

# Türkçe Sözlü Ders Anlatımları için Soru Cevaplama Sistemi

## A Question Answering System for Spoken Lectures in Turkish

Merve Ünü<sup>a,b</sup> Ebru Arısoy<sup>c</sup> Murat Saraçlar<sup>a</sup>

<sup>a</sup> Boğaziçi Üniversitesi, İstanbul, Türkiye

<sup>b</sup> Galatasaray Üniversitesi, İstanbul, Türkiye

<sup>c</sup> MEF Üniversitesi, İstanbul, Türkiye

**Özetçe**—Bu bildiri Türkçe sözlü ders anlatımları için bir soru cevaplama sistemi anlatılmaktadır. Bu çalışmada Türkçe sözlü ders anlatımlarının otomatik yazılandırılması ve geri getirmeye üzerine yapılan çalışmalar ileriye taşınarak, yazılı olarak sorulan bir soruyu cevaplamak için sözlü ders anlatımının içindeki cevabı gösteren bir sistem geliştirilmiştir. İlk aşamada, sözlü ders içerikleri, bir Türkçe konuşma tanıma sistemi kullanılarak yazıya çevrilir. Daha sonra yapay sinir ağları kullanılarak hem ders içerikleri hem de sorular için bir vektör gösterimi elde edilir. Bu gösterimler içerik içinde cevabın olduğu yeri belirlemek için kullanılır. Bu çalışmada Türkçe hukuk derslerinin hem elle hem de otomatik olarak yazıya çevrilmesiyle elde edilen veriler kullanılmıştır. Deneyler sonucunda konuşma tanıma sistemi çıkışının sistem başarımını sadece biraz düşürdüğü görülmüştür.

**Anahtar Kelimeler**—soru cevaplama, sözlü ders anlatımlarının işlenmesi.

**Özet**—This paper presents a question answering (QA) system for spoken lectures in Turkish. Extending the previous work on Turkish spoken lecture transcription and retrieval, we present a system that gives an answer to a written question by pointing to the answer within a spoken lecture. Spoken lectures are automatically transcribed using a Turkish speech recognition system. We make use of bidirectional recurrent neural networks to obtain vector representations for both the documents and the questions which are then used to determine an interval for the answer. The experiments are performed on both the manual and automatic transcription of Turkish spoken law lectures. The performance of the system degrades only slightly when the speech recognition outputs are used instead of the references.

**Keywords**—question answering, spoken lecture processing.

### I. GİRİŞ

Eğitim amaçlı ders videoları öğrencilerin sınıf dışında da esnek bir şekilde ve kolayca bilgiye erişebilmesini sağlamaktadır. Bu videolar gerek yaşam boyu eğitim gerekse örgün eğitim için yaygın olarak kullanılmaya başlanmıştır. Her ne kadar Türkçe için bu tür kaynaklar günümüzde nispeten sınırlı olsa da yakın gelecekte bu kaynaklar hızla artacaktır. Konuşma ve dil işleme yöntemlerinin gelişmesiyle bu tür kaynaklara erişim de kolaylaşmaktadır.

Konuşma tanıma sistemleri, sözlü ders anlatımlarının yazılandırılarak hem öğrencilerin konuşulanları daha rahat takip edebilmesi, hem de işitme engelli öğrencilerin bilgiye erişimi için kullanılmaktadır [7], [10], [15]. Konuşma tanıma sistemleri sadece yazılandırma için değil konuşma içeriğinin aranması ve geri getirmeye için de faydalıdır [5], [6].

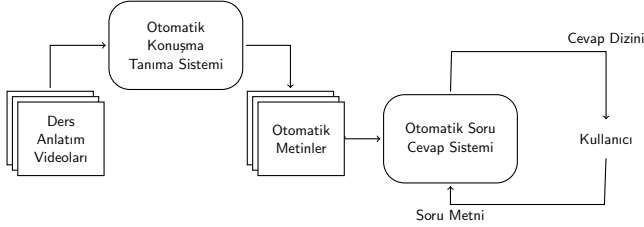
Bu çalışmada konuşma ve dil işleme yöntemleri kullanılarak derslerin yazılandırılması ve geri getirmeye tamamlayıcı şekilde bir soru cevaplama sistemi geliştirilmiştir. Bu bağlamda, soru cevaplama sistemleri yazılı veya sözlü bir soruya yazılı veya sözlü içeriğin içinden bir cevap bulmayı hedefler. Bu sistemler her ne kadar sorulan soruyu anlayarak cevap üretmeye de, sözlü içerik içinde anahtar sözcük arama veya verilen bir sorguya ilişkin bölümü geri getirmeye ötesinde bir işlem gerektirir.

