



Veri Kazıma Yoluyla Web Sitelerinden Elde Edilen Başlıkların Seslendirilmesine Yönelik Bir Metin Madenciliği Yaklaşımı

Yazılım Mühendisliği Ana Bilim Dalı
Tezsiz Yüksek Lisans Bitirme Projesi

Baran Şenel

Öğrenci No Y210234029

Bitirme Projesi Danışmanı: Prof. Dr. Ayşegül Alaybeyoğlu
Ocak 2023

Yazarlık Beyanı

Ben, **Baran Şenel**, başlığı **Veri Kazıma Yoluyla Web Sitelerinden Elde Edilen Başlıkların Seslendirilmesine Yönelik Bir Metin Madenciliği Yaklaşımı** olan bu bitirme projemin ve bitirme projemin içinde sunulan bilgilerin şahsıma ait olduğunu beyan ederim. Ayrıca:

- Bu çalışmanın bütünü veya esası bu üniversitede Yüksek Lisans / Doktora derecesi elde etmek üzere çalıştığım süre içinde gerçekleştirilmiştir.
- Daha önce bu bitirme projesinin herhangi bir kısmı başka bir derece veya yeterlik almak üzere bu üniversiteye veya başka bir kuruma sunulduysa bu açık biçimde ifade edilmiştir.
- Başkalarının yayımlanmış çalışmalarına başvurduğum durumlarda bu çalışmalara açık biçimde atıfta bulundum.
- Başkalarının çalışmalarından alıntıladığımda kaynağı her zaman belirttim. Bitirme projemin bu alıntılar dışında kalan kısmı tümüyle benim kendi çalışmamdır.
- Kayda değer yardım aldığım bütün kaynaklara teşekkür ettim.
- Bitirme projemde başkalarıyla birlikte gerçekleştirilen çalışmalar varsa onların katkısını ve kendi yaptıklarımı tam olarak açıkladım.

Tarih: 20.01.2023

Veri Kazıma Yoluyla Web Sitelerinden Elde Edilen Başlıkların Seseleştirilmesine Yönelik Bir Metin Madenciliği Yaklaşımı

ÖZ

Bu bitirme projesi, ses sentezleme ve veri kazıma konuları hakkında literatür taraması yapılarak bir web sunucusu üzerinde veri kazıma yöntemleri ile internet sitelerinin taranarak meta verilerin, html tagların ve içeriklerin elde edilmesi ve açık kaynak kodlu metinden ses / konuşma sentezleme kütüphaneleri kullanılarak elde edilen bilgilerin çıktı olarak üretilmesini hedeflenmektedir. Ses sentezleme işlemlerinde ingilizce modeller, açık kaynak materyal olarak daha verimli ve rahat bulunabildiği için çalışmada tercih edilecektir.

Anahtar Sözcükler: Web, veri kazıma, ses sentezleme, yapay zeka

A Text Mining Approach To Voice The Headlines Obtained From The Websites By Data Scraping

Abstract

This graduation project aims to obtain metadata, html tags and content by scanning websites with data scraping methods on a web server by scanning the literature on voice synthesis and data scraping, and outputting the information obtained from open source text using voice / speech synthesis libraries. intended to be produced. In sound synthesis processes, English models will be preferred in the study as they can be found more efficiently and comfortably as open source material.

Keywords: Web, data scraping, voice synthesis, artificial intelligence

Teşekkür

Bu bitirme projesi çalışmasına katkılarından dolayı eşime, arkadaşlarıma teşekkür ediyorum.

İçindekiler

Yazarlık Beyanı	i
Öz.	ii
Abstract	iii
Teşekkür	iv
Şekiller Listesi.....	vii
Tablolar Listesi.....	viii
Kısaltmalar Listesi.....	ix
Semboller Listesi	ix
1 ...Giriş	1
1.1 ...Bitirme Projesinin Amacı	1
1.2 ...Literatür Taraması	2
1.2.1 ...Veri Kazıma	2
1.2.2 ...Metinden Ses Sentezleme	2
1.2.2.1 ...Ses Hakkında Bilgiler	3
1.2.2.2 ...Metinden Ses Sentezleme Yöntemleri	3
2 ...Materyal ve Metot.....	10
2.1 ...Uygulama Gerçekleştirme	10
2.1.1 ...Uygulama Ana Ekranı	11
2.1.2 ...Verinin Flask API'ye iletilmesi	12
3 ...Bulgular ve Sonuç	14
Kaynaklar.....	15
Ekler.....	17

Ek A...Uygulama Gerçekleştirme Aşamasında Kullanılan Bazı Kodlamalar...	18
Ek B...Dış Cilt Kapağı	21
Özgeçmiş.....	24

Şekiller Listesi

Şekil 1.1Ses Dalgası	3
Şekil 1.2Metinden ses sentezleme aşamaları	4
Şekil 1.3Mel-spektogram.....	6
Şekil 1.4Ses sentezleme aşamaları, mimari ve modeller	8
Şekil 2.1Uygulama diagramı	11
Şekil 2.2Uygulama ana ekranı	11
Şekil 2.3Flask api request ve BeautifulSoup veri kazıma	12
Şekil 2.3Kazıma yapılan verilerin gösterildiği uygulama ekranı	12
Şekil 2.3Sentezlenen ses dosyasının otomatik olarak çalıştırılması	13

Tablolar Listesi

Tablo 3.1 ...Sentezlenen ses verisi hakkında bilgiler	14
---	----

Kısaltmalar Listesi

TTS	Text to speech
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
VITS	Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech

Semboller Listesi

S	Saniye
KB	Kilobyte

Bölüm 1

Giriş

Günümüzde yapay zeka ve makine öğrenmesi alanlarındaki gelişmelerin hızlanması ve gelişmesiyle metinden ses sentezleme çıktılarında hem performans hem de kalite açısından çeşitlilik sunan bir çok çalışma ortaya çıkmıştır.

