

# GRAF TABANLI METİN ÖZETLEME PROJESİ

Mustafa Furkan Öztürk 200202148

Kocaeli Üniversitesi Bilgisayar Mühendisliği

## 1-)Özet

Bu projede bizden istenen verilen bir dokümandaki cümlelerin graf yapısına dönüştürülmesi ve bu graf modelinin görselleştirilmesidir.

Ardından graf üzerindeki düğümler ile özet oluşturan bir algoritma oluşturulması beklenmektedir.

## 2) Giriş

Bu kısım projeye ilgili ön bilginin oluşması için yazılmıştır.

Projenin Amaç kısmı incelendiğinde bizden bir docx dokümanını yükleme yapmamız ve sonrasında bu docx dosyası üzerinden işlemler yapmamız istenmektedir.

İlk etapta proje için bir yol haritası çıkarılarak ön hazırlık sürecine girilmiştir. Projenin kodunu yazmaya başlamadan önce fonksiyonlar ve hangi dil ile yazacağımızla alakalı çeşitli araştırmalar yapıldı. Bu aşamada projenin isterlerine göre araştırmalar gerçekleştirilmesi adına internetten benzer projeler incelenmiş, projenin ana hatları ortaya çıkarılmış ve büyük ölçüde karşılaşılabilecek problemler saptanıp çözümlendirilmeye çalışıldıktan sonra IDE ortamında projenin ilk adımları ortaya atılmıştır.

Yapılan ön hazırlık sürecinde docx dokümanından verilere eriştikten sonra proje isterlerinin nasıl gerçekleştirilebileceği gibi problemler üzerinde durulmuştur. Bu konulara ve problemlere yönelik gerekli araştırmalar yapıldıktan sonra projeye şekil verme aşamasına gidilmiştir.

Proje ön hazırlık süreciyle birlikte yaklaşık 1 haftalık bir süreçte tamamlanmıştır.

## 3-)Yöntem

Proje 2 aşama olarak değerlendirildi. Back-end ve görsellik.

İlk olarak yapılacak dokümana erişmek için uygun bir modül bulmak oldu. Bunun için doc2txt modülü kullanıldı.

İkinci adımda projede istenilen graph haline getirme işlemi için neo4j veritabanı kullanıldı.

Metin cümlelere ayrıldı. Her node bir cümle olarak ele alındı.

Üçüncü adımda ise graph oluşturma işlemi tamamlandıktan sonra doğal dil ön işleme adımları uygulandı.

Bunlar Tokenization, Stemming, Stop-word ve noktalama işaretlerini kaldırdığımız adımlardı.

Daha sonra cümleler arası anlamsal ilişkiyi kurmak için Word Embedding kullanıldı.

Bu aşamadan sonra cümleler arası benzerlik hesaplandı. Benzerlik kosinüs benzerliği ile hesaplandı.

Bu aşamadan sonra cümle skoru hesaplamak için çeşitli parametreler oluşturuldu.

Bunlar: Cümle özel isim kontrolü, Cümlede numerik veri olup olmadığının kontrolü, Cümle benzerliği threshold'unu geçen node'ların bulunması, Cümlede başlıktaki kelimelerin olup olmadığının kontrolü ve Her kelimenin TF-IDF değerinin hesaplanması.

Bu parametreler ile cümle skoru hesaplama işlemi yapıldı.

Daha sonra cümle skoruyla metin özetleme işlemini gerçekleştirdik.

Son olarak rouge puanını hesapladık.

Projenin back end kısmını bitirdikten sonra masaüstü uygulama aşamasına geçtik.

Masaüstü uygulamada doküman yüklediğimiz, graph görüntülediğimiz ve metin özeti, beklenen özet, rouge skor gibi değerleri gördüğümüz arayüz yapıldı.

## 4-) Deneysel Sonuçlar

Bu kısım proje öncesi ve sonrası araştırmaları, deneysel çalışmaları ve karşılaşılan sorunlar için yazılmıştır.

Öncelikle projede kullanabilecek mobil programlama dilleri için araştırılma yapıldı. Doğal dil işleme alanında fazla bir tecrübeye sahip olunmadığı için internet kaynaklarından benzer projeler incelendi ve hangi programlama dilinin kullanım açısından daha uygun olduğu tespit edildi.

Daha sonra uygulamanın nasıl geliştirileceğiyle alakalı çeşitli araştırmalar yapıldı.

Öncelikle kısmen daha kolay olan backend kısmı yani fonksiyonel kısımlar için dizayn ve kodlamalar yapıldı. Bu aşamada istenilen sonuçlar alındı.

Yukarıda bahsedilen işlemin ardından bir diğer problem graphı uygulamada gösterme işlemiydi. Bunun için networkx kütüphanesi kullanıldı ve son kullanıcıyı tatmin eden bir görünüm verilmeye çalışıldı.

Bu kısım çok uzun ve detaylı araştırmalar neticesinde çözülmeye çalışıldı. Graph kısmı çok karmaşık bir halde gözüktüğü için düzgün bir tasarım vermek çok uğraştırıcı olmasına rağmen sonucu gördüğümüzde tüm

emeklerimize değdiđinin farkına vardık.

Projemizin isterlere uygun bir řekilde olabilmesi iin hassas bir alıřma yapıldı.

## **2.1-Proje Sırasnda Yararlanılan Teknolojiler**

Projeyi Python programlama dili kullanarak Pycharm geliřtirme ortamında ve Windows iřletim sistemi zerinde gerekleřtirdik.

Programı yaparken Python programlama dilinin bize sunduđu ktphanelerin zelliklerinden yararlanıldı. Pythonda kullanılacak eřitli ktphanelere ve algoritmalara karar verdikten sonra arařtırmalar yapıldı.

## **5-)Sonu**

Sonu olarak projemiz sayesinde hem Back-end hem de Front-end alanında ok deđerli bilgiler đrenilmiř oldu.

Farklı teknolojiler hakkında bilgiler edinildi. Farklı aralar tanındı ve kullanıldı.

Projede her ne kadar zorlayan kısımlar olsa da bu sorunları internetteki deđerli kaynaklar sayesinde özebildik.

Yapılan alıřmalar sonucunda đrenilen yeni teknolojiler sayesinde farklı bir bakıř aısı kazandık.

Proje sayesinde algoritma becerileri ve tasarım zerinde kendimizi geliřtirmemiz bize avantaj sađlayan diđer bir konudur.

zetle bu projenin grnen kısmı her ne kadar bir masast uygulaması gibi gzkse bile hem Back-end tarafında yapılan iřlemler hem de kullanılan farklı yapılar zerinde alıřma yapılarak ok deđerli bilgilere kavuřulmuř oldu . Bu sayede kendimize yeni pencereler amıř olduk.

## **6-)Tasarım**

**Akış řeması;**

Kısıım ektedir.(1)

## **7-)Kaynaka**

<https://neo4j.com/docs/>

<https://networkx.org/https://developer.android.com/reference/android/text/DynamicLayout>

<https://igraph.org/python/>

<https://www.javatpoint.com/android-tutorial>

<https://www.nltk.org/>

<https://docs.python.org/3/library/tkinter.html>

<https://www.w3schools.blog/android-tutorial>

<https://www.youtube.com/watch?v=Y5W8Ef2hh-s>

**AKIŞ ŞEMASI (EK):**

