



BURSA TEKNİK ÜNİV. MÜH. ve DOĞA BİL. FAK.  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

**BİTİRME PROJESİ ARA RAPOR FORMU**

**ÖĞRENCİ BİLGİLERİ**

	NUMARA	AD SOYAD	E-posta	İMZA <sup>1</sup>
1. ÖĞRENCİ	20360859111	FAZLİDDİN KHAMDAMOV	hfazliddin97@gmail.com	
2. ÖĞRENCİ	21360859114	HABİB SALİM	salimovhabib02@gmail.com	
3. ÖĞRENCİ				

<sup>1</sup> Her öğrenci tarafından imzalanmalıdır.

**DANIŞMAN ÖĞRETİM ELEMANI**

UNVAN, AD, SOYAD	Dr. Öğr. Üyesi Hayri Volkan AGUN	İMZA	
(VARSA) 2 . DANIŞMAN UNVAN, AD, SOYAD		İMZA	

**PROJE BİLGİLERİ**

DÖNEM	2022-2023 BAHAR
PROJE BAŞLIĞI	KELİMELER ARASI ÖRÜNTÜLERİN RAG TABANLI VEKTÖREL ANALİZİ
FİRMA İŞBİRLİĞİ	<input checked="" type="radio"/> Yok <input type="radio"/> Var
FİRMA İŞBİRLİĞİ VARSA FİRMA ADI VE İŞBİRLİĞİNİN KAPSAMI (Konu Önerisi, Veri Kümesi, Altyapı, Danışmanlık vb.)	
PROJE ÖZETİ * (Min 50 Maks 150 Kelime)	Bu çalışma, modern doğal dil işleme (NLP) teknikleri kullanarak kelimeler arasındaki semantik ilişkileri ve örüntüleri tespit etmeye yönelik bir araştırma projesidir. Çalışmada, vektör temsilleri (word embeddings) kullanarak kelimelerin anlamsal benzerliklerini ölçme, ilişkili kelime gruplarını tespit etme ve metinler içindeki semantik yapıları ortaya çıkarma amaçlanmaktadır. Projede Sentence Transformers kütüphanesi ile kelimelerin vektöre dönüştürülmesi, Langchain framework ve Chroma vektör veritabanı ile etkin sorgulamaların yapılması ve çıkan sonuçların görselleştirilmesi gerçekleştirilmiştir. Bu ara rapor, projenin mevcut durumunu, gerçekleştirilen çalışmaları ve elde edilen ilk bulguları sunmaktadır.



# BURSA TEKNİK ÜNİV. MÜH. ve DOĞA BİL. FAK. BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

## ARA RAPOR

### ARA RAPOR ZAMANINA KADAR YAPILMIŞ OLAN ÇALIŞMALAR \*

(Maddeler halinde net olarak yazılmalıdır.)

- **Çalışma Ortamı ve Gerekli Yazılımların Kurulumu**  
İlk aşamada, projenin geliştirme ortamı hazırlanmış ve gerekli kütüphaneler kurulmuştur.
- **Örnek Kelime Listesi ve Metin Kümesi Oluşturulması**  
Kelime listesi, Türkçe'nin en sık kullanılan 1000 kelimesinden oluşturulmuştur. Bu liste, TDK kaynakları ve çeşitli Türkçe korpuslar kullanılarak derlenmiştir. Metin kümesi ise farklı kategorilerden (edebiyat, bilim, günlük haberler, sosyal medya) toplam 500 metin parçasından oluşmaktadır.
- **Kelimelerin ve Metinlerin Vektörlere Dönüştürülmesi**  
Kelimelerin ve metinlerin vektörlere dönüştürülmesi için Sentence Transformers kütüphanesi kullanılmıştır. Özellikle Türkçe için ince ayar (fine-tuning) yapılmış olan "dbmdz/bert-base-turkish-cased" modeli tercih edilmiştir. Bu model, Türkçe'nin dil özelliklerine uygun olarak eğitilmiş bir BERT modelidir.
- **Vektör Veritabanı Kurulumu**  
Oluşturulan vektörlerin etkili bir şekilde sorgulanabilmesi için Langchain framework ve Chroma vektör veritabanı kullanılmıştır. Bu kombinasyon, semantik aramalar için optimize edilmiş bir altyapı sunmaktadır.
- **Kelime İlişkilerinin Analizi**  
Vektör veritabanı üzerinde çeşitli sorgular yapılarak kelimeler arasındaki ilişkiler analiz edilmiştir.
- **Sonuçların Görselleştirilmesi**  
Elde edilen analiz sonuçları, çeşitli görselleştirme teknikleri kullanılarak oluşturulmuştur.

### Ara Rapor Sonrası Dönem Sonuna Kadar Yapılması Planlanan Çalışmalar \*

(Maddeler halinde net olarak yazılmalıdır.)

- **Veri Kümesinin Genişletilmesi:** Kelime listesi 5000 kelimeye, metin kümesi ise 2000 metne çıkarılacaktır.
- **Model Karşılaştırması:** Farklı dil modellerinin (BERT, RoBERTa, XLM-R) performansları karşılaştırılacaktır.
- **Kelime İlişkileri Analizi:** Daha detaylı kelime ilişki analizleri (hiperonimler, hiponimler, meronimler vb.) yapılacaktır.
- **Web Uygulaması:** Elde edilen bulgulara dayalı interaktif bir web uygulaması geliştirilecektir.

### GENEL DEĞERLENDİRME \*

(Şu sorulara cevap veriniz:

1. Çalışma planladığınız şekilde ilerliyor mu?

1. Açıkçası projede biraz geride kaldık. Temel işlemleri (ortam kurulumu, vektörlere dönüştürme, veritabanı kurulumu) tamamladık ama vize sınavları ve diğer derslerin projeleri



# BURSA TEKNİK ÜNİV. MÜH. ve DOĞA BİL. FAK. BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

2. Yaşadığınız aksaklık veya sorun var mı?
3. Eğer planladığınız gibi gitmiyorsa veya aksaklıklar var ise bunların sizi etkilememesi için ne tür tedbirler alacaksınız veya aldınız? )

yüzünden zaman sıkıntısı yaşadık. Özellikle son iki haftada analiz kısmını hızlandırmamız gerekiyor. Şu an kabaca 1-2 adım gerideyiz planlanan takvime göre.

## 2. Birkaç sorunla karşılaştık:

Türkçe için iyi çalışan bir model bulmakta zorlandık. Kullandığımız model bazı kelime ilişkilerini yeterince iyi yakalayamıyor.

Chroma veritabanı büyük veri setlerinde yavaşlık gösteriyor.

Bilgisayarımın işlemci gücü, büyük veri setlerini işlemekte zorlanıyor, özellikle t-SNE görselleştirmelerinde.

Vize sınavları ve diğer dersler için harcadığımız zaman, projede ilerlememizi yavaşlattı.

## 3. Yaşadığımız sorunları çözmek için:

Daha hafif bir veri setiyle çalışıp, analiz ve görselleştirmeye odaklanacağız. İlk başta planladığımız 5000 kelimeye çıkmak yerine, 1500-2000 kelimeyle devam edip, çalışan bir prototip geliştireceğiz.

Performans sorununu çözmek için, bazı hesaplamaları daha küçük veri setleri üzerinde yapacağız.

Sınavlarda bittiği için çalışma tempomuzu arttırarak açığı kapatmayı planlıyoruz. Özellikle hafta sonları projeye daha fazla zaman ayıracacağız.

\* Doldurulması zorunludur.

Uygundur.  
Bölüm Başkanı  
İmza