Makine anlaması olarak ifade edilen soru cevaplama sistemleri, sisteme sunulan sorunun cevabını kısa metin paragrafları içinde bulmayı hedefler. Makine anlaması sistemlerinde son yıllarda kullanılan uçtan-uca yapay sinir ağları sayesinde yüksek başarımlar elde edilmektedir. Özellikle, İngilizce için toplanan ve makine anlaması sistemlerinin geliştirilmesinde yaygın olarak kullanılan SQuAD [12], [13] veri setinde insan başarımına yaklaşan sonuçlar elde edilmiştir [4], [19], [20]. Yazılı metinler dışında ses verileri kullanılarak da geliştirilen soru cevaplama sistemleri vardır [8], [16]. Ancak soru cevaplama sistemini ses verileri için geliştirmek yazılı metinler için geliştirmekten hem eğitim için kullanılacak kaynakların kısıtlı olması hem de ses verilerinin yazılandırılmasında kullanılan konuşma sistemlerinin hatalı metinler üretebilmesinden dolayı daha zor bir problemidir.

Konuşma ve dil işleme yöntemlerinin Türkçe ders metinleri ve videolarına uygulandığı çalışmalar sınırlıdır. Önceki çalışmalarda Türkçe ders anlatım videoları için konuşma tanıma ve geri getirme sistemleri [2] ve ders metinleri için doğal dil işleme ve bilgi geri getirmeye dayanan soru cevap sistemleri [3] geliştirilmiştir. Bu bildiriye anlatılan soru cevaplama sistemi ders anlatım videoları için uçtan-uca yapay sinir ağları kullanılarak geliştirilmiştir.

## II. YÖNTEM

Ders anlatım metinleri üzerinden geliştirilen soru cevaplama sisteminin detayları Şekil 1 üzerinde gösterilmiştir. İlk olarak ders anlatım videolarından otomatik konuşma tanıma sistemi kullanılarak anlatım metinleri elde edilmiş, sonra bu metinler soru cevaplama sisteminde kullanılmıştır. Kullanıcı ders anlatımları ile ilgili bir soru sorduğunda soru cevaplama sistemi sorunun cevabını metinler içerisinde arayıp, kullanıcıya cevap olarak bir metin alt dizisi geri döndürür.



Şekil 1: Soru cevaplama sistemi

Soru cevaplama sisteminin temel görevi kullanıcı tarafından girilen soruya en uygun cevabı verilen paragraf içerisinde bulmaktır. Bu cevap paragraftaki sıralı kelimelerden oluşur, yani paragrafın bir parçasıdır. Yapay sinir ağları kullanılarak geliştirilen sistemlerde kullanıcı bir soru girdiğinde ilk olarak soru ve paragraf kelimeleri arasında benzerlik hesabı yapılır. Sistem paragrafın ve sorunun kelimeleri arasındaki eşleşmeyi kullanarak paragraf içindeki en olası kelime dizisini geri döndürür. Bu modelin eğitilmesinde kullanılan veri setinde paragraf, soru ve cevap kelime dizisi bulunur. Paragraflar ders anlatım videolarının alt parçalarıdır. Sorunun cevabı kendisiyle birlikte girdi olarak verilen paragrafın içerisinde aranır.

Bu çalışmada soru cevaplama sistemi geliştirilirken uçtan-ucaya öğrenen bir yapay sinir ağı modeli olan AnswerPointer MatchLSTM kullanılmıştır [18]. Bu modelin temeli Uzun Kısa Vadeli Bellek ağlarına ve İşaretçi Ağlarına (Pointer Network [17]) dayanmaktadır. Stanford Üniversitesi tarafından geliştirilen soru-cevap veri seti (SQuAD) [12] üzerinde iyi sonuçlar elde etmiştir.