Bu teknoloji, özellikle e-kitap okuyucuları, mobil uygulamalar ve oyunlar gibi çeşitli uygulamalar, navigasyon cihazları, eşzamanlı dil çeviri uygulamaları için kullanılmaktadır. Ayrıca, sesli asistanlar ve sesli arama gibi sesli arama teknolojileri için de kullanılmaktadır.

1.1 Bitirme Projesinin Amacı

Bu bitirme projesinin amaçları şunlardır;

- Literatür taraması yaparak, metinden ses sentezleme alanında yapılmış çalışmaları incelemek ve proje konusuna uygun bir açık kaynak kütüphane bulmak.
- Python dilinde uygun bir yazılım çatısı ile web sunucusu çalıştırmak
- Kullanıcının belirlediği internet sitesi üzerinde veri kazıma işlemini başarı ile gerçekleştirmek.
- Veri kazıma metodu ile kullanıcının erişmek istediği internet sitesinden önemli başlıkları kazımak.
- Kazınan verinin, kullanıcı isteği üzerine metinden ses sentezleme kütüphanesine gönderildikten sonra ses dosyası çıktısı olarak web sayfası üzerinde çalıştırılması.

1.2 Literatür Taraması

1.2.1 Veri Kazıma

Veri kazıma, bir web sitesinden bilgi toplama işlemidir. Bu işlem genellikle bir program veya script tarafından gerçekleştirilir. Program, bir web sitesine bağlanır ve içeriği analiz eder, sonra istenen bilgileri çeşitli formatlarda kaydeder. Veri kazıma e-ticaret, finans, medya, sosyal medya, eğitim, sağlık gibi bir çok alanda kullanılabilir. Örneğin e-ticaret sitelerinden ürün, fiyat, stok ve yorum bilgileri toplanabilir, medya sitelerinden haber başlıkları, makaleler toplanabilir.

Python dilinde, veri kazıma işlemi için birçok kütüphane mevcuttur, ancak en yaygın olarak kullanılanları BeautifulSoup ve Scrapy'dir, Bozkuş [\[1\]](#) tez çalışmasını gerçekleştirmek için benzer bir araştırma yapmış ve çalışmasında BeautifulSoup kütüphanesini kullanmıştır.

Veri kazıma araçlarının performansları, kazıma yapılacak verinin büyüklüğü, internet kalitesi, sunucu performansları gibi bir çok etmenden etkilenebilir. Uzun [\[2\]](#) veri kazıma araçları hakkında yaptığı araştırmada bu araçların performans kriterlerini göz ardı ettiğini tespit etmiştir.

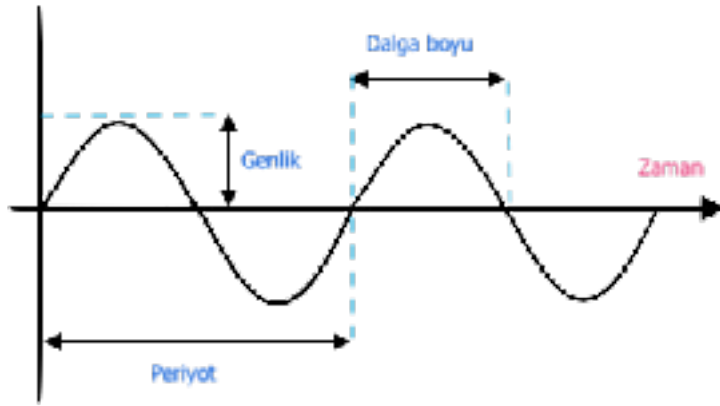
Bu projede internet sitelerinde bulunan başlık metinleri kazınacağından, performans ölçütlerinden etkilenmeyeceği öngörülmüştür.

1.2.2 Metinden Ses Sentezleme

Metinden ses sentezleme, yapay zeka ve makine öğrenmesi teknolojileri kullanarak seslerin üretilmesi veya sentezlenmesidir. Bu teknoloji, metin veya yazılı veriyi ses verisine dönüştürür.

1.2.2.1 Ses Hakkında Bilgiler

Ses, canlıların işitme organlarının algıladığı, bir maddedeki moleküllerin titreşimleri ile meydana gelen periyodik basınç değişimleri olarak tanımlanır. [3] yaptığı yüksek lisans çalışmasında sesin yapısı ele alınacak olunursa önem kazanan kavramların sesin dalga formu, frekansı, genliği ve gürültü seviyesi olduğunu belirtmiştir. Sesin bazı özellikleri Şekil 1.1’de görülmektedir.



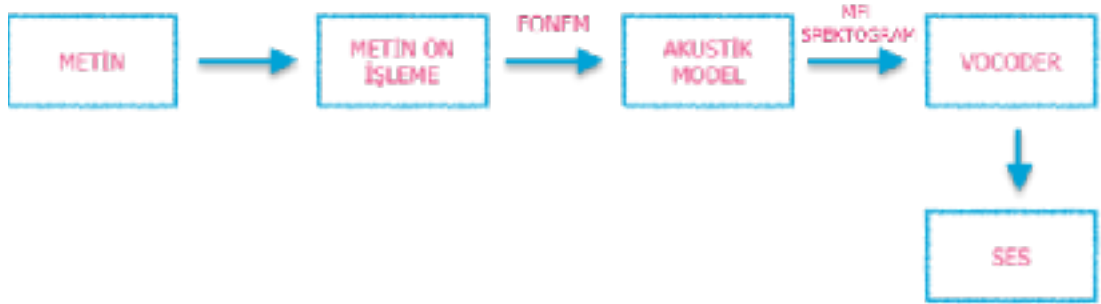
Şekil 1.1 Ses dalgası

1.2.2.2 Metinden Ses Sentezleme Yöntemleri

Ses sentezlemede bazı aşamalar bulunmaktadır bunlar;

- Metin ön işleme: Girdi olarak verilen metin, yazım veya dilbilgisi hatalarının düzeltilmesi veya temizlenmesi için ön işlemeye tabi tutulur.
- Sözcüklerin oluşturulması: Metin, sözcükler olarak katmanlara ayrılır. Bu katmanlar, sesli sözcüklerin fonolojik yapısını tanımlamak için kullanılır.
- Ses özelliklerinin modellenmesi (akustik modelleme): Sesli sözcükler, sesli özellikleri tanımlamak için kullanılan modeller tarafından ses tonu, ses hızı, frekansı gibi özellikleri analiz edilir.

