Metinden Konuşma Sentezinde Yeni Bir Geliştirme Çerçevesi Yaklaşımı

A Novel Approach Improvement Framework for Text to Speech Synthesis

Mehmet Ali KUTLUGÜN
Bilgisayar Bilimleri ve Mühendisliği
İstanbul Sabahattin Zaim Üniversitesi
İstanbul, Türkiye
mehmet.kutlugun@std.izu.edu.tr

Yahya ŞİRİN
Bilgisayar Bilimleri ve Mühendisliği
İstanbul Sabahattin Zaim Üniversitesi
İstanbul, Türkiye
yahya.sirin@izu.edu.tr

***Özetçe*—Metinden konuşma sentezleme uygulamaları çoğunlukla çoklu ortam araçlarında kullanıcı ile olan etkileşimin üst düzeylere çıkarılması amacıyla kullanılmaktadır. Bu uygulamalar genellikle yapay (robotik) sesler üretmektedirler. Bu çalışmada metinsel ifadelerin monoton, tek bir ses biçimi şeklinde seslendirilmesi yerine, türlere ayrılarak her türün farklı ses biçimleri şeklinde seslendirilmesi hedeflenmiştir. Kısaca, metinden konuşma sentezleme süreci bir metin sınıflandırma problemi olarak ele alınmıştır. Bu sınıflandırma sürecini gerçekleştirmek için makine öğrenmesi algoritmalarından yararlanılmıştır. Sınıflandırma sonucunda, doğru sınıflandırılan dokümanların ses dosyaları başlangıçta varsayılan olarak belirlenmiş biçimlerde elde edilirken, yanlış sınıflandırılan dokümanlar için kendi kategorisi dışında farklı ses biçimleri elde edilmiştir.**

Anahtar Kelimeler—metinden konuşma sentezleme; makine öğrenmesi; sınıflandırma; ses işleme; yapay zeka; veri madenciliği

***Abstract*— Text to speech applications are mostly used to extract interaction with the user in high-level multimedia tools. These applications usually produce artificial (robotic) sounds. In this study, instead of synthesizing textual expressions as monotone, single sound form, it is aimed to be separated into species and sounded as different sound forms. Briefly, the process of speech synthesis from the text is considered as a text classification problem. Machine learning algorithms have been used to perform this sorting process. As a result of the classification, sound files of correctly classified documents are obtained in the formats initially set as default, and different sound formats are obtained for misclassified documents except for their own category.**

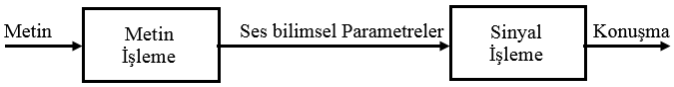
Keywords—text to speech synthesis; speech processing; machine learning; classification; artificial intelligence; data mining

I. GİRİŞ

A. Metinden Konuşma Sentezleme

Metinden Konuşma Sentezleme (MKS), yazılı bir metnin elektronik ortamda ses sinyallerine dönüştürülme işlemidir. Bu metin kaynağı, bir metin belgesi veya elektronik kitap olabileceği gibi bir web sayfası da olabilir. İdeal bir MKS sisteminin amacı insanın okuyabildiği her metni doğal insan sesi kalitesinde işleyebilmesidir. Başarımı yüksek bir MKS sistemi sayıları okuyabilmeli, kısaltmaları uygun formatta seslendirebilmeli ve bir kelimenin farklı yazımlarını ayırt edebilmelidir[1].

Böylece insan doğasına en uygun seslendirmenin yapılabileceği söylenebilir. MKS sistemleri genelde iki ana bölümden meydana gelmektedir. Bunlar **metin işleme** ve **sinyal işleme** olarak isimlendirilir.



Şekil. 1.MKS genel blok gösterimi

Metin işleme bölümü ile sentezlenecek dilin yapısına uygun bazı ön işlemler vasıtasıyla metin hakkında detaylı bilgiler elde edilir. Sinyal işleme bölümünde ise ses bilgisel parametreler kullanarak konuşma elde edilmeye çalışılır[1]. Özetle MKS sistemi yazı biçimindeki veriyi (text) girdi olarak alır ve konuşma diline uygun çıktı üretir. Bu işlemi, daha önceden hazırlanmış kayıtlı insan sesi parçalarını birleştirerek veya sinyal işleme tekniklerini kullanarak gerçekleştirmektedir.