MatchLSTM verilen bir paragraf içerisinde sorulan soruya cevap olacak diziyi yine paragraf içerisinde bulur. Cevabı ararken paragraf ve soru arasında kelime tabanlı eşleme modeli kullanır. Model temel olarak üç katmanda incelenebilir:

- 1) **Ön İşleme** Bu katmanın temel amacı verilen paragrafın ve sorunun her bir sözcüğünü modelde kullanılacak dahili gösterime dönüştürmektir. Paragraf  $\mathbf{P}$  ve soru  $\mathbf{Q}$  tek yönlü LSTM ile şu şekilde gösterilebilir:

$$\mathbf{H}^p = \overrightarrow{LSTM}(\mathbf{P}), \mathbf{H}^q = \overrightarrow{LSTM}(\mathbf{Q})$$

$\mathbf{H}^p$  paragraf için,  $\mathbf{H}^q$  ise soru için dahili gösterimi ifade eder.

- 2) **MatchLSTM** Bu katman çift yönlü LSTM modelinden oluşmaktadır. Paragraf ardışık olarak işlenir ve dikkatin çift yönlü olarak elde edildiği vektör üretilir. Dikkat vektörü  $\vec{\alpha}_{ij}$  paragrafın  $i$  konumundaki sözcük ile sorunun  $j$  konumundaki sözcük arasındaki eşleşmeyi gösterir.

Paragrafın  $i$  konumundaki sözcük için öncelikle  $\vec{\alpha}_i$  ileri yönde dikkat vektörü hesaplanır:

$$\vec{\mathbf{G}}_i = \tanh(\mathbf{W}^q \mathbf{H}^q + (\mathbf{W}^p \mathbf{h}_i^p + \mathbf{W}^r \vec{h}_{i-1}^r + \mathbf{b}^p) \otimes e_Q)$$

$$\vec{\alpha}_i = \text{softmax}(\mathbf{w}^T \vec{\mathbf{G}}_i + b \otimes e_Q)$$

$\mathbf{W}^q, \mathbf{W}^p, \mathbf{W}^r, \mathbf{b}^p, b$  modelde öğrenilecek parametreler ve  $\vec{h}_i^r$  tek yönlü matchLSTM katmanının saklı vektörüdür. Bu dikkat vektörü kullanılarak saklı vektör şu şekilde hesaplanır:

$$\vec{z}_i = \begin{bmatrix} \mathbf{h}_i^p \\ \mathbf{H}^q \vec{\alpha}_i^T \end{bmatrix}$$

$$\mathbf{h}_i^r = \overrightarrow{LSTM}(\vec{z}_i, \vec{h}_{i-1}^r)$$

Paragraftaki sözcüklerin iki yönlü dahili gösterimini elde etmek amacıyla ileri yöndeki ile aynı hesaplar ters yönde de uygulanır. İleri yöndeki saklı vektörler  $\vec{\mathbf{H}}^r \in \mathbb{R}^{l \times P}$  matrisinin sütunlarını, ters yöndeki saklı vektörler de  $\vec{\mathbf{H}}^r \in \mathbb{R}^{l \times P}$  matrisinin sütunlarını oluşturur. Bu iki matrisin alt alta birleştirilmesi  $\mathbf{H}^r$  gösterimini verir.

- 3) **Cevap İşaretçisi** Bu katmanın temelini işaretçi ağları [17] oluşturmaktadır. Bu katman  $\mathbf{H}^r$  matrisini girdi olarak alır ve iki tane sayı geri döndürür. Bu sayılar cevabın paragraftaki başlangıç ve bitiş sözcüklerinin konumlarıdır. Önceki katmandan gelen  $\mathbf{H}^r$  gösterimi kullanılarak paragraf için iki dikkat vektörü hesaplanır. Bu dikkat vektörleri paragraftaki her sözcük için cevabın başlangıç kelimesi ve bitiş kelimesi olma olasılığını gösterir. Bu hesaplamada tek yönlü LSTM modeli kullanılır. Başlangıç ve bitiş olasılıklarının çarpımının en yüksek olduğu ikili kelimelerin paragraftaki yerleri cevap olarak geri döner. Model eğitilirken maliyet hesabı negatif log olasılık fonksiyonu kullanılarak yapılır.

$$-\sum_{n=1}^N \log p(a_n | P_n, Q_n)$$

Verideki her bir paragraf soru ikilisi  $(P_n, Q_n)$  için en yüksek olasılıklı cevap değerleri bulunur.