- Ses verisinin üretilmesi: Ses özellikleri, ses üretim motorları tarafından kullanılarak ses verisi olarak üretilir. Bu motorlar, sözcüklerin fonolojik yapısını ve özelliklerini kullanarak ses verisini üretir.



Şekil 1.2 Metinden ses sentezleme aşamaları

Bu aşamalar arasında, modeller ve üretim motorları kullanılarak ses verisinin üretilmesi ve düzenlenmesi gerçekleştirilir. Bu modeller ve motorlar, yapay zeka ve makine öğrenmesi teknolojileri kullanılarak eğitilir.

Ses sentezlemede kullanılan ses verisi, üretilmek istenen sesin ton, hız, yükseklik ve diğer özelliklerini içermelidir. Bu veri, ses kayıtları, ses dosyaları veya metin verileri gibi farklı şekillerde elde edilebilir. Genellikle, ses sentezlemede kullanılan veri, çeşitli kaynaklardan elde edilen ve etiketlenmiş ses verileridir.

Kaliteli bir ses modeli oluşturmak için, veri miktarı ve kalitesi önemlidir. Genellikle, daha fazla ve daha çeşitli veri kullanılması, daha kaliteli bir ses sentezleme modeli oluşturulmasını sağlar. Ancak, veri miktarının çok fazla olması, eğitim ve modelleme işlemleri için çok fazla zaman ve kaynak gerektirebilir. Boyut olarak, çok fazla veriye ihtiyacın olması gerekmez, ancak çeşitli kaynaklardan ve çeşitli özelliklere sahip ses verileri kullanmak daha iyi sonuçlar verebilir.

Bu noktada bugüne kadar geliştirilmiş metinden ses sentezleme mimarileri kaynak gereksinimi, kalite, hız ve genişleyebilirlik / geliştirilebilirlik açısından farklılık göstermektedir. Günümüzde son teknoloji sentezleyicilerin büyük bir bölümü yapay sinir ağları kullanmaktadır. Metinden ses sentezleme alanında yapılan çalışmaların ve kullanılan araçların güzel bir özeti kaynak [4]'te verilmiştir. Yapay sinir ağları, metinden ses sentezleme teknolojilerinde büyük ilerlemeler sağlanmasına katkıda bulunmuştur. Bazı metinden ses sentezleme mimarilerini örnek verecek olursak;

- Artiküler ses sentezleme [4]
- Şekillendirici ses sentezleme [4]
- Eklemeli ses sentezleme [4-8]
- Parametrik ses sentezleme [4]
- Parametrik nöral ses sentezleme [4]

Eklemeli metinden ses sentezleme mimarisinde, gerçek ses kalitesine yakın sonuçlar elde edilse de, konuşmacının ses datasetine ihtiyaç duyduğu için gelişime kapalı bir yöntemdir. Bu noktadan kıyasla, yapılan son çalışmalarda parametrik sentezleme üzerine yapılan çalışmalarda artış görülmek ile birlikte, hibrit parametrik nöral ses sentezleme mimarilerinde mel-spektrogramların kullanımı yerine mel-spektrogramları temsil eden yaklaşımlar implemente edilmeye başlandığı görülmektedir [9].

Kaliteli bir ses sentezleme modeli oluşturmak için, son dönemde aşağıdaki teknik ve yöntemlerden faydalanılmaktadır [4];

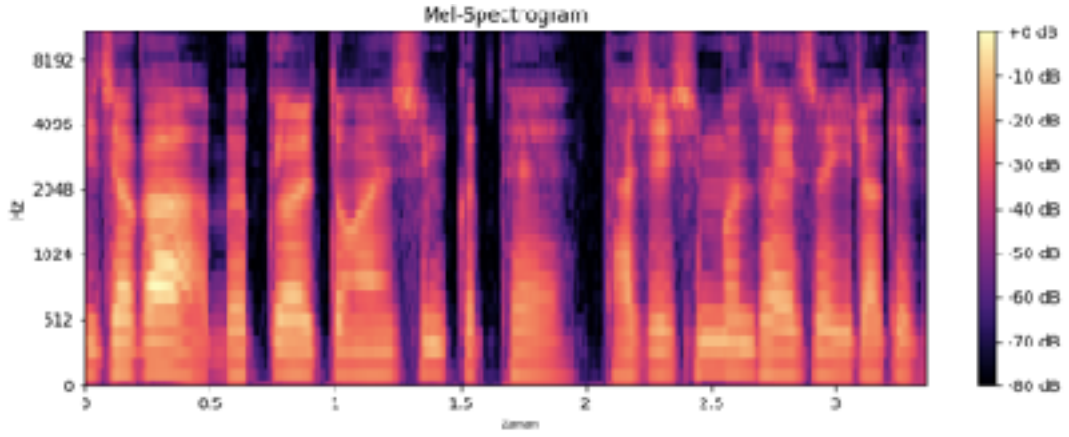
Yapay sinir ağları [10], üretken modeller (GANS, VAEs) gibi üretken modeller, fourier dönüşümü, kısa fourier dönüşümü ve mel-spektrogram gibi sinyal işleme teknikleri kullanılır.