Sinyal işleme teknikleri ile oluşturulan bir sistemde, ses sentezleyiciler genellikle yapay (robotik) sesler üretirler. Bu mekanik veya robotik ses insan sesinden kolaylıkla ayırt edilebilir. Bazı şartlar altında robotik ses tercih edilebilir fakat çoğu zaman sentezleyiciden gelen sesin daha kolay anlaşılabilir ve dinlenebilir olabilmesi için insan sesine benzemesi tercih edilir[2].

II. ÖNCEKİ ÇALIŞMALAR

Bu alandaki ilk İlk bilgisayar temelli ses sentezleme sistemi ise 1950'lerin sonunda üretilmiştir. 1968 yılında Japonya'da Noriko Umeda ve arkadaşları tarafından ilk İngilizce metinden ses sentezleme sistemi geliştirilmiştir. Bu sistemde sentezlenen ses, şimdiki sistemlerin kalitesinde olmasa da anlaşılabilir biçimde üretilebilmiştir[3]. 1960 sonrasında ise bilgisayar teknolojisinin kullanımı ile metinden konuşma sentezleme çalışmaları büyük bir ivme kazanmıştır. İlk yıllarda, başta

978-1-5386-1501-0/18/$31.00 ©2018 IEEE

İngilizce olmak üzere genelde Hint-Avrupa dil ailesindeki diller üzerinde yoğunlaşan çalışmalar, zamanla diğer dil ailelerine de uygulanmıştır[4]. Türkçe de dâhil birçok dil için hazırlanmış MKS sistemleri ticari olarak son kullanıcıya sunulmaktadır. Sunulan ticari sistemler dışında Türkçe Metinden Konuşma Sentezleme (TMKS) alanında akademik çalışmaların da yapıldığı görülmektedir[1,3-9]. MKS sistemleri teknik olarak kural tabanlı, söyleyiş ve eklemeli sentezleyiciler olmak üzere 3 farklı yöntem ile sınıflandırılır[6]. TMKS alanında yapılan akademik çalışmalar incelendiğinde oluşturulan birçok sistemde eklemeli sentezleme yöntemi kullanılmaktadır. Bu çalışmalarda sinyal işleme yönteminin ve kullanılan ses parçalarının senteze olan katkısının yanı sıra, frekans değişimi ve süre modellemesi de incelenmiştir[5].

Türkçe dili eklemeli bir dil olduğu için kelimeler hecelerin birleşmesiyle oluşmaktadır. Ayrıca bir kelimenin ek alarak farklı kelimelere türetilebildiği görülebilmektedir. Aşlıyan ve arkadaşları bu yüzden TMKS sistemleri için en uygun yöntemin 'Eklemeli Sentezleme' olduğunu belirtilmektedir. Yaptıkları çalışmada en küçük ses birimi olarak Türkçe dilinin doğal yapısı gereği heceleri kullanmışlar, TASA algoritması yardımıyla Türkçedeki farklı heceleri tespit edip kaydetmişlerdir. Bundan yola çıkarak önişlem süreci sonrası hece-ses veri tabanı oluşturmuşlar, bu ses veri tabanı kullanılarak vurgu ve tonlama özellikleri bakımından zayıf olsa da art arda bağlama yöntemi ile Türkçe metin sentezleme işlemi gerçekleştirmişlerdir[7].

Şayli, Türkçe MKS sistemleri için süre tabanlı bir model üzerinde çalışmış, fonem ve trifon tabanlı incelemelerin sonucunda ortalama süreleri baz almıştır. Bu çalışmadaki önemli sonuçlardan birisi; cümle içinde kullanıldıklarında fonem ve trifon ortalama sürelerinin belirli oranlarda düşmesidir. Bunun sebebi, daha uzun bir konuşmanın tek nefeste söylenebilmesi için, tüm birimlerin belirli oranlarda sıkıştırılmasıdır[8].

Öztürk, fonemler için süre tabanlı ve temel frekans eğrilerini esas alıp istatistiksel olarak fonemin türü, hecelerin sayısı, konumu ve vurgulu olup olmaması gibi özellikleri inceleyerek analiz yapmıştır. Bunun sonucunda ortalama sürenin en etkili parametre olduğunu rapor etmiştir. Öztürk, fonemler için süre ve F0: temel frekans eğrilerinin modellenmesini ele almıştır. İstatistiksel olarak metinsel özellikler (fonem türü, hece sayısı, hecenin konumu, hecenin vurgu alıp almaması vb.) incelenmiş ve regresyon analizi yapılmıştır. Çalışmasının sonunda bu modellerin duyumsal olarak değerlendirilmesini önermektedir[9].

