamamaktadır. Uzaktan eğitim, öğrencilerin öğrenme süreçlerini planlama sorumluluğu yüklerken, uzaktan eğitim platformunun etkin kullanılma becerisini de zorunlu kılmaktadır (Dinçer, 2016). Bu sebepler dolayısıyla yüz yüze eğitim ile uzaktan eğitimin birlikte uygulandığı Harmanlanmış Öğrenme uygulamaları da tercih edilmeye başlanmıştır (Çolakoğlu, ve Akdemir, 2010).

Uzaktan eğitimde yukarıda sayılan sorunlar dışında ayrıca kişiselleştirilmiş eğitim de uygulanmamaktadır (acaba). Oysa ki öğrenme ortamının ve materyalinin kişiselleştirilmesinin, gereksiz bilgi yüklemesini azaltarak öğrencilerin öğrenme gücünü giderdiği, merakını ve motivasyonunu artırdığı ve böylece akademik başarısını artırdığı, yapılan çalışmalarla ortaya konmuştur (Rivera, 2005; Liang, Lai ve Ku, 2007; Sezer, 2015). Kişiselleştirilmenin yapılabilmesi için de her bir bireyin özelliklerini, ilgilerini, istek ve ihtiyaçlarını tespit edip ona göre materyalleri ve öğrenme sürecini uyarlayabilen teknolojilere ihtiyaç duyulmaktadır. Bu teknolojilerin başında da yapay zekâ gelmektedir.

Gerek kişiye özgü eğitim materyalinin oluşturulması ve kişiselleştirilmiş eğitim sürecinin uygulanmasında ve gerekse öğreticiden kaynaklı sorunların giderilebilmesi için ders anlatan yapay

zekâ projesinin hayata geçirilmesi gerektiği değerlendirilmektedir. Bu çalışmada Türkiye’de öğretmenin yerine geçebilecek yapay zekâ uygulamasının geliştirilme sürecinde gelinen mevcut durum hakkında (doğal dil işleme çalışmaları bağlamında) bilgi verildikten sonra bu büyük projenin nasıl alt başlıklara ayrılacağına ve tasarımında izlenecek yol hakkında bazı öneriler paylaşılmaktadır.

DOĞAL DİL İŞLEME

Makinelerin, insanların dilini anlaması ve iletişim kurabilmesi için doğal dil işleme çalışmaları yapılmaktadır. Bu bağlamda insanlar tarafından konuşulan bir dilin (doğal dil) anlaşılması ve analiz edilmesi için bilgisayar marifetiyle çeşitli çalışmalar yürütülmesine, Doğal Dil İşleme (Natural Language Processing) denilmektedir. Bu çalışmalar, dilden bağımsız ve dil bağımlı çalışmalar olmak üzere iki ana eksende yürütülmektedir. Dil bağımsız çalışmalar, makine öğrenmesi ve istatistiksel yöntemler ile yürütülürken, kural tabanlı olarak yürütülen dil bağımlı çalışmalar, dört ana odak etrafında şekillenmektedir: kelime düzeyinde morfolojik (biçim bilimsel) çalışmalar, söz dizimsel düzeyde cümle analizi çalışmaları, anlam düzeyindeki çalışmalar ve söylev (pragmatik) düzeyindeki çalışmalar (Şeker, 2015).

Literatür incelendiğinde çoğu doğal dil işleme çalışmasının, İngilizce için yapıldığı görülmektedir. İngilizcenin yaygın kullanımı ve İngilizce metinlerin işlenmesinin kolaylığı bunda etkilidir. İngilizce çalışmalarda 30 bin kelime bir sözlük kullanılması yeterli gelirken aynı tür çalışmalar, Türkçe için yapıldığında, kelimenin tüm olası ekleriyle sözlükte bulunması gerekmektedir ve bu da neredeyse olanaksızdır (Vural vd., 2004). Çünkü bir kelime, kök ve gövdesine eklenen yapımlarla yeni anlamlar kazanmaktadır ve bir kelimeye 3 yapımla eki eklendiğinde bir milyondan fazla kelime oluşturulabilmektedir. Hatta öyle ki eklerin kelimeye eklenme sırası dahi önemlidir. “pazarlıkçı” ve “pazarlıklık” kelimelerine eklenen “lık” ve “cı” ekleri bu duruma örnek verilebilir (Dönmez, 2016).

Kelime Düzeyinde

Kelime düzeyinde doğal dil işleme çalışmaları yapılırken öncelikle metinler, tokenlerine (kelimelerine) ayrılır. Bu ayırma işlemi Türkçe için kolaydır zira Türkçede kelimeler arasında boşluk karakteri ve noktalama işaretleri bulunur (Japonca gibi bazı dillerde kelimeler arasında boşluk bulunmaz).

Kelimeler, metin içinden tespit edildikten sonra iki tür ayrıştırma daha yapılabilmektedir. Bunlardan biri kelimenin hecelenmesi, ikincisi ise kelimenin ek ve köklerine ayrıştırılması (morfolojik ayrıştırma) işlemidir. Ayrıştırmanın faydalarından biri, kelimenin yanlış yazılıp yazılmadığına karar vermektir. Çünkü hecelere ve/veya ek ve köklerine ayrılamayan bir kelime Türkçe dil kurallarına uygun olmadığı için yanlış yazıldığına hükmedilebilir. Ayrıca ekler ve kökler ile ilgili ses olayları hece düzeyinde gerçekleştiği için morfolojik ayrıştırma öncesinde kelimenin hecelere ayrılabilmesi önemlidir.

Kelimenin yeniden türetilmesi de ayrıştırma işleminin doğru yapıldığının sağlaması niteliğindedir. Öztürk (2016) de bu bağlamda morfolojik türetme çalışması yürütmüş ve %82 doğrulukla kelimeleri denetimsiz olarak türetilmesini bildirmiştir.

Çabuk ve arkadaşları (2003), kelime kökünü bulurken, kelime köklerinin bulunduğu bir sözlükten faydalanmıştır. İlgili kelimeyi ilk harfinden başlayarak ve her seferinde bir harf daha ilave ederek sözlükte aratan, olası kökleri işaretleyen ve kelimenin aldığı eklerle ve cümle içindeki konumuna göre en olası köke karar veren bir sistem geliştirmiştir.

Cebiroğlu (2002) ise sonlu durum makinelerini ters işleterek ayrıştırmayı sağdan sola doğru yaptırmıştır. Kelime sonuna gelen ekleri tek tek çıkardıktan sonra köke ulaştığı bu yöntem ile alıntı kelimeler için de köke inebilmiştir. Solak ve Oflazer (1993) tarafından gerçekleştirilen yazım denetimi çalışmasında ise 23 bin kelime kökünden oluşan sözlükten faydalanılmıştır ve kelime kökü bulurken ekler çıkarıldıktan sonra kelime kökünün ekler dolayısıyla değişime uğraması durumu (sesli düşmesi, sessiz benzeşmesi vs.) da dikkate alınmıştır.

Türkçe kurallı bir dil olmasına rağmen kelimeler, POS (Part of Speech – Metin Parçası) etiketlerine (isim, sıfat, fiil vs.) ve cümledeki kullanım amacına göre farklı şekillerde ek ve köklerine ayrılabilir. Çünkü biçim bilimsel çözümleyicilerde (morfolojik ayrıştırıcı) isim ve fiil olma durumuna göre farklı sonlu durum makinelerinin çalıştırılması gerekmektedir. Kelimenin cümledeki konumu ise kelimenin türünü belirginleştirmektedir. Dolayısıyla kelimenin eklerine ayrılması işi, cümledeki konumundan bağımsız olarak yapılamamaktadır. Örneğin “Ayşe buraya ne zaman gelir” cümlesindeki “gelir” kelimesi, fiil olarak kullanılmış ve bir adet çekim eki almışken; “Gelir kalemlerine birlikte göz atalım” cümlesindeki “gelir” kelimesi isim olarak kullanılmıştır ve çekim eki almamıştır. Bu belirsizlikleri gidermek için Sak, Güngör ve Saraçlar, (2011), biçim bilimsel belirsizlik giderici uygulaması geliştirmiştir.