Fourier dönüşümü, bir ses sinyali veya ses verisi için zamana göre sesin frekans bileşenlerini analiz etmek için kullanılan bir matematiksel tekniktir. Bu teknik, ses verisini frekans bileşenlerine ayırarak, sesin hangi frekanslarda daha yoğun olduğunu ve hangi frekanslarda daha az yoğun olduğunu gösterir.

Kısa Fourier dönüşümü, ses verisi için Fourier dönüşümünün kısa zaman aralıkları için uygulanmış bir versiyonudur. Bu teknik, ses verisini kısa zaman aralıkları için

frekans bileşenlerine ayırarak, sesin zaman içinde nasıl değiştiğini analiz etmek için kullanılır.

Mel-Spektrogram, ses verisi için zaman ve mel-skala bileşenlerini gösteren bir grafikdir. Mel skala, insan işitme duyusu için uyarlanmış bir frekans ölçeğidir ve ses verisi için hangi mel-frekanslarda ne zaman yoğun olduğunu gösterir. Bu grafik, sesteki mel-frekans değişimlerinin ve sesin zaman içindeki evriminin analizi için kullanılır. Bir mel-spektrogram örneği Şekil 1.3'te verilmiştir.



Şekil 1.3 Mel-spektrogram

Fonem, bir dil için sesli harflerin en küçük birimi olarak kabul edilir. Bir sesin anlamını belirleyen bu küçük birimler, ses sentezlemede sesli sözcüklerin oluşturulması ve ses verisinin üretilmesi için kullanılır.

Ses sentezlemede yakın geçmişte yapılan araştırmalar sonucu ortaya çıkan modeller arasında Concatenative Synthesis [5-8] ve Parametric Synthesis modelleri ile WaveNet, Tacotron, MelNet gibi derin öğrenme tabanlı modeller sayılabilir. Concatenative Synthesis modelleri, kayıtlı ses parçalarının birleştirilmesi ile sentezlenir, Parametric Synthesis modelleri ise matematiksel modelleme ile sentezlenir. Her modelin avantaj ve dezavantajları mevcuttur. Örneğin, WaveNet gerçekçi ses üretebilir ancak üretim hızı daha yavaş olabilir. Tacotron ise hızlı üretim yapabilir ancak gerçekçi ses üretebilme olasılığı daha düşük ve kalitesiz olabilir.

MelNet ise mel-spektrogramları kullanarak üretebilir ancak gerçek ses verileri ile karşılaştırıldığında gerçekçi ses üretebilme olasılığı daha azdır.

Metinden ses sentezleme aşamalarını tekrar ele alacak olursak bunlar sırasıyla; preprocessor, encoder, decoder ve vocoder olacaktır.

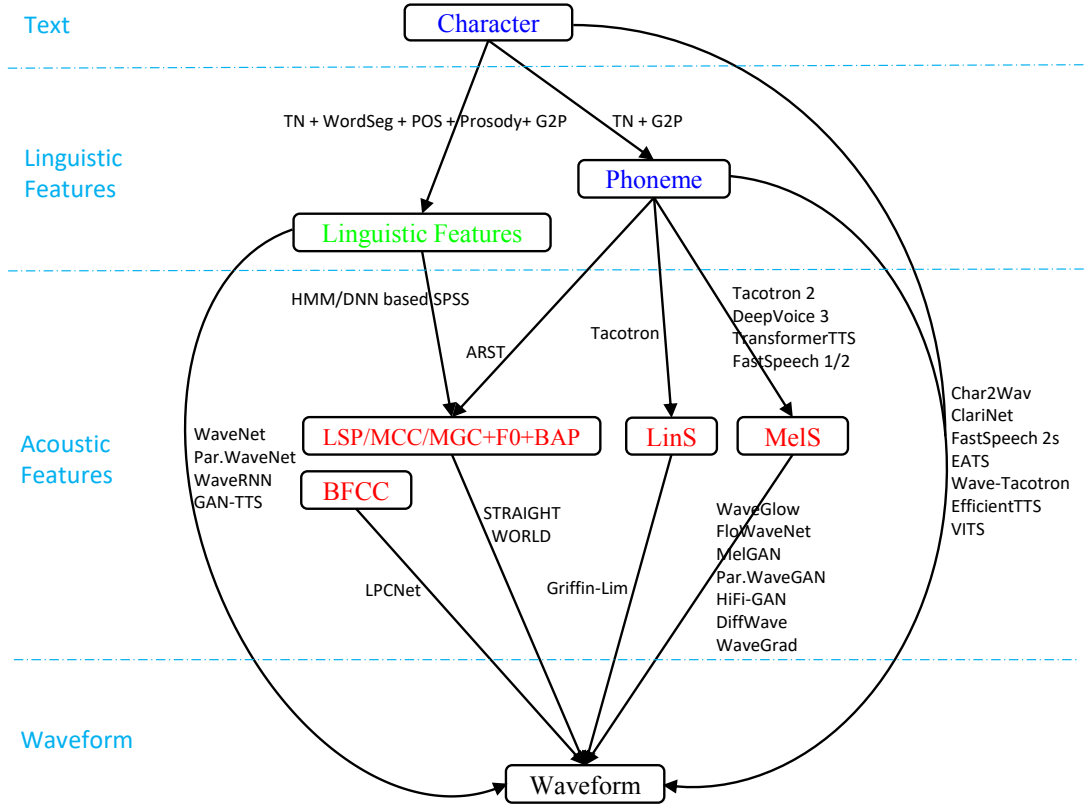
Preprocessor aşaması, metnin dili, sözcük veya fonem dizilimleri gibi özelliklerin belirlenmesi için LSTM, GRU gibi yapay sinir ağları kullanır. LSTM (Long Short-Term Memory) ve GRU (Gated Recurrent Unit) yapay sinir ağları preprocessor (ön işleme) aşamasında metin verisi için dil modelleme, sözcük veya foneme dizilimleri gibi özelliklerin belirlenmesi için kullanılırlar. Bu yapay sinir ağları, metin verisi için zaman serisi verilerinde uzun süreli bağımlılıkları tanımlayabilen ve bu verileri işleyebilen yapay sinir ağlarıdır. LSTM ve GRU yapay sinir ağları arasındaki fark, LSTM'de kullanılan "memory cell" ve "input, forget, output gates" kavramları ile GRU'da kullanılan "update ve reset gates" kavramlarıdır.