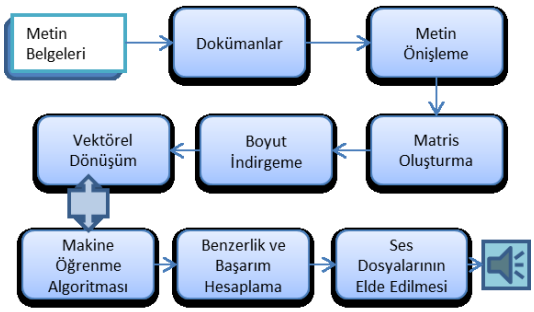
III. SUNULAN ÇALIŞMA

A. Amaç

Düz bir metinden türlerine göre ses biçimleri elde etmek için metnin doğru metotlarla işlenmesine ve yorumlanmasına ihtiyaç vardır. Bu çalışmada monoton, robotik bir metin seslendirme yerine farklı metin türlerinin kendi kategorilerine göre farklı ses biçimleri şeklinde seslendirilmesi amaçlanmıştır.

Verinin incelenip, içerisinden işe yarayan bilginin çıkarılmasına veri madenciliği (data mining) adı verilmektedir. Özellikle doğal dil tanıma probleminde en sık kullanılan makine öğrenmesi tekniği ise sınıflandırmadır ve metin sınıflandırma bu alanda önemli bir yere sahiptir[10]. MKS için makine öğrenme metotları yardımıyla metinler çeşitli aşamalardan geçirilerek türlerine göre sınıflara ayrılmış ve her sınıfın birbirinden farklı ses biçimleri şeklinde seslendirilebilmesi hedeflenmiştir. Kısaca, metinden konuşma sentezleme süreci bir metin sınıflandırma problemi olarak ele alınmıştır. Örneğin; haber içerikli metinler belirli bir formatta seslendirilirken, spor içerikli metinler daha farklı bir formatta seslendirilmek istenmiştir.

Makine öğrenmesi yaklaşımlarında sistemin metni anlaması, belirsiz durumların belirsizliklerini gidererek eğitilmesi esasına dayanır. Bu yaklaşımlarda, öğrenilen durumlardan birine daha önce görülmemiş örnekler verilerek sınıflandırma yapılması istenir. Bu yaklaşımlar, eğitim materyalinin türüne, ne kadar materyale ihtiyaç duyulduğuna, kullanılan dil bilgisinin çeşidine ve üretilmek istenen çıktıya göre değişiklikler gösterir[11]. Yapılan çalışmaya ait süreçler Şekil.2'de gösterilmiştir.



Şekil. 2.Süreç blok diyagramı

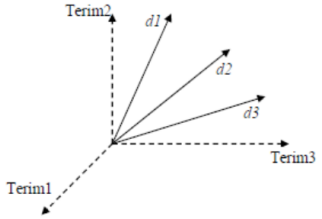
IV. DENEYSEL ÇALIŞMA

A. Veri Seti

Veri seti akademik dil ile yazılmış olan makale örneklerinden oluşmaktadır. Belli dilbilgisi kurallarına uygun olmadan yazılan metinler bu çalışma kapsamına dâhil edilmemiştir. Akışkanlar dinamiği, bilimsel endeksler ve tıbbi konulardaki 3 farklı türe ait toplamda 3891 adet makale içeren, literatürde "classic3" veri seti olarak adlandırılan veri kümesi kullanılmıştır[12].

B. Yöntem

Bir metin dokümanını makine öğrenmesi metotları kullanarak sınıflandırmak veya kümelemek için öncelikle dokümanın hazırlık işleminden geçmesi gerekir. İşleme hazırlama dokümanı makine öğrenmesi teknikleri için uygun bir duruma getirme sürecidir[12]. Hazırlama işlemi için terim sayma modeli (term count model) veya vektör uzayı-modeli (vector-space model) kullanılmaktadır[13].



Şekil. 3.Dokümanların vektörel gösterimi

Bu çalışmada vektör uzayı modeli kullanılmış olup bu modelde, her doküman vektör ‘d’ ile gösterilmiştir. Vektör ‘d’deki her boyut dokümanların terim uzayında farklı bir terimini belirtir.