Biçim bilimsel çözümleyici ayrıca çekimsel, türetimsel (yapım eki almış kelimeler için) ve birleşik (birleşik iki kelime için) kelimeler için üç farklı şekilde uygulanmaktadır. Türkçe için sınırlı sayıda biçim bilimsel çözümleyici (Dönmez, 2016; Sak, Güngör ve Saraçlar, 2011; Oflazer, 2005) geliştirilmiştir: Zembekte 20 bin kök kelime kullanılırken; Oflazer (1994)’in geliştirdiği yazılımda (PC-KIMMO) 23 bin kök kelime kullanılmaktadır.

Cümle Düzeyinde

Kelimelerin cümle içindeki yeri, cümlenin anlamını değiştirebilmektedir. Bu bakımdan isim/sıfat tamlamalarında ve cümle öğelerinin dizilişinde, kelimelerin konumu önemlidir. Kelimelerin cümle içindeki konumunu ve anlamını belirginleştirir, çekim ekleridir. Bu bakımdan kelime düzeyinde yapılan analizlerin (morfolojik analiz) çıktısı, cümle düzeyindeki analizi beslemektedir. Yani bu iki alanda yapılan çalışmalar birbiri içine geçiş durumdadır.

Türkçede cümleler, yapıları itibarıyla, basit, sıralı, bağlı ve birleşik olmak üzere dört türe ayrılmaktadır. Basit cümlede tek yargı (fiil) bulunurken, diğer üç tür cümlede birden fazla yargı bulunmaktadır. Birleşik cümlede esas yargıyı bildiren bir adet temel cümle ve bir veya birden fazla yan cümle bulunmaktadır. Bağlı cümle yapısında ise anlam yakınlığı nedeniyle iki cümle birbirine bağlanır. Dolayısıyla basit olmayan cümle yapılarında ana cümlenin ve her bir yan cümlenin ayrı ayrı incelenmesi gerekmektedir.

Cümle içinde geçen kelimeler genellikle (%95) sağa bağımlı olarak birbiriyle ilişkilidir (Eryiğit vd., 2006). Sıfatın isimden önce gelmesi, zarfın fiilden önce gelmesi durumu gibi. Ancak çoğunlukla bağlı kelime ikilileri ardışık olarak gelmeyip araya başka kelimeler alabilmektedir. Ayrıca bağlı olduğu kelimenin bütünüyle değil de daha küçük bir parçasıyla ilişkili olabilmektedir. “Küçük odadayım” ifadesi ele alındığında “küçük” kelimesi “odadayım” kelimesini nitelememektedir. Çünkü “odadayım” ifadesi fiildir. “küçük” kelimesi, sıfattır ve dolayısıyla “odadayım” kelimesinin bir parçası olan “oda” varlığını nitelemektedir.

Türkçe cümlelerin bağıllık analizi ile ilgili yapılan ilk akademik çalışma, Oflazer (2003) tarafından geliştirilen kural tabanlı gerekirci ayrıştırıcısıdır. Eryiğit ve arkadaşları (2006) ise ilk olasılık tabanlı bağıllık ayrıştırıcıyı geliştirmiştir. Bu yöntemde önce birimler arası bağıllık olasılıkları bulunur sonra da en olası bağıllık ağacı bulunmaya çalışılır. Nivre ve Nilsson (2005) ise “Maltparser” adını verdiği ayrıştırıcıda sistemin eğitilmesi için Destek Vektör Makinesinden faydalanmıştır. Türkçe metinlerin cümle düzeyinde dil bilgisi çözümlemesi alanında yapılan diğer çalışmalara, Doğan ve Karaağaç (2012) ve Aygül, Karaalioğlu ve Amasyalı (2014) tarafından yapılan çalışmalar örnek verilebilir.

Anlam Düzeyinde

Anlam düzeyindeki çalışmalar, metnin anlamıyla ilgilenir ve dilin gerçek dünyayla irtibatını sağlar (Delibaş, 2008). Metinlerin ayrıştırılması ve dil bilgisi denetimine tabi tutulması üzerine yapılan çalışmalar, 50 yılı aşkın süredir devam ederken; anlam düzeyindeki çalışmalar son 20 yıldır yapılmaktadır. Bu bağlamda özellikle Türkçe için yapılan çalışmalar, WordNet, PropBank, FrameNet ve VerbNet gibi anlamsal sözlüklerin Türkçeye uyarlanamaması veya eksik uyarlanması gibi sebepler dolayısıyla ne yazık ki yetersiz kalmaktadır. Anlam düzeyinde yapılan çalışmalar için temel teşkil eden anlamsal sözlükler, metin sınıflandırma, özetleme, bilgi çıkarımı ve soru cevaplama gibi çalışmalarda kullanıldığı için de bu sözlüklerin geliştirilmesi önem arz etmektedir (Şahin, 2016).

Türkçe çalışmalarda kelimeler (kavramlar) arasında alt-üst, parça-bütün, eş anlam ve zıt anlam ilişkilerini tespit için WordNet'ten faydalanılmaktadır. Bu alanda yapılan ilk çalışma olan ve "WordNet" 'in Türkçeye çevrilmesi amaçlanan "BalkaNet" çalışmasında 117 bin eş küme bulunmaktadır. Yazıcı ve Amasyalı (2011) tarafından yürütülen çalışmada kullanılan küme sayısının (11.628 küme) neredeyse 10 katıdır. Çalışmaların istenen seviyede başarıyı elde edememesinin sebebi, İngilizce kelimelerin Türkçeye tam mutabık düşmemesidir.

Türkçe için anlamsal sözlüklerin oluşturulmasında kullanılan diğer bir yöntemse, derlemlerden faydalanılarak kelimeler arası ilişkilerin otomatik olarak çıkarılmasıdır. 12 milyon cümleden oluşan "BOUN" derlemi ve 2 milyon cümleden oluşan "Kemik" grubu derlemi bu tür çalışmalara örnek verilebilir.

Metinler veya derlemler analiz edilerek ilişkili kelimelerin belirlenmesinde ise bazı kelime veya kelime grupları tarafından bağlanan kelimelerin otomatik olarak ilişkilendirildiği şablon yöntemi kullanılmaktadır (Şahin ve Amasyalı, 2014). Bu yöntemde örneğin, "elma vb. meyveler" ifadesi analiz edildiğinde elmanın meyve kelimesine alt üst ilişkiyle bağlı olduğu görülmektedir. "köpek ya da kedi" ifadesinde de "ya da" şablonu (kelime grubu) kardeş iki kelimeyi birbirine bağlamaktadır. Şablon yöntemi kullanılarak anlamsal sözlük oluşturma çalışmaları incelendiğinde O. Güngör ve T. Güngör (2007)'ün kelimeleri anlamsal olarak kümeleme çalışması yaptığı görülmektedir. Bunun dışında kelimelerin alt üst anlamsal ilişkilerinin çalışıldığı (Şahin, Diri ve Yıldız, 2015; Yıldırım ve Yıldız, 2012), parça bütün ilişkilerinin araştırıldığı (Ittoo ve Bouma, 2010; Yıldız, Yıldırım ve Diri, 2013) ve anlamsal ilişkiler üzerine yapılan (Fırat, Kılıçaslan ve Uçar, 2013; Amasyalı, 2012) çalışmalar bulunmaktadır.