Encoder aşaması, akustik özelliklerin belirlenmesi için CNN (Convolutional Neural Network) , Transformer gibi yapay sinir ağlarını kullanır. Örneğin, mel-spektrogramları gibi sesli özellikler elde edilir.

Decoder aşaması, bazı mimarilerde metinsel olarak belirlenmiş akustik özellikleri kullanarak gerçek zamanlı ses verisi üretmek için mel-spektrogramlarını üretir. RNN (Recurrent Neural Network), LSTM gibi yapay sinir ağlarını kullanır.

Vocoder ise elde edilen tüm çıktıları kullanarak, WaveNet, WaveGlow ve benzeri yapay sinir ağları ile ses verisi üretimi yapar.

Yukarıda anlatılan bir çok detay Şekil 1.4'te verilmiştir.



Şekil 1.4 Ses sentezleme aşamaları, mimari ve modeller [4]

Bu modellerin performanslarını arttırmak için şu işlemler yapılabilir;

Data augmentation: Eğitim verilerinin çeşitlendirilmesi, modelin çeşitli ses durumlarına dayanıklı hale gelmesini sağlar. Örneğin, ses dosyalarının yavaşlatılması, hızlandırılması, yükseltilmesi gibi işlemler yapılabilir.

Transfer learning: Eğitilmiş bir modelin ağırlıklarının yeni bir model için başlangıç değerleri olarak kullanılması, modelin hızlı bir şekilde eğitilmesini ve daha iyi performans göstermesini sağlar.

Attention mechanism: Ses sentezleme sürecinde, metin verisi için sesli özelliklerin belirlenmesi için kullanılan yapay sinir ağlarında, attention mekanizması kullanılabilir. Attention mekanizması, modelin belirli sesli özellikleri daha ayrıntılı olarak incelemesini sağlar ve daha gerçekçi ses üretebilmesini sağlar.

Multi-speaker model: Bu tekniğin kullanılması ile aynı metin verisi için farklı konuşmacıların seslerinin üretilmesi sağlanır. Bu teknik, ses sentezleme sisteminin gerçekçi ve çok konuşmacılı ses üretebilmesini sağlar.

Fine-tuning: Bu teknik kullanıldığında, eğitilmiş bir model üzerinde daha spesifik bir veri setiyle daha az miktarda veri ile eğitilir. Bu teknik, modelin daha spesifik bir veri seti için daha iyi performans göstermesini sağlar ve önceden eğitilmiş bir modelin özelliklerini koruyarak, üretilen sesin gerçekliğini arttırmak için kullanılabilir.

Araştırmalar [\[10-16\]](#) bir çok yeni teknolojiyi ele almaktadır.

Metinden ses sentezleme araştırmalarındaki hidden markov model [\[15\]](#) gibi bazı teknikler çok eski yıllara dayanmaktadır.

Son yıllarda metinden ses sentezleme sistemleri için ortaya çıkmış popüler metotlar arasında farklı sinir ağı optimizasyonlarını kullanan, end-to-end tacotron2, transformer-tts, fastspeech, fastspeech2, VITS, JETS metinden ses sentezleme mimarileri gösterilebilir [\[4,16\]](#).

Son yıllarda seslendiren kişilerin belirli olduğu ön tanımlı veri setleri üzerinden oluşturulan sentezleme modellerinin yanı sıra oluşturulmuş mevcut modellerdeki konuşmacı seslerinin kısa süreli veri setleri üzerinden klonlanarak yeni ses modelleri oluşturulması üzerine yapılan çalışmalar artış göstermektedir [\[16\]](#).

Ses sentezlemede sesin kalitesini ölçmek amacıyla MOS “mean opinion score” değeri baz alınmaktadır.

Bölüm 2

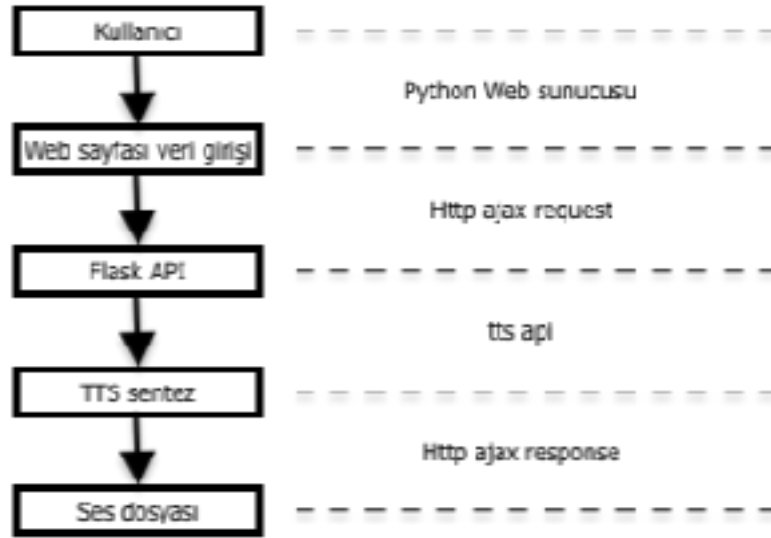
Materyal ve Metot

Bu bitirme projesinde web uygulama katmanı Python Flask mikro web çerçevesi üzerinde çalışan, kullanıcının girdi olarak verdiği internet sitesindeki başlık etiketlerini kazıyarak, metin girdilerini arka planda VITS [16] ve çeşitli metinden ses sentezleme modellerini barındıran açık kaynaklı end-to-end metinden ses sentezleme uygulaması olan coqui.ai'a ileterek ses verisine çevirdikten sonra kullanıcının dinlemesini sağlayan bir web uygulaması gerçekleştirilmektedir.