## III. DENEYLER

Bu çalışmada geliştirilen soru cevaplama sistemi ders anlatım metinlerinden oluşturulan soru-cevap veri seti kullanılarak eğitilmiş ve sistemin başarımı ders anlatım videoları üzerinde sınanmıştır. Ders anlatım videoları ilk olarak otomatik konuşma tanıma sistemi kullanılarak yazılandırılmış, ardından otomatik olarak üretilen metinler üzerinde sisteme sunulan sorunun cevabı aranmıştır.

Bu bölüm sırasıyla soru cevaplama ve konuşma tanıma sistemlerinde kullanılan verileri, bu sistemlerin detaylarını ve başarımlarını anlatmaktadır.

### A. Veri Seti

Ders anlatım videoları MEF Üniversitesi bünyesindeki stüdyoda yüksek kaliteli ekipmanlar kullanılarak hazırlanmıştır. Bu videolar 10-20 dakika uzunluğundadır ve Hukuk Fakültesi'nden Ceza Hukuku, Borçlar Hukuku, İdare Hukuku,

**Paragraf:** Her suç ahlaksızlık her ahlaksızlık da bu nedenle suç değildir. Hukuk ile ahlakın ortak yönleri olduğu gibi birbirinden ayrı yönleri de bulunmaktadır. Üçüncü ceza hukukunun tamamlayıcı kaynakları ise mahkeme içtihatlarıdır ve bugün Türkiye’de yüksek mahkeme olarak adli yargı mahkemelerince verilen kararların son inceleme yeri olan yargıtay bulunmaktadır. Özellikle ilk derece mahkemelerinin kararlarına karşı yasa yoluna temyiz yoluna başvurunca yargıtay ilgili dairesi bu kararı incelemekte ve kararı bozmakta veyahut onamaktadır. Bozma üzerine ilk derece mahkemesine bu karar geldiğinde ilk derece mahkemesinin bu karara karşı direnme hakkı bulunmaktadır ve direnme kararı verirse ilk derece mahkemesi bu tekrar temyize gidildiğinde bunu ilgili daire değil ceza genel kurulu incelemektedir ve ceza genel kurulunca bozma üzerine verilip ilk derece mahkemesince direnme verilen karara karşı ceza genel kurulunun verdiği karar bağlayıcıdır. Bu nedenle bu direnme kararından sonra ceza genel kurulunun verdiği karara ilk derece mahkemesi uymak zorundadır.

**Soru 1:** Yargıtay nedir?

**Cevap 1:** Türkiye’de yüksek mahkeme olarak adli yargı mahkemelerince verilen kararların son inceleme yeri

**Soru 2:** İlk derece mahkemesi ceza genel kurulunun verdiği karara neden uymak zorundadır?

**Cevap 2:** Ceza genel kurulunca bozma üzerine verilip ilk derece mahkemesince direnme verilen karara karşı ceza genel kurulunun verdiği karar bağlayıcıdır

Şekil 2: Örnek bir paragraf için soru-cevap ikilisi.

Medeni Hukuk, Kriminoloji ve Siyaset Bilimi alanlarında anlatılan dersleri içermektedir. Bu videolardan çıkarılan ses kayıtları elle yazılandırılmış ve bu kayıtlara karşılık gelen referans metinler elde edilmiştir<sup>1</sup>. Güz döneminde verilen 6 farklı dersin ses kayıtları (12.5 saat) ve referans metinleri (75,000 kelime) eğitim verisi, bahar döneminde verilen 2 farklı dersin ses kayıtları (4.3 saat) ve referans metinleri (24,443 kelime) ise sınama verisi olarak ayrılmıştır. Sınama verisi Ceza Hukuku II ve Borçlar Hukuku II derslerini içermekte, Ceza Hukuku I ve Borçlar Hukuku I dersleri ise eğitim verisinde yer almaktadır. Bu veri ile ilgili daha fazla detay Kaynakça [2]’de verilmiştir.