Anlam düzeyinde yapılan diğer bir çalışma varlık ismi tanımadır. Bu çalışmalar "Bilgi Çıkarımı" bölümünde anlatılacaktır.

Söylev Düzeyinde

Yorum ve kullanım olarak da literatürde karşılık bulan söylev düzeyinde yapılan çalışmalarda, anlatımda kullanılan kelimelerin bağlamıyla ilgilenilir. Bir kelimenin sözlükte onlarca farklı anlamı ve kullanım örnekleri bulunabilmektedir. Söylev düzeyindeki çalışmalarda bu kelimenin hangi anlamda ve amaç ile söylendiği tespit edilmeye çalışılır. Dolayısıyla, anlamsal sözlüklerin kullanılmasının ötesinde, eş sesli kelimelerin hangi anlamda kullanıldığının tespiti ve terim anlamının anlaşılması gerekmektedir. Bu bağlamda, kelimenin, cümledeki ve kelime grubu içindeki konumu dikkate alınarak, gündelik kullanımdaki anlamının ve mecaz anlamının da tespit edilmesi gerekmektedir.

Örneğin herhangi bir cümlede "balık" kelimesi geçtiğinde bunun yemek mi yoksa akvaryum hayvanı mı olduğunu belirlemek, söylev düzeyindeki çalışmaların ilgi alanındadır. "Balık etli olmak" ve "balık eti yemek" ifadelerindeki "balık" kelimeleri söylev düzeyindeki ayrışmaya bir örnek verilebilir.

TÜRKÇE DERS ANLATAN YAPAY ZEKÂ PROJESİNİN ALT BAŞLIKLARI

Konuşma Tanıma

Konuşan kişi ile iletişim kurulabilmesi için öncelikle konuşma sesi analiz edilerek olasılığı en yüksek metne dönüştürülmektedir. Bu aşamada seslerin, cümle bağlamının da dikkate alınarak kelimelere dönüştürülmesi başarıyı artıracaktır. Sesin yazıya dönüştürülmesi aşamasından sonra artık metin işleme çalışmaları yapılabilmektedir.

El Yazısı Tanıma

El yazısı tanıma çalışmaları, literatürde Optik Karakter Tanıma (Optical Character Recognition - OCR) olarak geçmektedir. Bu çalışmalar, matbu veya el ile yazılmış yazıların, görüntü işleme teknikleri kullanılarak, resim içinden çözümlenmesini ve yazıya dönüştürülmesini amaçlar. "Google Drive", "OmniPage Ultimate", "Abby FineReader", "Adobe Acrobat Pro DC", "LightPDF", "OCR.Space", "i2OCR", "NewOCR.com", "To-Text OCR Converter" ve "OCRConvert.com" (Viilup, 2019; Nevres, 2019) OCR uygulamalarına örnek verilebilir. Bu uygulamalar, matbu yazılar

için dahi kusursuz çözümler sunamazken, el yazısı söz konusu olduğunda başarı yüzdeleri iyice düşmektedir.

Microsoft, bu sorunu çözebilmek için “Seeing AI” uygulamasını güncelleyerek el yazısı tanıma özelliği eklemiştir (Perkinselearning, 2019). “Pen to Print”, uygulaması da bu amaç için özel özel geliştirilmiş uygulamalardandır. Türkçe el yazısını resim dosyası içinden tespit için de Mavi (2019) tarafından makine öğrenmesi uygulaması (karar ağaçları ile) geliştirilmiştir.

Yalnızca görüntü işleme tekniklerinin kullanıldığı bu çalışmalar, çevrim dışı çalışma olarak adlandırılmaktadır. Yazı oluşturma aşamasında hareket takibinin (dokunmatik ekranlar üzerinde) de yapılabildiği çevrim içi yazı tespiti çalışmalarında ise el yazılarının tespitinin daha kolay olduğu söylenebilir. Bu kapsamda geliştirilen uygulamalardan “Google Handwriting Input”, android işletim sisteminde; “Nebo” ise Windows işletim isteminde çalışarak ekrana yazılan el yazılarını tespit edebilmektedir. “Ink” uygulaması ise sadece grafik tablet üzerinden kullanılabilir (Viilup, 2019).

Vural ve arkadaşları (2004) tarafından geliştirilen Türkçe çevrim içi el yazısı tespiti sisteminde ise gizli markov modeli uygulanmış ve %90 başarı elde edilmiştir. Çalışmada, karakter tespitindeki hataların bir kısmının “ç”, “ş”, “ğ” ve “ü” gibi Türkçeye özgü harflerin yazım süresinin uzunluğundan kaynaklandığı belirtilmiştir.

Yazım Yanlışlarının Düzeltmesi

Eski dönemlerde musahhihler tarafından yapılan yazım yanlışlarının düzeltilmesi işini, bugün bilgisayarlar devralmış bulunmaktadır. Yazım ve dil bilgisi denetiminde, eklemeli olmayan bir dilde yazılan herhangi bir kelimenin doğruluğunun tespiti için kelimelerin doğru yazılışlarının tutulduğu bir tablodan faydalanılabilirken, eklemeli (bağlantılı) diller sınıfında olan Türkçe söz konusu olduğunda bu yöntem yetersiz kalmaktadır. Türk Dil Kurumu’nun sözlüğünde 100,353 adet kelime bulunduğu (Dilsiz, 2005) düşünüldüğünde yapım ekleri ve çekim ekleriyle bu tablonun boyutunun ne kadar büyüyebileceği, çoğu Türkçe doğal dil işleme çalışmasında dile getirilmiştir.

Burada yeri gelmişken söylemekte fayda var, yazım yanlışlarının tespiti ve düzeltilmesine odaklanan çalışmalar kelime düzeyinde yapılan çalışmalarla birlikte yürütülmektedir. Dolayısıyla daha önceden de bahsedildiği gibi kelimenin hecelerine, ek ve köklerine ayrılabilmesi, kelimenin Türkçe olduğu ve doğru yazıldığı hakkında bilgi verebilmektedir. Bunun için ayrıştırma işlemi sonucunda bulunan kök ifadesinin, kelime kök hallerinin tutulduğu sözlükte arattırılması gerekmektedir. Diğer bir deyişle, bu yöntemde, sözcük denetimi için, kelimelerin ekli hallerinin sözlükte tutulmasına gerek duyulmamaktadır.

Sözcük denetimi ile ilgili yapılan çalışmalardan biri, Oflazer ve Solak (1992)’a aittir. Çalışmalarında 23 bin kök kelimededen oluşan bir sözlük kullanan bir yazılım geliştirmişlerdir. Yazılım tarafından denetlenen kelime, sözlükte bulunamazsa son harfi çıkarılıp tekrar sözlükte aranmakta; son harf kalana veya kelime sözlükte bulunana kadar bu işlem devam ettirilmektedir. Son harfe gelindiğinde ise kelimenin hatalı yazıldığı sonucuna varılmaktadır. Ancak bu yöntem her zaman doğru köke ulaşmayı sağlamamaktadır. Örneğin, “çarpıştı” kelimesinin son harfleri atıldığında “çarpı” kelimesi (doğru kök “çarp”) kök olarak kabul edilmektedir. Benzer bir biçimde ek gelerek, ses düşmesine veya sessiz yumuşamasına uğrayan kökler için de kelimenin yanlış yazıldığı sonucuna ulaşılabilmektedir.

Delibaş (2008), bu sorunları aşmak için yaptığı çalışmada kelimenin ilk harfinden başlanarak kelime sözlüğündeki olası tüm kelimeleri taramaktadır. Çalışmasında, kelime kökünün ses olaylarına uğramış versiyonlarının da tutulduğu ağaç yapısını kullanmıştır. Güzey ve Oflazer (1994) tarafından yapılan başka bir çalışmada ise köklerin ses düşmesine veya sessiz yumuşamasına uğrama durumları da tespit edilebilmiştir.

