

KOCAELİ ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

YAZILIM LABORATUVARI II

2022-2023 Bahar Dönemi III. Proje

FURKAN AYDOĞAN - ALPER TALHA KÜÇÜK- OZAN
BOZUKLU

180202085@kocaeli.edu.tr / 180202034@kocaeli.edu.tr /
180202006@kocaeli.edu.tr

1.Özet

Bu projede verilen bir dokümandaki cümlelerin graf yapısına dönüştürülmesi ve bu graf modelinin görselleştirilmesi sağlanmıştır. Graf üzerindeki düğümler ile özet oluşturan bir algoritma oluşturulması sağlanmıştır

2.Giriş

Projede bir masaüstü uygulama geliştirilmiştir. Masaüstü uygulamada ilk olarak doküman yükleme işlemi gerçekleştirilmiş ardından yüklenen dokümandaki cümleleri graf yapısı haline getirilip bu graf yapısı görselleştirilmiştir. Bu grapta her bir cümle bir düğümü temsil etmektedir. Cümleler arasındaki anlamsal ilişki kurulup, cümleler skorlanmıştır. Belirli parametreleri kullanarak cümle skorunun hesaplama algoritmasını geliştirilmiş ve cümle skorlarına göre metin özeti çıkarma algoritmada geliştirilmiştir.

3.Yöntem

Bu kod, metin belgelerinin özetlenmesi için bir belge özetleme aracı sunmaktadır. Tkinter kullanılarak basit bir kullanıcı arayüzü oluşturulmuştur. Kullanıcı bir metin belgesi yükler ve belgedeki cümleleri özetlemek için farklı yöntemleri kullanabilir. İlk adımda,

yüklenen belgeyi işlemek için metin ön işleme işlemleri yapılır. Belge cümlelere ayrılır, küçük harflere dönüştürülür, noktalama işaretleri ve durak kelimeleri (stop words) çıkarılır. Ardından, cümlelerin ön işlenmiş halleri kullanılarak cümle skorları hesaplanır. Cümle skorları, cümlelerin çeşitli özelliklerine dayanarak hesaplanır. Özellikler arasında cümledeki özel isimlerin oranı, sayısal verilerin oranı, cümleler arasındaki benzerlik, cümledeki başlık kelimelerinin sayısı ve cümledeki tema kelimelerinin sayısı bulunur. Her özelliğe ağırlık verilerek cümle skoru hesaplanır.

Belge başlığındaki kelimeler tema kelimeleri olarak kabul edilir ve TF-IDF (term frequency-inverse document frequency) yöntemi kullanılarak belgedeki tema kelimeleri belirlenir. Cümle skorları hesaplandıktan sonra, skora göre cümleler sıralanır ve en yüksek skora sahip cümleler özet olarak seçilir.

Seçilen cümlelerin yanı sıra, cümleler arasındaki benzerlikler de dikkate alınır. Belirli bir benzerlik eşliğini geçen cümleler arasında bir bağlantı kurulur ve bir graf oluşturulur. Graf görselleştirilir ve görsel olarak sunulur.

4.Algoritma

Bu projede, belge özetleme aracı olarak kullanılan bir metin işleme uygulamasını temsil etmektedir. Tkinter kütüphanesi kullanılarak basit bir kullanıcı arayüzü oluşturulmuştur. Kullanıcı, bir metin belgesini yükleyebilir ve belgenin özetini elde edebilir.

İşleme adımları ise şu şekilde gerçekleştirilmektedir:

Metin içeriği ön işleme işleminden geçirilir. Cümlelere bölünür, kökleri çıkarılır, gereksiz kelimeler (stop words) ve noktalama işaretleri çıkarılır. Cümle skorları hesaplanır. Her cümle için bir skor hesaplanır ve bu skorlar cümlelerin önem düzeylerini temsil eder. Skorlar, cümledeki isimlerin oranı (P1), sayısal verilerin oranı (P2), cümledeki diğer cümlelere benzerliği (P3), cümlede başlıkta bulunan kelimelerin oranı (P4) ve cümledeki belirli bir temaya ait kelimelerin oranı (P5) gibi faktörlere dayanır. En yüksek skorlara sahip cümleler özet olarak seçilir. Özet, en yüksek skorlara sahip belirli bir sayıdaki cümleyi içerir. Belirli bir benzerlik eşiği üzerindeki cümleler arasında bir graf oluşturulur. Graf, cümleler arasındaki bağlantıları ve benzerlikleri temsil eder. Graf, ağırlıklı bir graf olarak görselleştirilir. Cümleler düğümler olarak temsil edilir ve benzerlik ağırlıkları kenarlar arasında gösterilir. Kullanıcı, özeti görüntüleyebilir ve bir referans metniyle karşılaştırarak ROUGE skorlarını hesaplayabilir. ROUGE, otomatik özetleme sistemlerinin performansını değerlendirmek için yaygın olarak kullanılan bir ölçüttür.