Türkçe hukuk ders anlatımlarından elde edilen referans metinler kullanılarak bir soru-cevap veri seti oluşturulmuştur. Soru-cevap veri seti oluşturulurken ilk olarak referans metinler elle anlamlı paragraflara bölünmüş, ardından her paragraf dikkatlice okunarak bu paragrafta cevabı bulunan bir veya birden fazla soru üretilmeye çalışılmıştır. Her soruya karşılık gelen cevap işaretlenmiştir. Soru-cevap veri seti oluşturulması elle yapılan bir iş olduğu için oldukça zahmetlidir. Bu yüzden eğitim setindeki hukuk ders anlatım referans metinlerinin sadece 4 farklı derse ait bölümü (54,461 kelime) soru-cevap veri seti oluşturmak için işaretlenmiştir. Bu metinler 484 paragrafa bölünmüş ve bu paragraflardan 650 tane soru-cevap ikilisi oluşturulmuştur. Eğitim verisindeki ortalama paragraf uzunluğu 116 kelimedir. Sınama verisi olarak ayrılan iki dersin referans metinleri için de aynı işlem yapılmış ve sınama verisi için 259 paragraf ve bu paragraflardan 286 tane soru cevap ikilisi oluşturulmuştur. Sınama verisindeki ortalama paragraf uzunluğu 96 kelimedir. Veri setinden örnek bir paragraf ve bu paragraftan üretilen soru-cevap ikilileri Şekil 2’de verilmiştir.

### B. Konuşma Tanıma Sistemi

Bu çalışmada kullanılan otomatik konuşma tanıma sistemi Türkçe ders anlatım videolarının yazıya çevrilmesi için geliştirilmiştir [2]. Konuşma tanıma sisteminin akustik model eğitiminde çok katmanlı yapay sinir ağları tabanlı gelişmiş yöntemler kullanılmıştır. Akustik model eğitim verisi olarak 184 saatlik yayımlanmış Türkçe haber programları veritabanı [1] ve Bölüm III-A’de detayları verilen 12.5 saatlik

hukuk ders anlatım veritabanı kullanılmıştır. Akustik model açık kaynaklı Kaldi yazılımı [11] kullanılarak eğitilmiştir. Konuşma tanıma sisteminin dil modeli eğitiminde ise 182 milyon kelimelik gazete haberleri metin veritabanı ve hukuk ders anlatımları eğitim verisinin elle yazılandırılmış referans metinleri (75,000 kelime) kullanılmıştır. Bu metin verileri ile en büyük olasılık kestirimi ile hesaplanan 3’lü dizilimli dil modelleri SRILM [14] yazılımı kullanılarak eğitilmiş ve bu modeller doğrusal ara değer bulma yöntemi ile birleştirilmiştir. Bu dil modelleri hukuk verisindeki tüm kelimeleri ve haber metinlerindeki sık geçen kelimeleri içeren 200 bin kelimelik dağarcık kullanılarak eğitilmiştir. Kullanılan konuşma tanıma sistemi ile ilgili daha fazla detay Kaynakça [2]’de verilmiştir. Bu konuşma tanıma sisteminin test verisi üzerindeki başarımları %15.05 kelime hata oranı (KHO) olarak hesaplanmıştır.

### C. Soru Cevaplama Sistemi

Soru cevap sistemi PyTorch [9] kütüphanesi kullanılarak gerçekleştirilmiştir. Modelin kodu geliştirilirken açık kaynak kodlu bir projeden<sup>2</sup> yararlanılmış ve gerekli uyarlamalar yapılmıştır. Modelin saklı vektör boyutu 150, kelime vektör boyutu 200 olarak seçilmiştir. Karakter vektörleri ve paragraf-soru arasındaki kelime bazlı tam eşleme vektörü ek öznitelikler olarak eğitim sırasında kullanılmıştır. Eğitim sırasında kullanılan eniyileme Adam algoritmasıdır. Eğitim sırasında 32 örneklili gruplar kullanılmış ve tüm veri üzerinden toplan 50 defa geçilmiştir.