Dilsiz (2005) ise olası tüm kelimeler için bir sözlük oluşturulmasının zorluğu dolayısıyla kelimeleri sadece kök halleriyle sözlükte saklarken, biçim bilimsel çözümleyici marifetiyle eğitim veri setindeki kelimeleri ek ve köklerine ayırdıktan sonra çalışmasında, köklerin bulunma sıklıklarını

(istatistikleri) ve köke gelen eklerin sırasını da kaydetmektedir. Ek sırasını kaydederken ardışık ek ikililerinin bulunma sıklıklarını da kaydetmektedir.

Türkçe kural tabanlı bir dil olduğu için yazım yanlışlarının tespitinde sözlüksüz yazım denetimi çalışmaları da kullanılabilir. Çınar (2003), kelime içindeki hece uyumu, kelimenin ve hecenin sonundaki sessizler ve sessiz uyumları üzerinden yazım denetimi yapabilmektedir. Ancak sesli uyumu üzerinden yaptığı kontrollerin, Türkçede kullanılan ve sesli uyumuna uymayan kelimelerin çokluğu dolayısıyla yetersiz kaldığını ifade etmiştir.

Cebiroğlu (2002) ise sözlük kullanmadan, kelimeyi, ek ve köklerine ayırabildiği (tasarladığı sonlu durum makinelerini kullanarak ekleri sondan başlayarak kırptığında kelime köküne ulaşabildiği) çalışmasında, kelimenin isim ve fiil olma durumuna bağlı olarak ek içinde kullanılan harfler üzerinden de yazım denetiminin yapılabileceğini göstermiştir. Sonlu durum makineleri geliştirilerek yazım ve dil bilgisi denetimi yapan çalışmalardan biri, Eryiğit, Oflazer ve Adalı (2006)'ya aittir. Yazım yanlışlarının düzeltilmesi üzerine yapılan bir diğer çalışma da Delibaş(2008)'a aittir.

Kelimenin, yanlış yazıldığı tespit edildiği, en olası kelimeyle değiştirilmesi gerekmektedir. En uygun kelimenin seçimi içinse yaygın kullanılan yöntemlerden biri, uzaklık (distance) fonksiyonunun uygulanmasıdır. Kul (2018), sözlük kullandığı çalışmasında klavye farklılıklarını (Q ve F) ve harflerin klavyedeki uzaklıklarını (tombul parmak hatası) da dikkate alarak sözlükte aratmaktadır. Sözlükte uygun bir kelimeyle uyuşmadığında ise Türkçe karakterler için ASCII eşlenikleri de değiştirilerek sözlükte tekrar aratmaktadır.

Eşleşmeyen harf sayısına ve harflerin klavyedeki uzaklıklarına ek olarak harflerin/seslerin/kelimelerin söyleniş benzerlikleri de dikkate alınabilmektedir. Daha isabetli tercih yapılabilmesi içinse kelimenin, ilgili kelime grubundaki, cümledeki ve hatta tüm metindeki bağlamıyla birlikte istatistiksel yöntemlerle değerlendirilmektedir.

Sözlüksüz ve kural bazlı olmayan çalışmalardan biri de karakter düzeyde di-gram, tri-gram ve four-gram tablolarından faydalanılmasıdır. Bu tablolarda ilgili harf grubunun kaçınıcı seviyede bulunduğu bilgisi kelimenin hatalı yazılıp yazılmadığı konusunda bilgi verebilmektedir. Ancak bu yöntemin uygulanabilmesi için kullanımda olan tüm kelimelerin ekli hallerinin de tablolara kaydedilmesi gerekmektedir (Çınar, 2003; Aşlıyan ve Günel, 2007).

Metin Özetleme

Metin özetleme çalışmaları, doğal dil işlemenin en zorlayıcı çalışmalarındandır. Kaynak metnin yerine geçebilecek önemli içerikleri barındıran daha kısa bir metne çevrilmesi olarak tanımlanabilen metin özetlemede, önceden tanımlı bir değer fonksiyonu ve anlamsal sözlükler kullanılarak, her bir cümle için bir sayısal değer hesaplanır ve bu sayısal değerler temel alınarak cümlelerin anlamsal benzerlikleri anlaşılmasına çalışılır (Tülek, 2007). Metin özetleme yapılırken ayrıca benzer cümlelerin sıklıkları da hesaba katılmaktadır (Hu ve Liu, 2004; Bawakid ve Oussalah, 2008).

Çoğunlukla istatistiksel yöntemlerin kullanılarak gereksiz detayların gizlendiği metin özetleme çalışmalarıyla dijital çağda bilgi yığınlarına boğulan okuyucuya zaman kazandırılmaya çalışılmaktadır. Web sitelerindeki bilgilerin özetlendiği çalışmalarda birden fazla belge birlikte özetlenebilmektedir. Otomatik sınav sorusu hazırlayan projede özet metnin kullanılması süreci kolaylaştıracağı ve verimi artıracacağı değerlendirilmektedir.

Cümle Öğelerine Ayırma

Türkçede cümle öğelerinin yaygın dizilişi, özne, nesne ve yüklem şeklindedir. Ancak Türkçede farklı anlamlara gelebilen farklı öğe dizilişleri de kullanılmaktadır. Yani Türkçe, serbest söz dizilimine sahiptir, denilebilir.

Cümle öğelerinden biri olan nesne, belirtili ve belirtisiz nesne olmak üzere ikiye ayrılmaktadır. Bunun dışında cümlenin, dolaylı tümleç, zarf tümleci ve ilgeç tümleci olmak üzere üç öğesi daha bulunmaktadır. Cümle öğelerinin belirlenmesi çalışmasında, kelime kök ve gövdelerine gelen çekim ekleri analiz edilerek, ilgili kelimenin, cümlenin hangi öğesini teşkil ettiği çözümlenmeye çalışılmaktadır. Ancak bu yöntem, tek başına yeterli değildir. Çünkü aynı tür ekleri almasına rağmen

farklı anlamda ve bağlamdaki kelimeler, cümlenin farklı öğelerini teşkil edebilmektedir. Örneğin, “hastalıktan öldü” cümlesinde “hastalıktan” kelimesi sebep ifade ettiği için cümlede zarf tümleci; “denizden çıktı” cümlesinde ise “denizden” kelimesi dolaylı tümleç görevi görmektedir (Çabuk vd., 2003).

Öğelerin tespitinde genellikle kural tabanlı çalışmalar yürütülmektedir. Kuralların elle belirlenmesi veya eğitim veri seti içinden öğrenilmesi (makine öğrenmesi) ise bu çalışmaları iki alt sınıfa ayırmaktadır. Makine öğrenmesi kullanan yöntemlerden biri, Özköse ve Amasyalı (2012) tarafından, bir diğeri ise Aygül, Karaalioğlu ve Amasyalı (2014) tarafından (2000 adet basit cümleyi %75 doğruluk ile öğelerine ayırabilmiş) yapılmıştır.

Bilgi Çıkarımı

Bilgi çıkarımı, kaynak metni sıralı kelimeler olarak ele alarak anlamsal özelliklerini ortaya çıkarmaktadır. Yani cümle öğelerinin, cümledeki yeri de dikkate alınarak varlık isimlerinin tanınması, daha önceden belirlenmiş senaryoların çıkarılması ve varlıklar ile eylemler arası ilişkilerin çıkarılmasıdır. Metin içinden bilgi çıkarımı çalışmalarından biri Adalı (2009) tarafından yapılmıştır. Adalı çalışmasında, biçimbirimsel analiz modülü, belge yapı analiz modülü, alan ontolojisi ve çıkarım ontolojisinden oluşan bir model önermiştir. Alan ontolojisi çıkarırken de Wordnet’i temel almıştır.