Çalışma linux işletim sisteminde gerçekleştirilmiştir.

2.1 Uygulama Gerçekleştirme

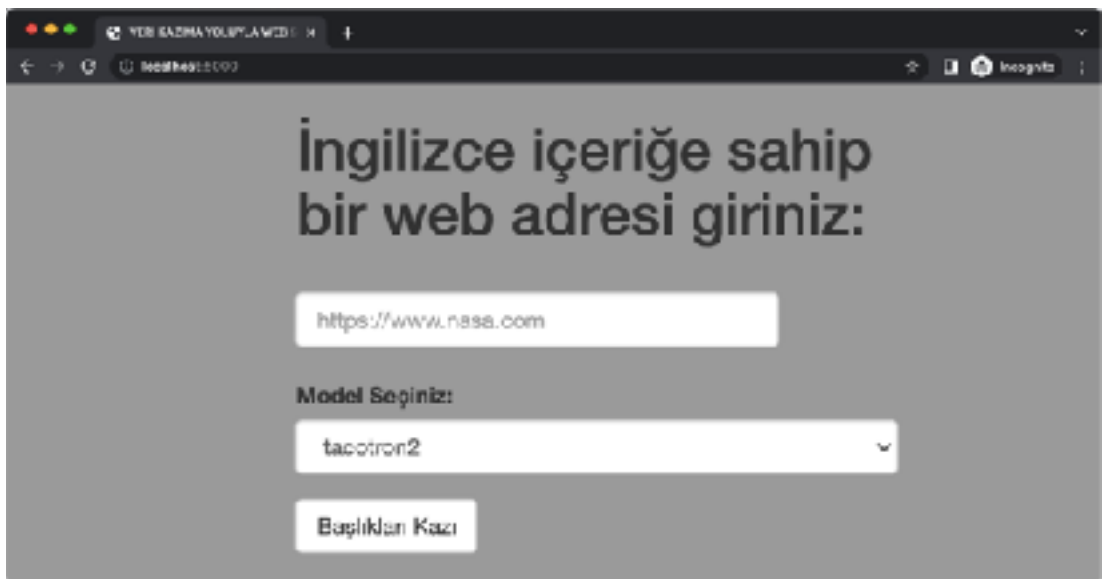
Python günümüzde çok popüler olan bir programlama dilidir ve beraberinde bir çok gelişmiş arayüz ve kütüphaneyi barındırmaktadır. Requests kütüphanesi http isteklerini yerine getirmek için kullanılmıştır. BeautifulSoup kütüphanesi elde edilen DOM ağacı üzerinde gezinerek istenilen html tag'larını ve içeriklerini döndürmektedir. Benzer bir çalışma [1] tarafından yapılmıştır. Bitirme projesinde yapılan çalışmaya ait basit bir diagram Şekil 2.1'de verilmiştir.



Şekil 2.1 Uygulama diagramı

2.1.1 Uygulama Ana Ekranı

Uygulama ilk çalıştırıldığında Şekil 2.2'deki ekran kullanıcıya gösterilmektedir. Kullanıcının ingilizce dilinde hazırlanmış bir web sayfası adresi girmesi ve metinden ses sentezleme modeli seçmesi istenmektedir.