Soru cevaplama sisteminin başarımları ilk olarak sınama verisinin referans metinleri üzerinde denenmiştir. Sisteme sunulan sorunun doğru cevabı ile sistemin cevap olarak döndüğü kelime diziliminin örtüşmesi, soru cevaplama sisteminin başarımları, F1 skoru kullanılarak ölçülmüştür. Sonuçlar Tablo I’da verilmiştir. Referans metinler üzerinde F1 skoru 51.66 olarak bulunmuştur.

Soru cevaplama sisteminin başarımları sözlü ders anlatımları üzerinde de sınanmıştır. Bunun için sınama verisindeki ders anlatımlarının otomatik konuşma tanıma ile yazılandırılmasından üretilen ve %15.05 kelime hata oranı veren metinler kullanılmıştır. Sistemin başarımlarının bu metinler üzerinde ölçülebilmesi için bu metinler için de soru-cevap ikililerine

<sup>1</sup>Bu veri TÜBİTAK BİDEB 2232 projesi (Proje No: 115C037) kapsamında hazırlanmıştır.

<sup>2</sup><https://github.com/laddie132/Match-LSTM>

Tablo I: Soru Cevaplama Sistemi Başarım Tablosu

Sınama verisi	F1 skoru	KHO (%)
Referans Metinler	51.66	0
Konuşma Tanıma Metinleri	48.01	15.05

ihtiyaç vardır. Bu ikilileri elle üretmek bu metinlerin yeniden işaretlenmesini gerektirmektedir ve bu da zahmetli bir iştir. Konuşma tanıma çıktısındaki soru-cevap ikililerini otomatik olarak bulmak için sınama verisinin referans metinleri ile konuşma tanıma ile elde edilen metinler hizalanmış ve hizalama sonucunda konuşma tanıma metinleri için soru-cevap ikilileri yaratılmıştır. Böylece bu metinler üzerinde F1 skoru 48.01 olarak bulunmuştur. Konuşma tanıma sisteminin hatalı metinler ürettiği düşünüldüğünde soru cevaplama sisteminin referans metinlere göre daha düşük başarımlar vermesi beklenen bir durumdur.

Bu çalışmada geliştirilen soru cevaplama sisteminin referans metinler ve sözlü ders anlatımları üzerindeki başarımları umut vericidir. Fakat Tablo I'deki deneylerde kullanılan sınama verisinde hem referans hem de konuşma tanıma metinlerinde paragraf sınırlarının bilindiği ve her paragrafta ilgili sorunun sadece o paragrafın içinde arandığı varsayımı yapılmıştır. Bu varsayım kısa paragraflar içinde cevabı bulmayı hedefleyen makine anlaması sistemleri için uygundur. Örneğin metin üzerinde soru cevaplama sistemlerinin geliştirilmesine yönelik oluşturulan SQuAD veri seti [13] ortalama 120 kelimelik paragraflardan ve ortalama 3 kelimelik cevaplardan oluşmaktadır. Geliştirilen en güncel soru cevaplama sistemleri [4], [20] SQuAD verisi üzerinde neredeyse insan başarımlarına ulaşmıştır. Sözlü ders anlatımları için geliştirdiğimiz soru cevaplama sisteminde ise bu varsayım sistemin gerçekleşmesi durumunda gerçekçi bir senaryo oluşturamamaktadır. Gerçekçi bir senaryo olarak her bir ders anlatım videosunun konuşma tanıma metinleri uzun bir paragraf olarak düşünülüp; o video ile ilgili tüm soruların cevapları bu paragraf içinde aranabilir. Bu durumda sınama verisindeki ortalama paragraf uzunluğu 917 kelimedir ve sistemin başarımları, F1 skoru, sınama verisi üzerinde 12.08'e düşmüştür.