Cucerzan ve Yarowsky (1999) tarafından yapılan dil bağımsız çalışma, Türkçe metinlerdeki varlık isimlerinin tespiti için de kullanılabilir. Bunun dışında kişi isimlerinin tespit edildiği çalışmalar ve kural tabanlı varlık ismi tanımlama çalışmaları bulunmaktadır (Bayraktar ve Taşkaya Temizel, 2008; Küçük ve Yazıcı, 2009; Sekine ve Nobata, 2004; Şeker ve Eryiğit, 2012; Özkaya ve Diri, 2012).

Özger ve Diri (2012) ise varlık ismi tanımlama ile ilgili yaptıkları çalışmayla ilgili sonuçlarını detaylı olarak paylaşmıştır. Kişi ismi, yer ismi, kurum ismi, tarih, saat ve para ile ilgili kelimeleri yazılımsal olarak tespit ettikleri çalışmalarında kişi isimlerini %84, yer isimlerini %83, kurum isimlerini %86, tarihsel ifadeleri %92, saat ifadelerini %94 ve parasal ifadeleri ise %96 doğrulukla tespit edebildiklerini belirtmişlerdir. Çalışmalarında varlık isimlerinin bulunduğu herhangi bir sözlükten faydalanmamışlar, önceden belirledikleri anahtar kelimelerden önce veya sonra gelen kelimeleri varlık ismi olarak etiketleme yoluna gitmişlerdir. Özel isimleri tespit ederken mesela, büyük harf ile başlayan kelimeleri alarak bu kelimedenden önce veya sonra unvan veya hitap ifadelerinin gelme durumuna göre özel isim yani kişi ismi olarak etiketlemişlerdir.

Bilgi çıkarımı alanında cümle öğelerine rollerin otomatik olarak atandığı (Swier ve Stevenson, 2004) ve rol kütüphanelerinin kullanıldığı (Palmer, Gildea ve Xue, 2010) çalışmalar da bulunmaktadır. Metindeki bilgiler çıkarıldıktan sonra bilgi, yapısal forma sokularak veri tabanlarına kaydedilebilmektedir.

Soru Tespiti

Soru işareti ile biten ve içinde soru kelimesi veya soru eki (“mi” ve türevleri) bulunan cümlelerin, soru anlamı içerdiği söylenebilir. Türkçede soru tespitinde kullanılan kelimelere “niye”, “kaç”, “nerede”, “ne”, “kim”, “nasıl”, “neden” ve “niçin” kelimeleri örnek verilebilir. Bu kelimeler, cümle içinde başka kelimeler ile grup oluşturarak veya sonlarına ek alarak da karşımıza çıkabilmektedir. Başka kelimelerle grup oluşturduğunda ise bazen soru anlamı içermemektedir (“kaçıncı bu ya!” ifadesindeki gibi). Özger ve Diri (2012), çalışmasında soru kalıplarını ve örüntüleri tespit ederek bu örüntülerin bulunduğu cümleleri soru cümlesi olarak işaretleme yoluna gitmiştir (%87 başarı oranıyla) ve soru cümlesine benzemesine rağmen soru olmayan durumları 6 başlık altında sıralamıştır:

1. Davet
2. Öneri alma
3. Fikir danışma
4. Cevaplı soru (soru cümlesi ardınca cevabının bulunması)
5. Rica
6. Retorik soru (sorunun cevap beklememesi)

Soru yanıtlama çalışması öncesinde metinde geçen soru ifadesinin tespitinin önem arz ettiği kanaatindeyiz. Sonraki aşamada, yani soru tespit edildikten sonra, soruya en uygun bilginin getirilmesi aşaması gelmektedir. Öncelikle bilgi tabanından kayıt sorgulanarak ilgili cevap üretilebilir. Bilgi tabanı olmadığı durumda ise soru yanıtlamanın aşamaları, metin parçalarının sınıflandırılması, bilgi çıkarımı yapılması ve soruya uygun cevap üretilmesi şeklinde olacaktır. Soru cevaplama sistemlerine, SCISOR yazılımı örnek verilebilir. Derici ve arkadaşları (2014), yarı otomatik soru cevaplama sistemi tasarlamıştır.

Yazıyı Seslendirme

Yazıyı seslendirme çalışmalarında, yapay ses üretimi ve doğal seslerin kaydedilmesi yöntemleri uygulanmaktadır. Metinden ses üretiminde kullanılan yapay ses üretimi sistemleri, 70'li yıllardan beri geliştirilmektedir. Bu sistemler, geliştirilme dönemlerine göre: Bell Labs TTS (1973), MITalk (1979), Infovox (1982), CNET PSOLA (1985), ETI Eloquence (1988), Festival TTS (1998), MBROLA (1998) ve Whistler (2000) olarak sıralanabilir.

Türkçe için doğal seslerin kaydedilmesi yöntemi ise TDK tarafından her bir kelimenin seslendirilmesi ile uygulanmıştır (TDK, 2019). Ancak bu sözlükte bazı kelimeler için erkek sesi kullanılırken bazı kelimeler içinse kadın sesi kullanılmıştır. Dolayısıyla ses üretiminde bu sözlüğün kullanılması sorun teşkil edebilmektedir.

Sesin vurgu ve tonlamasının ayarlanarak doğal bir ses elde edilebilmesi için ve anlaşılabilirliğini artırmak için sesin şiddetinin ve süresinin ayarlanması gerekmektedir. Kurallı bir yapının bulunmaması dolayısıyla vurgu ve tonlama için matematiksel bir model uygulanamamaktadır. Belirlenen genel kural, hece içindeki sessizlerin 60 ms, seslilerin ise 90 ms olması gerektiğidir. Ancak Şentürk ve Adalı'ya göre bu kural, sözcüğün son hecesi söz konusu olduğunda, sessizlerin süresinin 90 ms'ye çıkması, seslilerin süresinin ise 120 ms olması şeklinde düzenlenmelidir (Şentürk ve Adalı, 2011).

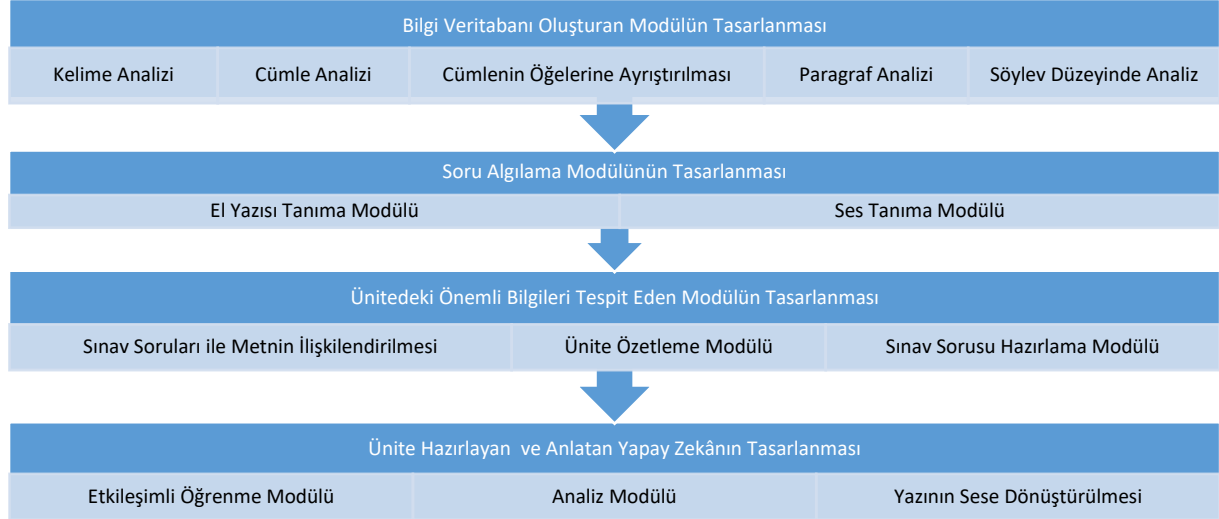
Şentürk ve Adalı (2011) tarafından yapılan aynı çalışmada Türkçe sesler, "MBROLA" programı kullanılarak dijital olarak oluşturulmuş ve oluşturulan ses dosyaları arasına genlik dengeleme (vurgu için) uygulanmıştır. En anlaşılır ses uzunluğunun ise 65 ms olduğunu tespit edilmiştir. 2000'li yıllarda SESTEK firması tarafından geliştirilmiş olan "TTS" yazılımı sayesinde ise Türkçe için oluşturulan sesin doğallığı hususunda ciddi başarı elde edildiği ifade edilmiştir (Sestek, 2019).