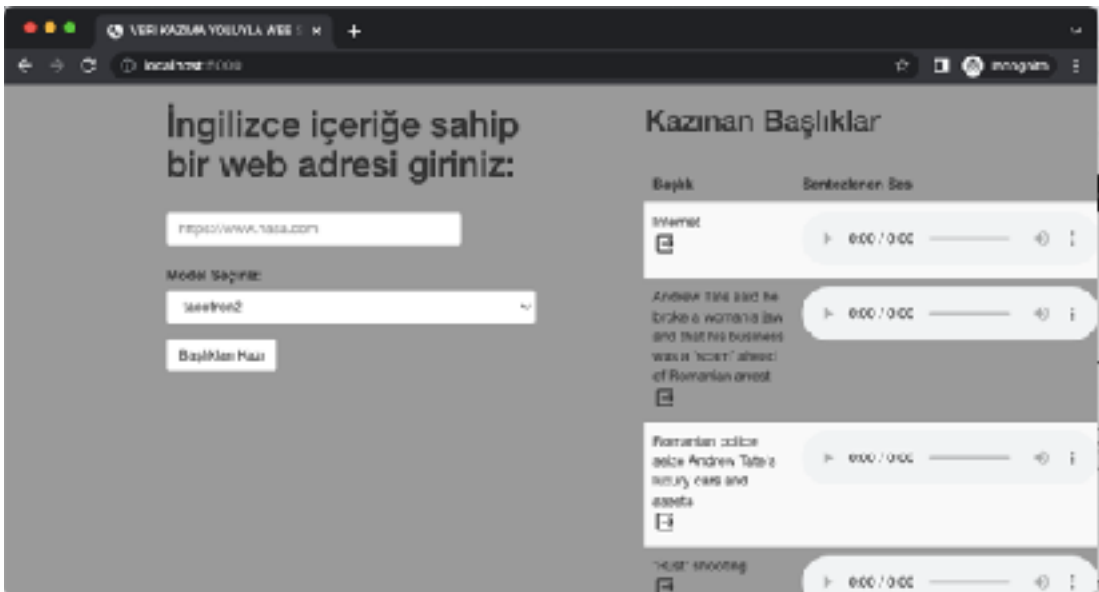
Şekil 2.2 Uygulama ana ekranı

2.1.2 Verinin Flask API'ye İletilmesi

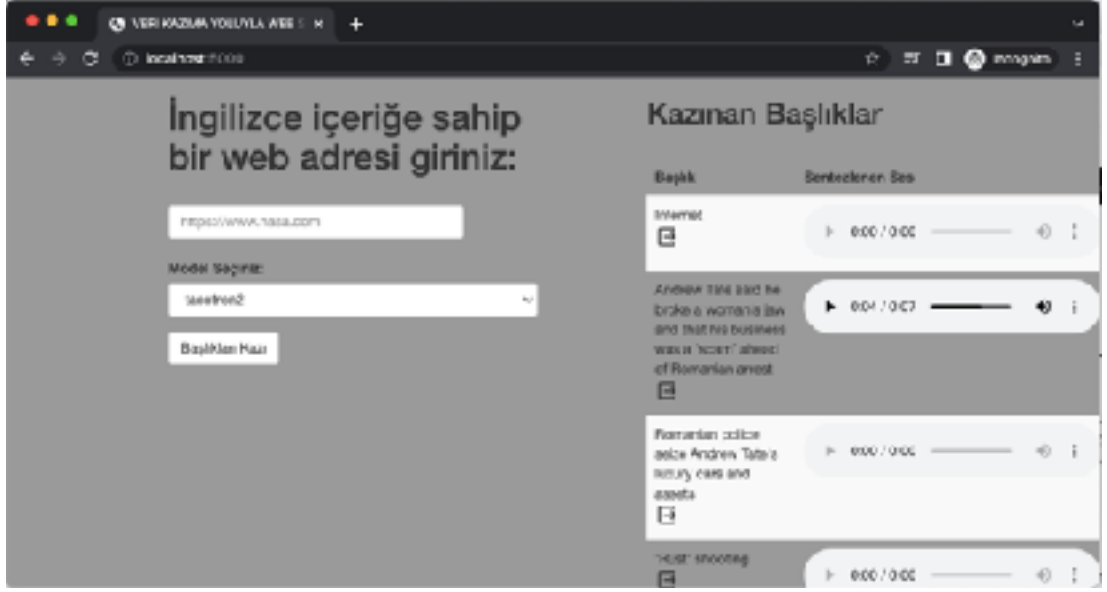
BeautifulSoup vasıtasıyla kazınan h1 - h6 başlıkları tekrar web sayfasına döndürülerek kullanıcıya listelenir. İlgili ekran görüntüleri Şekil 2.3 ve Şekil 2.4'te verilmiştir.



Şekil 2.3 Flask api request ve BeautifulSoup veri kazıma



Şekil 2.4 Kazıma yapılan verilerin gösterildiği uygulama ekranı



Şekil 2.5 Sentezlenen ses dosyasının otomatik olarak çalıştırılması

Şekil 2.5’te ajax sorgusu sonucu coqui tts tarafında sentezlenen sesin javascript kodlarıyla otomatik olarak çalıştırılarak kullanıcının dinlemesi sağlanmaktadır.

Bölüm 3

Bulgular ve Sonuç

Veri kazıma işleminin çok hızlı gerçekleştiği, ancak ses sentezleme işleminin donanıma bağımlı olduğu uygulama gerçekleştirimi esnasında gözlemlenmiş ve yapılan güncel çalışmalarda NVIDIA ekran kartlarının çoğunlukla kullanıldığı gözlemlenmiştir [4,16].

Veri kazıma denemeleri sırasında, BeautifulSoup kütüphanesinin dinamik web sitelerini kazımakta tek başına yetersiz olduğu, html içeriğinin sayfada ön yüklü gelmediği durumlarda web sitelerinin kazınmadığı ve veri elde edilemediği gözlemlenmiştir.

Ses sentezleme aşamasında 'Andrew Tate said he broke a woman's jaw and that his business was a scam ahead of Romanian arrest' cümlesi 3 model üzerinden sentezlenmiştir. Tüm sentezlemeler insan sesinden ayırt edilmesi zor sesler üretmiştir. Modellerin sentez süresi ve üretilen dosya boyutları Tablo 3.1'de verilmiştir. Ayrıca Tacotron2 modeli ile sentezlenen sesin az da olsa daha kaliteli olduğu gözlemlenmiştir.

Tablo 3.1 Sentezlenen ses verisi hakkında bilgiler

TTS MODELİ	Sentez süresi	Dosya boyutu
YOUR_TTS (VITS)	1,808 s	238 KB
TACOTRON2	4,168 s	305 KB
GLOW-TTS	1,574 s	290 KB

Kaynaklar

1. Bozkuş AY. Veri kazıma yöntemleri ile siber tehdit istihbaratı aracının geliştirilmesi (yüksek lisans tezi). Elazığ: Fırat Üniversitesi; 2021. <https://tez.yok.gov.tr/>
2. Uzun E. A novel web scraping approach using the additional information obtained from web pages. IEEE Access 2020; 8: 61726-61740. doi.org:10.1109/ACCESS.2020.2984503.
3. Kutlugün MA. Gözetimli makine öğrenmesi yoluyla türe göre metinden ses sentezleme (yüksek lisans tezi). İstanbul: Sabahattin Zaim Üniversitesi; 2017. <https://tez.yok.gov.tr>
4. Tan X, Qin T, Soong F, Liu TY. A survey on neural speech synthesis, arXiv preprint doi.org:10.48550/arXiv:2106.15561, 2021.
5. Stylianou Y. Applying the harmonic plus noise model in concatenative speech synthesis. IEEE Transactions on Speech and Audio Processing 2001; 9(1): 21-29. doi.org:10.1109/89.890068.
6. Schwarz D. Concatenative sound synthesis: the early years. Journal of New Music Research, 2006, 35 (1): 3-22. doi.org:10.1080/09298210600696857
7. Sturm B. Adaptive concatenative sound synthesis and its application to micromontage composition. Computer Music Journal 2006; 30: 46-66. doi.org:10.1162/comj.2006.30.4.46.
8. Kayte S, Mal M, Kayte C. A corpus-based concatenative speech synthesis system for marathi. IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) 2015; 5: 7. doi.org:10.9790/4200-05612026.