#### IV. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada Türkçe ders anlatım videoları için bir soru cevaplama sistemi geliştirilmiştir. Sistemin başarımları hem sınama verisinin referans metinleri, hem de konuşma tanıma ile otomatik yazılandırılan metinler üzerinde denenmiştir. Geliştirilen sistem paragraf sınırlarının ve her paragraf-soru eşleşmesinin bilindiği durumlarda umut verici sonuçlar verse de gerçekleşen bir sistemde bu bilgilere erişim olmayacaktır. Bu yüzden Kaynakça [21]'dekine benzer bir yaklaşımla paragraflar yerine yapay bölütler ve sorunun hangi bölütle ilgili olduğuna karar veren bir yapay sinir ağı kullanılabilir.

#### TEŞEKKÜR

Hazırladıkları ders anlatım videolarının araştırma amaçlı kullanımına izin verdikleri için MEF Üniversitesi Hukuk Fakültesi öğretim üyelerine teşekkür ederiz. Bu bildiride yapılan çalışmalar TÜBİTAK ARDEB 3501 Programı (Proje No: 117E202) tarafından desteklenmektedir.

#### KAYNAKLAR

- [1] E. Arısoy, D. Can, S. Parlak, H. Sak, and M. Saraçlar. Turkish Broadcast News Transcription and Retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(5):874–883, 2009.
- [2] Ebru Arısoy. Developing an automatic transcription and retrieval system for spoken lectures in Turkish. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, 2017.
- [3] Caner Derici, Yiğit Aydın, Çiğdem Yenialaca, Nihal Yağmur Aydın, Günizi Kartal, Arzucan Özgür, and Tunga Güngör. A closed-domain question answering framework using reliable resources to assist students. *Natural Language Engineering*, 24(5):725–762, 2018.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [5] James Glass, Timothy J Hazen, Scott Cyphers, Igor Malioutov, David Huynh, and Regina Barzilay. Recent progress in the MIT spoken lecture processing project. In *Proceedings of Interspeech*, 2007.
- [6] James R Glass, Timothy J Hazen, D Scott Cyphers, Ken Schutte, and Alex Park. The MIT spoken lecture processing project. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 28–29, 2005.
- [7] Lori Lamel, Gilles Adda, Eric Bilinski, and Jean-Luc Gauvain. Transcribing lectures and seminars. In *Proceedings of Interspeech*, pages 1657–1660, 2005.
- [8] Chia-Hsuan Lee, Shang-Ming Wang, Huan-Cheng Chang, and Hung-Yi Lee. Odsqa: Open-domain spoken question answering dataset, 2018.
- [9] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.
- [10] Thomas Pellegrini, Helena Moniz, Fernando Batista, Isabel Trancoso, and Ramon Astudillo. Extension of the LECTRA corpus: classroom lecture transcriptions in European Portuguese. *Speech and Corpora*, 2012.
- [11] Daniel Povey et al. The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, December 2011.
- [12] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for squad. In *Proceedings of the Association for Computational Linguistics*, 2018.
- [13] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2016.
- [14] Andreas Stolcke. SRILM—An extensible language modeling toolkit. In *Proceedings of ICSLP*, volume 2, pages 901–904, Denver, 2002.
- [15] Isabel Trancoso, Rui Martins, Helena Moniz, Ana Isabel Mata, and M Céu Viana. The LECTRA corpus—classroom lecture transcriptions in European Portuguese. In *Proceedings of International Conference on Language Resources and Evaluation*, 2008.
- [16] Bo-Hsiang Tseng, Sheng-syun Shen, Hung-yi Lee, and Lin-Shan Lee. Towards machine comprehension of spoken content: Initial TOEFL listening comprehension test by machine. *CoRR*, abs/1608.06378, 2016.
- [17] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2692–2700. Curran Associates, Inc., 2015.
- [18] Shuohang Wang and Jing Jiang. Machine comprehension using match-1stm and answer pointer. *CoRR*, abs/1608.07905, 2016.
- [19] Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. Gated self-matching networks for reading comprehension and question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 189–198, 2017.
- [20] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V. Le. Qanet: Combining local convolution with global self-attention for reading comprehension. *CoRR*, abs/1804.09541, 2018.
- [21] Merve Ünlü, Ebru Arısoy, and Murat Saraçlar. Question answering for spoken lecture processing. In *Proceedings of ICASSP*, 2019.