Açık Kaynak Kodlu DDİ Çalışmaları

- İTÜ NLP: <http://tools.nlp.itu.edu.tr>
- TS Corpus: <https://tscorpus.com/ts-nlp-toolkit/>
- Zemberek: <http://zembereknlp.blogspot.com>, <https://github.com/ahmetaa/zemberek-nlp>
- Kemik: <http://www.kemik.yildiz.edu.tr>
- Işık Üniversitesi – Olcay Taner Yıldız
- [Http://ddil.isikun.edu.tr/mortur/](http://ddil.isikun.edu.tr/mortur/)
- <http://haydut.isikun.edu.tr/nlptoolkit.html>
- <http://haydut.isikun.edu.tr/nlptoolkit.ui-1.0/#!Main>
- WordNet: <https://wordnet.princeton.edu/download/current-version>
- WS4J, WordNet' de bulunan benzerlik fonksiyonları için API hizmeti: <https://code.google.com/archive/p/ws4j/>

MODEL ÖNERİSİ

Ders anlatan yapay zekâ modülü tasarlanırken öncelikle metin içinden bilgi çeken ve bilgi veri tabanını oluşturan/besleyen modülün tasarlanması gerekmektedir (Şekil 1). Basit cümle yapılarından başlanarak cümle, öğelerine ayrıştırılmalıdır. Cümlenin her bir öğesinin bilgi bütünü içinde bir parçayı temsil ettiği gerçeğinden hareketle de veri tabanı oluşturabilmektedir. Daha sonra birleşik ve karmaşık cümle yapıları ve ardışık cümleler de çalışmaya katılarak kapsam genişletilmelidir.

Şekil 1. Ders anlatan yapay zekâ yol haritası

Yapay zekâ modülü, öğrenciyle etkileşimli olarak tasarlanacağı için öğrenci tarafından sorulan soruların tespiti ve sistem tarafından sorulan sorulara öğrenci tarafından verilen cevapların değerlendirilebilmesi gerekmektedir. Bunun için ses tanıma ve el yazısı tanımlama modüllerinin tasarlanması gerekmektedir.

Elde edilmek istenen öğrenme çıktılarıyla uyumlu olmayan soru hazırlanmasını önlemek için soru hazırlamaya geçmeden önce üniteye önemli bilgilerin bulunduğu metin parçalarının tespit edilmesi gerekmektedir. Bunu yapmak için de daha önce alan uzmanları tarafından hazırlanmış sınav soruları analiz edilerek yaygın kullanılan soru kalıpları tespit edilebilir. Soru havuzundaki bu sorulardan ilgili kalıplar çıkarıldığında geri kalan kelimelerin birlikte geçtiği metin bloğu, birliktelik analizleri kullanılarak, ünite içinde bulunur. Böylece sınav sorularının hazırlandığı metin bloklarının ortak özellikleri tespit edilmiş olur. Böylece öğretim çıktılarıyla doğrudan ilişkili olan içerikler tespit edilebilir ve bunlar birleştirilerek, sınav sorularının otomatik olarak hazırlanmasında kullanılacak ünite özetleri hazırlanabilir.

Soru metninin hazırlanmasında bulunabilen çeldirici sayısı dikkate alınarak, hazır soru kalıplarından faydalanılabilir. Metnin hangi bölümünün soruya dönüştürülmesinin daha doğru olacağını tespiti için de farklı varyasyonlarla hazırlanan sorular kullanıcılara oylatılabilir. Çeldiricilerin tespit edilebilmesi için de anlamsal sözlüklerden ve terim ağırlıklandırılmadan faydalanılabilir. İlgili konuda sıklıkla geçen ancak kitap bütününde daha az sıklıkta geçen ifadeler alana özgü kelimeler olarak kabul edilebilir. Doğru cevapla aynı paralelde ve yan anlamda olan ifadeler çeldirici olarak kullanılabilir.

Yapay zekânın ünite hazırlayabilmesi için öncesinde etkileşimli öğrenme modülü tasarlanması ve bu modülün kullanım verileri analiz edilerek en uygun ünitenin düzenlenmesi gerekmektedir. Ünitenin seslendirilmesinde ise yazının sese dönüştürülmesi modülünden faydalanılabilir.

TARTIŞMA VE SONUÇ

Ders anlatan yapay zekâ projesi Türkçe için tasarlanmak istendiğinde Türkçe doğal dil işleme çalışmaları üzerine temellendirilmelidir. Bu bağlamda bu çalışma ile Türkçede yapılan çalışmalar detaylı olarak incelenmiş ve mevcut durum tespiti yapılmıştır. Kural tabanlı olarak yürütülen dil bağımlı çalışmaları, kelime düzeyinde, söz dizimsel düzeyde, anlam ve söylev düzeyinde olmak üzere 4 düzeyde yürütülmektedir.

Kelime düzeyinde yapılan çalışmalar, özellikle kelime kökünün ve eklerinin bulunmasına odaklanmaktadır. Böylece kelimenin doğru yazıldığı kontrol edilebilirken, kelime türüne de karar verilebilmektedir. Kelime türü ve kelimeye gelen ekler ise kelimenin cümledeki görevini belirginleştirmektedir. Bu bakımdan cümle analizi düzeyindeki çalışmalar, kelime düzeyindeki çalışmalardan beslenmektedir. Cümle düzeyinde, cümlelerin bağıllık analizi çalışmaları yapılmaktadır.

Anlam düzeyinde yapılan çalışmalar ise anlamsal sözlüklerden beslendiği için Türkçe üzerine yapılan çalışmalar, özellikle İngilizce için yapılan anlamsal sözlüklerin Türkçeye çevrilmesi ve derlemlerden faydalanılarak anlamsal sözlüklerin oluşturulması olmak üzere iki ana eksen üzerinde yürütülmektedir. Bu ikinci çalışma ise derlem oluşturma çalışmalarından beslenmektedir.

Cümle içinde geçen kelimelerin hangi terim anlamında veya mecazi anlamda kullanıldığının anlaşılmasına odaklanan çalışmalar ise söylev düzeyinde yürütülmektedir. Bu bağlamda, kelimenin, cümledeki ve kelime grubu içindeki konumu dikkate alınarak, yaygın ve gündelik kullanımının tespit edilmesi önemlidir.