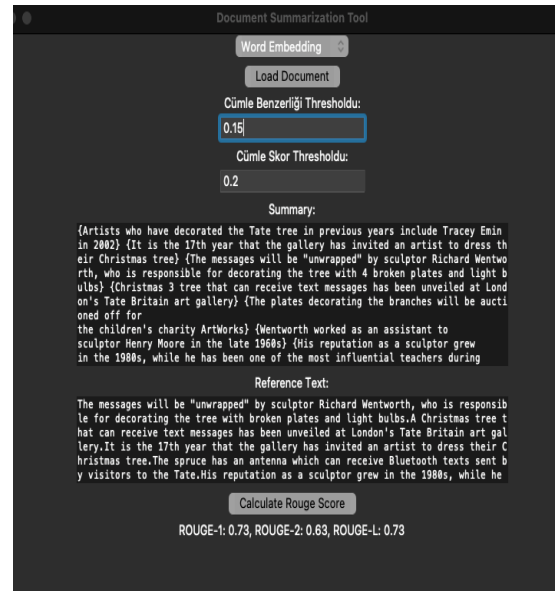
Bu kod, metin özetleme algoritmasının basit bir uygulamasını sunmaktadır. Özetleme işlemi için belirli bir metin belgesi yüklenir, cümle skorları hesaplanır, en önemli cümleler seçilir ve özet oluşturulur. Grafik görselleştirme ve ROUGE skorları gibi ek işlevler de sunulmaktadır. Bu kodun kullanıcı dostu bir arayüzle sunulması, metin özetleme sürecini kullanıcılar için daha erişilebilir hale getirir.

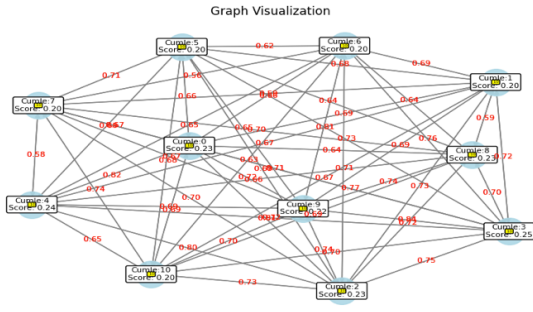
5. Sözde Kod (PseudoCode)

1. `preprocess_text(content)`: İçeriği cümlelere ayırma. Stemming işlemi yapma. Stop kelimeleri kaldırma. İşlenmiş cümleleri döndürme.
2. `calculate_sentence_scores(sentences, title, content)`: Cümle skorlarını hesaplama.

3. Cümlelerin orijinal hallerini saklama. Tema kelimelerini belirleme. TF-IDF değerlerini hesaplama. Cümle skorlarını döndürme.
4. `get_title(paragraph)`: Paragraftan başlığı alma. Başlığı cümlelere ayırma. Başlığı döndürme.
5. `bert_method(sentences)`: Cümlelerin gömme (embedding) vektörlerini hesaplama. BertTokenizer ve BertModel kullanarak cümle gömme vektörlerini döndürme.
6. `word_embedding(sentences)`: Cümlelerin gömme (embedding) vektörlerini hesaplama. Word2Vec modelini kullanarak cümle gömme vektörlerini döndürme.
7. `load_document()`: Belgeyi yükleme. İçeriği ve başlığı alarak işleme sokma. Cümle gömme vektörlerini hesaplama. Cümle skorlarını ve özetlemeyi hesaplama. Özeti güncelleme. Eşik değerini güncelleme. Grafik görselleştirme.
8. `calculate_rouge_score()`: Referans metni ve özeti alarak ROUGE skorlarını hesaplama. ROUGE skorlarını etikete yazma.

6. Çıktılar



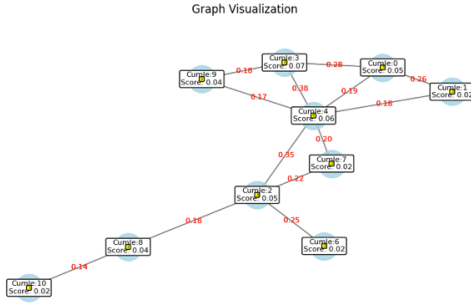


<https://www.javatpoint.com/graph-algorithms>

<https://www.freecodecamp.org/news/graph-algorithms-and-data-structures-explained-with-java-and-c-examples/>

<https://www.educative.io/blog/graph-algorithms-tutorial>

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3439



7. Sonuçlar

Bu proje, Python'daki Tkinter kütüphanesi kullanılarak uygulanan bir belge özetleme işlevini gerçekleştirir. Metni ön işleme, cümle puanlarını hesaplama, Word Embedding ve BERT yöntemlerini kullanarak cümle yerleştirmeleri oluşturma, bir belge yükleme, cümle benzerliklerinin grafiğini görselleştirme ve oluşturulan özet için ROUGE puanlarını hesaplama fonksiyonlarını içerir. Sonuç olarak bu araç, kullanıcıların belgeleri yüklemesine, cümle puanlarına ve yerleştirmeye dayalı özetler oluşturmaya, cümle benzerlik ilişkilerini görselleştirmesine ve ROUGE puanlarını kullanarak özetlerin kalitesini değerlendirmesine olanak tanıyan kullanıcı dostu bir belge özetleme arayüzü sağlar.

8. Kaynakça

<https://www.geeksforgeeks.org/graph-data-structure-and-algorithms/>

<https://towardsdatascience.com/10-graph-algorithms-visually-explained-e57faa1336f3>