9. Purohit H, Endo T, Yamamoto M, Kawaguchi Y. Hierarchical Conditional Variational Autoencoder Based Acoustic Anomaly Detection. 2022; doi.org:10.48550/arXiv.2206.05460.
10. Awad M, Khanna R. Deep neural networks. Efficient Learning Machines 2015; 127-147 [doi.org:10.1007/978-1-4302-5990-9_7](https://doi.org/10.1007/978-1-4302-5990-9_7)
11. Kim J, Kim S & Kong, Jungil & Yoon, Sungroh. Glow-tts: a generative flow for text-to-speech via monotonic alignment search. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada. [doi.org:10.48550/arXiv.2005.111129](https://doi.org/10.48550/arXiv.2005.111129)
12. Peng J, Ping W, Song Z, Zhao K. Non-autoregressive neural text-to-speech. In International Conference on Machine Learning, 2020; 10192–10204.
13. Ping W, Peng K, Gibiansky A, Arik SÖ, Kannan A, Narang S, ve diğ. Deep voice 3: scaling text-to-speech with convolutional sequence learning. In International Conference on Learning Representations, 2018.
14. Prenger R, Valle R, Catanzaro B. Waveglow: A flow-based generative network for speech synthesis. In International Conference on Acoustics, Speech and Signal Processing, pages 3617–3621, 2019.
15. Rabiner LR. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257–286, 1989.
16. Edresson Casanova, Julian Weber, Christopher Dane Shulby, Arnaldo Cândido Júnior, Eren Gölge, and Moacir A. Ponti, Yourtts: Towards zero-shot multispeaker TTS and zero-shot voice conversion for everyone. ICML 2022; doi.org:10.48550/arXiv.2112.02418.

Ekler

Ek A

Uygulama Gerçekleştirme Aşamasında Kullanılan Bazı Kodlamalar

```
import requests
import uuid
import re
from flask import Flask, render_template, request, send_from_directory
from bs4 import BeautifulSoup
from TTS.api import TTS
app = Flask(__name__)
app.config["TEMPLATES_AUTO_RELOAD"] = True
@app.route('/', methods=['GET', 'POST'])
def index():
    errors = []
    headings = []
    _tts = ""
    if request.method == "POST":
        # get url that the person has entered
        try:
            url = request.form['url']
            _tts = request.form.get('tts')
            r = requests.get(url)
            print(url)
        except:
```

```

        errors.append( "URL parse edilemedi")
        return render_template('index.html', errors=errors)
    if r:
        # text processing
        raw = BeautifulSoup(r.text, 'html.parser')
        raw.decode('utf-8','ignore')
        headings = [str(w.get_text()) for w in raw.find_all(re.compile("^h[1-6]$"))]
        print(headings)

    return render_template('index.html', errors=errors, results=headings, tts=_tts)

@app.route('/sound', methods=['GET', 'POST'])
def sound():
    if request.method == 'POST':
        _text = request.form.get('text')
        _tts = request.form.get('tts')
        model_name = "tts_models/multilingual/multi-dataset/your_tts"
        if _tts == "your_tts":
            model_name = "tts_models/multilingual/multi-dataset/your_tts"
            tts = TTS(model_name)
            _suffix = "your_tts"
            # Init TTS with the target model name
            _id =str(uuid.uuid4())
            tts.tts_to_file(text=_text, speaker=tts.speakers[0], language=tts.languages[0],
file_path="static/sound/"+_id+_suffix+".wav")
        elif _tts == "tacotron2":
            model_name = "tts_models/en/ljspeech/tacotron2-DDC"
            _suffix = "tacotron2"
            tts = TTS(model_name)
            # Init TTS with the target model name
            _id =str(uuid.uuid4())
            tts.tts_to_file(text=_text, file_path="static/sound/"+_id+_suffix+".wav")
        elif _tts == "glow-tts":

```

```

    model_name = "tts_models/en/ljspeech/glow-tts"
    _suffix = "glow-tts"
    tts = TTS(model_name)
    # Init TTS with the target model name
    _id =str(uuid.uuid4())
    tts.tts_to_file(text=_text, file_path="static/sound/"+_id+_suffix+".wav")
else:
    return "Model belirtiniz!"
return _id+_suffix
else:
    return "Bu sayfayı görmeye yetkiniz yok!"
@app.route('/static/sound/<path:filepath>')
def data(filepath):
    return send_from_directory('static/sound/', filepath)
if __name__ == '__main__':
    app.run(debug=True)

```

Ek B

Dıř Cilt Kapađı



Veri Kazıma Yoluyla Web Sitelerinden Elde Edilen Başlıkların Seslendirilmesine Yönelik Bir Metin Madenciliği Yaklaşımı

Yazılım Mühendisliği Ana Bilim Dalı
Tezsiz Yüksek Lisans Bitirme Projesi

Baran Şenel

Öğrenci No Y210234029

Tez Danışmanı: Prof. Dr. Ayşegül Alaybeyoğlu
Ocak 2023

ŞENEL VERİ KAZIMA YOLUYLA WEB SİTELERİNDEN ELDE EDİLEN BAŞLIKLARIN
SESLİNDİRİLMESİNE YÖNELİK BİR METİN MADENCİLİĞİ YAKLAŞIMI
TEZSİZ YÜKSEK LİSANS BİTİRME PROJESİ 2023

Özgeçmiş

Adı Soyadı: Baran Şenel
E-mail (1): baransenel@ikcu.edu.tr
E-mail (2): baransenel@hotmail.com

Eğitim:
2015–2009 Adnan Menderes Üniversitesi, Ziraat Müh. Bölümü

İş Deneyimi:
2015 – 2023 Softek bilişim teknoloji ve yazılım iç ve dış ticaret A.Ş