Yapay zekânın insan ile etkileşiminde sesin yazıya ve yazının sese çevrilmesi çalışmaları ile el yazısının ve matbu yazının tespit edilmesi çalışmaları kullanılabilir. Bu aşamadan sonra varsa yazım yanlışlarının tespiti ve düzeltilmesinden sonra soru ifadelerinin tespiti önem arz etmektedir. Metnin öğelerinin tespiti, metnin anlaşılması ve uygun cevap metinlerinin oluşturulması çalışmaları gelmektedir. Ders ünitelerinin özetlenmesi ve metindeki bilgilerin tespit edilmesi çalışmaları ise kendi kendine ünite hazırlayabilen yapay zekâ çalışmasını besleyecektir. Bu bağlamda yapay zekâyâ giden yolda geliştirilmesi gereken modüller ve teknik detaylarından bahsedilmiştir. Bu çalışma ile ayrıca, bireysel veya grup içinde, Türkçe için açık kaynak kodlu olarak geliştirilen ve/veya servis olarak kullanıma sunulan yazılımların erişim adresleri listelenmiştir.

KAYNAKÇA

Adalı, Ş. (2009). Türkçe Belgelerden Bilgi Çıkarımı için Bir Tümeleştirmeci Mimari. Doktora Tezi. İTÜ, Fen Bilimleri Enstitüsü, İstanbul.

Amasyalı, M. F. (2012). Kavramlar Arası Anlamsal İlişkilerin Türkçe Sözlük Tanımları Kullanılarak Otomatik Olarak Çıkarılması/matic Extraction of Semantic Relationships using Turkish Dictionary Definitions. EMO Bilimsel Dergi, 1(1): 1-14.

Aşliyan, R. ve Günel, K. (2007). YakhnoDetecting Misspelled Words in Turkish Text Using Syllable N-Gram Frequencies. Pattern Recognition and Machine Intelligence, 18-22 December, Kolkata, India.

Aygül, M., Karaalioglu, G. ve Amasyalı, F. (2014). Prediction of Function Tags of the Simple Turkish Sentences by Conditional Random fields. Sigma, 32: 23-30.

Bawakid, A. ve Oussalah, M. (2008). A Semantic Summarization System: University of Birmingham at TAC 2008. First Text Analysis Conference (TAC 2008), 17-19 November, Maryland, USA.

Bayraktar, Ö. ve Taşkaya Temizel, T. (2008). Person Name Extraction From Turkish Financial News Text Using Local GrammarBased Approach”, The 23rd International Symposium on Computer and Information Sciences-ISCIS, 27-29 October, İstanbul, Turkey.

Benitti, F. B. V. (2012). Exploring the Educational Potential of Robotics in Schools: A Systematic Review. Computers and Education, 58(3): 978-988.

Cebiroglu, G. (2002). Sözlüksüz Köke Ulaşma Yöntemi. Doktora Tezi. İTÜ, Fen Bilimleri Enstitüsü, İstanbul.

Cucerzan, S. ve Yarowsky, D. (1999). Language Independent Named Entity Recognition Combining Morphological and Contextual Evidence. Empirical Methods in Natural Language Processing and Very Large Corpora, 21-22 June, MD, USA.

Çabuk, H., Yüksel, Ç., Mocan, Z., Diri, B. ve Amasyalı, M. F. (2003), Metin Analizi ve Sorgulama (MAvS), Koç Üniversitesi İstanbul, 1-11.

Çınar, M. (2003). Veri Ambarlarında Verilerin Temizlenmesi. Doktora Tezi. İTÜ, Fen Bilimleri Enstitüsü, İstanbul.

Çolakoğlu, Ö. M., & Akdemir, Ö. (2010). Motivational Measure of The Instruction Compared: Instruction Based On The ARCS Motivation Theory v.s. Traditional Instruction In Blended Courses, Turkish Online Journal of Distance Education, 11, 2, 73-89.

Delibaş, A. (2008). Doğal Dil İşleme ile Türkçe Yazım Hatalarının Denetlenmesi. Yüksek Lisans Tezi. İTÜ, Fen Bilimleri, İstanbul.

Derici, C., Çelik, K., Özgür, A., Güngör, T., Kutbay, E., Aydın, Y. ve Kartal, G. (2014). Rule-Based Focus Extraction in Turkish Question Answering Systems. 22nd Signal Processing and Communications Applications Conference (SIU), 23 - 25 April, Trabzon, Turkey.

- Dilsiz, S. (2005). Bulanık Mantık ve Yapay Sinir Ağları ile Türkçe Yazım Denetleyicisi, Yüksek Lisans Tezi, İTÜ, Fen Bilimleri Enstitüsü, İstanbul.
- Diñçer, Ö. G. S. (2016). Bilgisayar Destekli Eğitim Ve Uzaktan Eğitime Genel Bir Bakış. Adana, Seyhan, Türkiye.
- Doğan, S. ve Karaağaç, G. (2012). Dilbilgisel Bağdaşmazlık ve Anlamsal Tutarsızlık. İstanbul Aydın Üniversitesi Dergisi, 16: 25-42.
- Dönmez, İ. (2016). Türkçe Tümcelerin Yüklem Odaklı Anlam ve Dilbilgisi Çözümlemesi, Doktora Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Eryiğit, G., Oflazer, K. ve Adalı, E. (2006). Türkçe Cümlelerin Kural Tabanlı Bağlılık Analizi. TAINN 2006 Fifteenth Turkish Symposium on Artificial Intelligence and Neural Networks, 21-23 Haziran, Akyaka, Muğla, Türkiye.
- Fırat, Y., Kılıçaslan, Y. ve Uçar, Ö. (2013). Bilgisayar Ortamında Biçimsel Ontoloji Oluşturulması. Journal of Higher Education & Science/Yükseköğretim ve Bilim Dergisi, 3(1).
- Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., ve Ivaldi, S. (2016). Trust as Indicator of Robot Functional and Social Acceptance. An Experimental Study On User Conformation to ICub Answers. Computers in Human Behavior, 61, 633-655.
- Güngör, O. ve Güngör, T. (2007). Türkçe bir Sözlükteki Tanımlardan Kavramlar Arasındaki Üst-Kavram İlişkilerinin Çıkarılması. Akademik Bilişim Konferansı, 1(1): 1-13.
- Güzey, C. ve Oflazer, K. (1994). Spelling Correction in Agglutinative Languages. Bilkent University Department Of Computer Engineering and Information Systems Technical Report, BU-CEIS-94-01, Ankara, Turkey.
- Hashimoto, T., Kato, N. ve Kobayashi, H. (2011). Development of Educational System with the Android Robot SAYA and Evaluation. International Journal Advanced Robotic Systems, 8(3): 51-61.
- Hu, M. ve Liu, B. (2004). Mining and Summarizing Customer Reviews. Tenth ACM SIGKDD International Conference on Knowledge discovery and data mining, 22-25 August, Washington, U.S.A.
- Ittoo, A., Bouma, G., (2010). On Learning Subtypes of The Part-whole Relation: Do Not Mix Your Seeds. Association for Computational Linguistics, 11-16 July, Uppsala, Sweden.
- İşman, A. (2008). Uzaktan eğitim. Pegem Akademi.
- Kanda, T., Hirano, T., Eaton, D. ve Ishiguro, H. (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. Human Computer Interaction, 9: 61-84.
- Kul, S. (2018). Türkçe Tweetlerin Analiz Edilebilmesi için Hadoop/Hive Kullanan Melez Bir Sistemin Geliştirilmesi. Doktora Tezi. Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü, Erzurum.
- Küçük, D. ve Yazıcı, A. (2008). Identification of Coreferential Chains in Video Texts for Semantic Annotation of News Videos”, The 23rd International Symposium on Computer and Information Sciences-ISCIS, 27-29 October, Istanbul, Turkey.
- Küçük, D. ve Yazıcı, A. (2009). Exploiting Information Extraction Techniques for Semantic Annotation of Videos in Turkish. The 14th International Conference on Applications of Natural Language to Information Systems-NLDB, 24-26 June, Saarland, Germany, 2009.
- Liang, T. P., Lai, H. J., & Ku, Y. C. (2007). Personalized Content Recommendation and User Satisfaction: Theoretical Synthesis and Empirical Findings. Journal of Management Information Systems , 23 (3), 45-70.
- Mavi, A. Makine Öğrenmesi ile El Yazısı Tanıma Programı. ardamavi.com/2017/04/el-yazisi-tanma.html, (12.11.2019).
- Movellan, J., Eckhart, M., Virnes, M. ve Rodriguez, A. (2009) Sociable Robot Improves Toddler Vocabulary Skills. Proceedings of 2009 International Conference on Human Robot Interaction, California, USA.
- Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A. ve Dong, J. J. (2013). A Review of the Applicability of Robots in Education. Technology for Education and Learning, 1: 1-7.
- Nevres, B. İşinize Yarayacak En İyi 4 OCR Yazılımı. <https://mediatrend.mediamarkt.com.tr/isinize-yarayacak-en-iyi-4-ocr-yazilimi>, (01.11.2019).