C. Önişleme

Metin önişleme süreci, dokümanların makine öğrenmesi teknikleri için elverişli duruma getirilmesi işlemidir. Önişleme, aşağıdaki işlemler ile gerçekleştirilmektedir.

1) Veri Temizleme

Zamirler, edatlar ve bağlaçlar metinlerin birbirleriyle karşılaştırılmasında ayrıştırıcı özelliğe sahip olmadıklarından bunların belirlenip temizlenmesi gerekmektedir. Ayrıca tüm kelimeler küçük harf olarak tek satırlar haline dönüştürülmüş, noktalama ve rakamsal veriler ilgili dokümanlardan çıkartılmıştır. Bu işlemler için “stop-words removal” [14] algoritması kullanılmıştır.

2) Kelime Köklerinin Tespiti

Metinler içinde geçen kelimeler aynı anlam içermesine rağmen cümle içerisinde farklı ekler alabildiklerinden aynı kökten gelen kelimelerin tespit edilmesi gerekmektedir. Bu işlem için “Porter-Stemming” [15] algoritması kullanılmıştır. Bu işlem sonunda öznitelikler elde edilmiştir.

3) Doküman-Terim Matrisinin Oluşturulması

Bu aşamada tüm dokümanların isimleri satırlarda, tüm kelimeler de sütunlarda olacak şekilde denklem (1) ‘de ifade edildiği gibi bir matris oluşturulmuştur.

$$\begin{array}{c|cccc} & T_1 & T_2 & \cdots & T_t \\ D_1 & d_{11} & d_{12} & \cdots & d_{1t} \\ D_2 & d_{21} & d_{22} & \cdots & d_{2t} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ D_n & d_{n1} & d_{n2} & \cdots & d_{nt} \end{array}$$

Şekil. 4.Doküman-Terim matrisi

$$d = (w_1, w_2, \ldots, w_{|N|}) \quad (1)$$

w : i.nci öznitelik (terim),
N : Toplam öznitelik (terim) sayısı.

TABLO I. DOKÜMAN-TERİM MATRİS ÖRNEĞİ

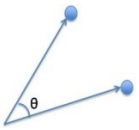
	Edit	dewey	decim	classif	…
cisi.000001	4	3	2	1	…
cisi.000002	1	0	0	0	…
cisi.000003	0	0	1	0	…
cisi.000004	0	0	0	0	…
…	…	…	…	…	…
Toplam	55	17	26	228	…

Metin sınıflandırmada farklı yöntemler ile karşılaştırıldığında hesaplama karmaşıklığı bakımından en sade ve en etkili yöntemin doküman frekansı yöntemi olduğunu belirtilmiştir[16]. Bu sebeple, kelimelerin hangi dokümanda kaç defa geçtiği hesaplanıp bu matriste gösterilmiş ve son satırında her bir kelimenin tüm doküman kümesinde toplam kaç defa geçtiği hesaplanmıştır. Bu aşama sonunda 3892 satır ve yaklaşık 29000 sütundan (öznitelik) oluşan bir matris elde edilmiştir [17]. Aşağıdaki tabloda bunun kısa bir örnek gösterimi verilmiştir.

4) Benzerliklerin Tespit Edilmesi

Bu aşamada ayrı ayrı vektörel olarak ifade edilen dokümanlar (satırlar), kümeleme işlemi ile eğitim ve test kümesi şeklinde gruplanmıştır. Tüm doküman kümesinin yüzde yirmilik bölümü eğitim kümesi, geri kalanlar test kümesi olacak şekilde seçilmiştir. Eğitim kümesinin her elemanı test kümesinin her elemanı ile hesaplamaya girerek benzerlik oranı elde edilmiştir. K-En yakın komşu algoritmasına göre benzerlik oranının hesaplanması için denklem (2) ve (3) ‘deki “Kosinüs Benzerliği” formüllerinden yararlanılmıştır [18].

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

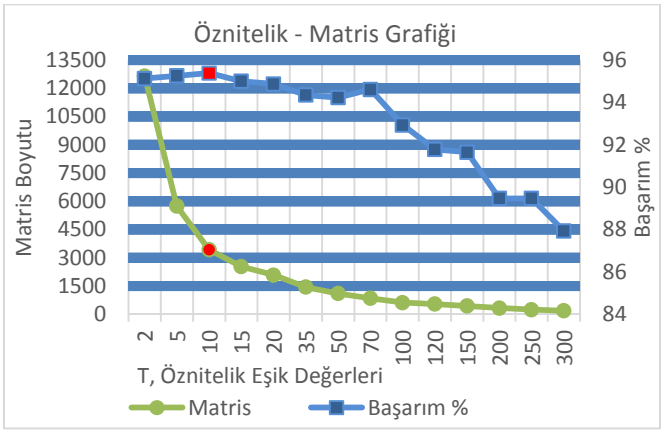


Şekil. 5.Kosinüs benzerliği formülü

Dokümanlar arasında benzerlik ölçümünü belirlenmesinde Θ değerine göre benzerlik oranı hesaplanır. Θ değerinin sıfır çıkması dokümanlar arasında herhangi bir ilişkinin olmadığı, 1’e yakın çıkmasında ise en fazla benzerlik oranına sahip olduğu anlamına gelir[19].

5) Boyut Azaltma

Boyut azaltma başarımı etkileyen önemli bir adımdır. Çünkü bu aşamada tüm dokümanlar içerisinde ayırt edici özelliği bulunmayan kelimeler tespit edilerek ilgili matristen çıkarılır ve sütun sayısının boyutu azaltılır[2,19].



Şekil. 6.Boyut azaltmanın başarıma etkisi

Yapılan uygulamada Şekil. 6’da görüldüğü gibi tüm doküman kümesinde toplamda geçen kelime sayısı T ile ifade edilecek şekilde eşik değerleri belirlenmiştir. Bu eşik değerinin altında kalan kelimeler matristen çıkarılarak matrisin boyutunda azaltmaya gidilmiştir. Uygulamada 10 eşik değerinde en yüksek başarıma ulaşıldığı görülmüş, boyut daha da azaltıldığında başarım da azalmıştır. Bu aşama sonunda

3891 satır (doküman) ve yaklaşık 3400 sütun (kelime) içeren boyutu indirgenmiş yeni bir matris elde edilmiştir.

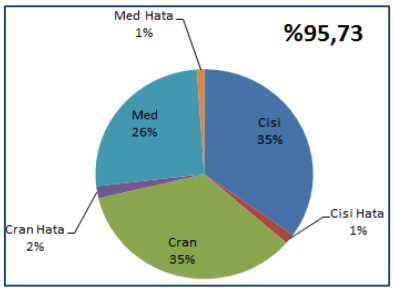
D. Başarım Oranının Hesaplanması

Performans değerlendirme işlemi için karmaşıklık matrisinden (confusion matrix) yararlanılmış, yöntemin başarımı F-ölçütü (*F-measure*) ile değerlendirilmiştir. F-ölçütü kullanılan algoritmaların başarımlarının belirlenmesi ve karşılaştırılmasında güvenilir bir ölçüt olarak literatürde oldukça yoğun olarak kullanılmaktadır[20].

TABLO II. KARMAŞIKLIK MATRİSİ

	Cisi	Cran	Med
Cisi	1398	24	38
Cran	39	1326	35
Med	16	28	989

Ayrıca farklı k değerleri için de k-en yakın komşu algoritması uygulanmış, bu veri seti için başarım oranı yüzde 95,373 olarak k=1 değerinde elde edilmiştir.



Şekil. 7.Başarım oranı

1) Ses Dosyalarının Elde Edilmesi

Ses dosyaları, "FreeTTS" kütüphaneleri ile "MBROLA" ses veri tabanı kullanılarak elde edilmiştir[21]. Başlangıç değerleri için 3 türe ait Tablo 3'e benzer değerler tanımlanmıştır. Bu değerlerden elde edilen ses dosyaları "wav" ses biçiminde oluşturulmuştur.

TABLO III. CISI TÜRÜNDEKİ DOKÜMANLAR İÇİN;

Ses Dosyası	:	mbrola_us1 - 16kHz
Cinsiyet	:	Bayan Sesi
Okuma Hız Değeri	:	1.0f
Perde Frekans Değeri	:	180f
Perde Frekans Aralığı	:	22.0f
Ses Yoğunluğu	:	1.0f

V. SONUÇ VE ÖNERİLER

Bu çalışmada, metinden konuşma sentezlemeye yeni bir yaklaşım getirilerek, farklı metin türlerinin farklı ses biçimleri şekilde seslendirilmesi önerilmiştir. Bu sayede, örneğin bir web sitesi gibi birçok metin türünün bir arada bulunduğu ortamlarda kategorilere ayrılmış farklı seslendirmeler yapılarak konular daha belirgin hale getirilebilir. Belirli standartlar ile özellikle görme engelliler için, seslendirme biçimine göre hangi konudan bahsedildiği daha anlamlı hale getirilebilir.