- Nivre, J. ve Nilsson, J. (2005). Pseudo-Projective Dependency Parsing. The 43rd Annual Meeting on Association for Computational Linguistics, Haziran, Stroudsburg, USA.
- Oflazer, K. (1994). Two-level description of Turkish morphology. *Literary and linguistic computing*, 9(2): 137–148.
- Oflazer, K. (2003). Dependency Parsing with an Extended Finite-State Approach. *Computational Linguistics*, 29(4): 515-544.
- Oflazer, K. (2005). Biçimbirimsel Çözümleyici Sonuçları Edinimi.
- Oflazer, K. ve Solak, A. (1992). Parsing Agglutinative Word Structures And Its Application to Spelling Checking for Turkish. In *Proceedings of the 15th International Conference On Computational Linguistics*, Nantes, France, August 23-28, 39-45.
- Özger, Z. B. ve Diri, B. (2012). Türkçe Dokümanlar için Kural Tabanlı Varlık İsmi Tanıma. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2): 1-11.
- Özkaya, S. ve Diri, B. (2011). Named Entity Recognition by Conditional Random Fields from Turkish Informal Texts". In *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)*, 20-22 April, Antalya, Turkey.
- Özköse, C. ve Amasyalı, M. F. (2012). Tümce Ögelerinden Hayat Bilgisi Çıkarımı. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2): 1-11.
- Öztürk, M. B. (2016). Türkçede Morfolojik Analiz Yapan Bir Sistemin Morfolojik Türetme için Kullanılması. Yüksek Lisans Tezi. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Palmer, M., Gildea, D. ve Xue, N. (2010). Semantic Role Labeling. *Synthesis Lectures on Human Language Technologies*, 3(1): 1–103.
- Perkinselearning. "Seeing AI - How About that Handwriting Feature?", <https://www.perkinselearning.org/technology/blog/seeing-ai-how-about-handwriting-feature>, (12.11.2019).
- Rivera, D. (2005). The effect of content customization on learnability and perceived workload. *Conference on Human Factors in Computing Systems*, Portland, Oregon, USA, 1749–1752.
- Rs., "The Jobs AI Will Take Over First", <https://uk.rs-online.com/web/generalDisplay.html?id=the-jobs-ai-will-take-over-first>, (01.01.2020).
- Sak, H., Güngör, T. ve Saraçlar, M. (2011). Resources for Turkish Morphological Processing. *Language Resources and Evaluation*, 45(2): 249–261.
- Sekine, S. ve Nobata, C. (2004). Definition, Dictionaries and Tagger for Extended Named Entity Hierarchy. *The 4th International Conference on Language Resources and Evaluation*, May, Lisbon, Portugal.
- Sestek, "Sestek" <https://ttsdemo.sestek.com/demo.aspx>, (20.11.2019).
- Sezer, B. (2015). Kişiselleştirilmiş Matematik Problemlerinin Akademik Başarıya Etkisi. *Eğitim Teknolojisi Kuram ve Uygulama*, 5(2), 71-88.
- Solak, A. ve Oflazer, K. (1993). Design and Implementation of a Spelling Checker for Turkish. *Literary and Linguistic Computing*, 8(3): 113–130.
- Swier, R. S. ve Stevenson, S. (2004). Unsupervised Semantic Role Labelling, *EMNLP*, 25-26 July, Barcelona.
- Şahin, G. (2016). Türkçe Derlemden Anlamsal İlişkilerin Çıkarılması. Yüksek Lisans Tezi. Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Şahin, G. ve Amasyalı, F. (2014). Geniş Metin Koleksiyonlarından İteratif Bilgi Çıkarımı Iterative Information Extraction from Large Text Collections. *EMO Bilimsel Dergi*, 4(7): 13-20.
- Şahin, G., Diri, B. ve Yıldız, T. (2015). Pattern and Semantic Similarity Based Automatic extraction of hyponym-hypernym relation from Turkish corpus. *Signal Processing and Communications Applications Conference (SIU)*, 674–677.
- Şeker, Ş. E. Doğal Dil İşleme (Natural Language Processing), *YBS Ansiklopedi*, 2(4), 2015.
- Şeker, G. A. ve Eryiğit, G. (2012). Initial Explorations on Using CRFs for Turkish Named Entity Recognition". In *Proceedings of COLING 2012*, December, Mumbai, India.
- Şentürk, F. "Yetkin Bul ve Değiştir", *TBV BBHD*, 3, 2010.

- Şentürk, T. ve Adalı, E. (2011). Türkçe Metin Seslendirme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 4(1).
- Şerbetçi, A., Orhan, Z. ve Pehlivan, İ. (2011). Extraction of Semantic Word Relations in Turkish From Dictionary Definitions. ACL 2011 Workshop on Relational Models of Semantics, June, Oregon, USA.
- TDK, “Türk Dil Kurumu Sözlükleri”, <https://sozluk.gov.tr/?kelime=SESL%C4%B0>, (14.11.2019).
- Tülek, M. (2007). Türkçe için metin özetleme (Doctoral dissertation, Fen Bilimleri Enstitüsü).
- Viilup, A. Ücretiz En İyi 6 OCR Aracı. <https://lightpdf.com/tr/ucretsiz-ocr-programi.html>, (01.11.2019).
- Viilup, A. En Harika 4 El Yazısı Tanıma Uygulaması. <https://lightpdf.com/tr/el-yazisi-tanima-uygulamasi.html>, (15.11.2019).
- Vural, E., Erdogan, H., Oflazer, K. ve Yanikoglu, B. (2004). An Online Handwriting Recognition System for Turkish. The IEEE 12th Signal Processing and Communications Applications Conference, 25-30 April, Kusadasi, Turkey.
- Yazıcı, E. ve Amasyalı, M. F. (2011). Automatic Extraction of Semantic Relationships Using Turkish Dictionary Definitions. EMO Bilimsel Dergi, 1(1): 1-13.
- Yıldırım, S. ve Yıldız, T. (2012). Automatic Extraction of Turkish Hypernym-Hyponym Pairs From Large Corpus. 24th International Conference on Computational Linguistics, 8 -15 December, IIT Bombay, India.
- Yıldız, T., Yıldırım, S. ve Diri, B. (2013). Extraction of Part-Whole Relations from Turkish Corpora. Springer, 126–138, 2013.
- Yun, S., Shin, J., Kim, D., Kim, C. G., Kim, M. ve Choi, M. T. (2011). Engkey: Tele-Education robot. In International Conference on Social Robotics, 142-152.