KAYNAKLAR

[1] Sel, İ., "*Türkçe Metinler İçin Hece Tabanlı Metinden Konuşma Sesntezleme Sistemi Yüksek Lisans Tezi",* Fırat Üniversitesi, Elazığ, 2013.

[2] Şirin Y. ve Kutlugün M.A., "Konuşma Sentezinde Artan Doğallık için Boyut Azaltma Seçimi", 25th Signal Processing and Communications Applications Conference, Antalya, (IEEE : s. 10.1109/SIU.2017.7960573), 2017.

[3] Erdemir, C., "Türkçe Metin Seslendirme İçin Doğal Konuşma Sentezleme", Yüksek Lisans Tezi, İstanbul Üniversitesi, 2010.

[4] Yılmaz, A. E., *"Türkçe Metinden Konuşma Sentezleme Uygulamaları için bir Veri Sözlük Seti ve Yazılım Çerçevesi Önerisi"*, IEEE 17. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Antalya, 2009.

[5] Eker, B., "*Turkish Text to Speech System, Yüksek Lisans Tezi",* Bilkent Üniversitesi Mühendislik ve Fen Bilimleri Enstitüsü, Ankara, 2002.

[6] Sel İ. ve ark., "Beyin Bilgisayar Arayüzleri İçin Türkçe Metinden Konuşma Sentezleme Sistemi", Fırat Üniversitesi Elektrik-Elektronik Bilgisayar Sempozyumu, Elazığ, 2011.

[7] Aşlıyan, R. ve ark., "Türkçe Otomatik Heceleme Sistemi ve Hece İstatistikleri", Akademik Bilişim 2006 BilgiTek IV. Denizli: Pamukkale Üniversitesi, 2006.

[8] Şayli, Ö., "Duration analysis and modelling for Turkish text-to-speech synthesis", Yüksek Lisans Tezi, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü, 2002.

[9] Öztürk, Ö., "Modelling phoneme durations and fundamental frequency contours in Turkish speech", Doktora Tezi, ODTÜ Fen Bilimleri Enstitüsü, 2005.

[10] Bayrak, Ş. ve ark., "Makine Öğrenme Yöntemleriyle N-Gram Tabanlı Dil Tanıma", Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, 2012.

[11] Jurafsky, D. ve Martin, J. H., "Speech and Language Processing", Prentice Hall, 2008.

[12] Özgür A., *"Supervised and unsupervised machine learning techniques for text document categorization Yüksek Lisans Tezi"*, Boğaziçi Üniversitesi, İstanbul, 2002.

[13] Salton, G., *"Automatic Information Organization and Retrieval"*, McGraw Hill Text, ISBN:0070544859, 1968.

[14] Ronald, *"Remove Stop Words from a File",* https://javaextreme.wordpress.com, 2014.

[15] Porter, M., *"The English Porter stemming algorithm",* https://snowballstem.org, 2015.

[16] Uzun, E., "İnternet Tabanlı Bilgi Erişimi Destekli Bir Otomatik Öğrenme Sistemi", Doktora Tezi, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne, 2007.

[17] Dasan, S., *"Program code for building TDM",* https://snowballstem.org, 2010.

[18] Han, J., Kamber, M., Pei, J., "Data mining: concepts and techniques: concepts and techniques", Elsevier, 2011.

[19] Kutlugün, M.A., "Gözetimli Makine Öğrenmesi Yoluyla Türe göre Metinden Ses Sentezleme", Yüksek Lisans Tezi, İstanbul Sabahattin Zaim Üniversitesi, 2017.

[20] Üstüner, M. ve Bilgin, G., "Mitosis detection on histopathological images using statistical detection algorithms", Signal Processing and Communications Applications Conference (SIU), 2015 23th. IEEE, 2015.

[21] Walker, W., Lamer, P. ve Kwok, P., *"FreeTTS 1.2 - A speech synthesizer written entirely in the Java programming language",* http://freetts.sourceforge.net, 2005.