

Genetik Algoritmalar Kullanılarak Metinlerde Otomatik Özet Çıkarma

Automatic Summary Extraction in Texts Using Genetic Algorithms

Abdullah Ammar KARCIOĞLU
Bilgisayar Mühendisliği Bölümü
Ege Üniversitesi İzmir, Türkiye
abdullah.ammar.karcioglu@ege.edu.tr

Ahmet Cahit YAŞA
Bilgisayar Mühendisliği Bölümü
Celal Bayar Üniversitesi Manisa, Türkiye
ahmet.yasa@cbu.edu.tr

Özetçe—Otomatik metin özetleme üzerinde uzun süredir çalışılan doğal dil işlemenin uygulama alanlarından biridir. Web kaynaklarındaki bilgi miktarındaki artış, otomatik metin özetleme yöntemlerine ihtiyacın artmasına neden olmuştur. İnsan eliyle oluşturulan özetler üretmek için bir sistem tasarlamak zordur ve bu nedenle birçok araştırmacı bir tür özetleme olan cümle veya paragraf çıkarmaya odaklanmıştır. Bu çalışmada, bu tür özetleri oluşturmak için genetik algoritmalar kullanarak oluşturulan bir yöntem tanıtıyoruz. Metinler ön işlemeden geçirildikten sonra söz dağarcığı oluşturulup önerilen yönteme girdi olarak verilmektedir. Genetik Algoritma temelli cümle seçimi, özet yapmak için kullanılır ve özet oluşturulduktan sonra uygunluk fonksiyonu kullanılarak değerlendirilir. Uygunluk fonksiyonu ilk modelimizde her bir kelimenin frekansına ve kelime çifti frekanslarına dayanmaktadır. Uygulanan modelin sonuçları aynı veri setini tf-idf tabanlı başka bir yöntemde kullanarak anma(recall), duyarlılık (precision), f-skoru ve Rouge metrikleri ile tartışılmıştır.

Anahtar Kelimeler — Doğal Dil İşleme, Genetik Algoritmalar, Seçilmiş Otomatik Metin Özetleme

Abstract—Automatic text summarization is one of the applications of natural language processing that has been studied for a long time. The increase in the amount of information in web resources has increased the need for automatic text summarization methods. It is difficult to design a system to produce abstracts created by human hands. For this reason, many researchers have focused on extracting sentences or paragraphs, which is a kind of summary. In this study, we introduce a method that was created using genetic algorithms to generate such summaries. After the texts are preprocessed, vocabulary is created and given as input to the proposed method. The sentence selection based on Genetic Algorithm is used to summarize and after that the summary is created, it is evaluated using the fitness function. In our first model, the fitness function is based on the frequency of each Word and the word pair frequencies. The results of the applied model are discussed using the same dataset in another method based on tf-idf, with precision, recall, f-score and Rouge metrics.

Keywords — Natural Language Processing, Genetic Algorithms, Extractive Automatic Summarization

I. GİRİŞ

Otomatik metin özetleme, son yıllardaki teknolojik gelişmeler sayesinde verinin ve bilginin hızlı bir şekilde büyümesine bağlı olarak bilgisayar programları tarafından bu

bilgi birikimi üzerinde işlemler yapabilmek için kullanılan yaklaşımlardan birisidir. Otomatik metin özetleri oluşturma, bu problemi uzun yıllar boyunca ele alarak, bir belgede en alakalı fikirleri sunan kısa metinler elde etmeye çalışmakla açıklanabilir. Otomatik özetleme işleminde metin bilgileri korunurken metnin boyutu küçültülmüş olur. Bir metin özetleyici, son kullanıcının amacına göre girdilerinin yoğunlaştırılmış temsilini üreten bir sistemdir [1]. Özet üretimi derin doğal dil işleme yöntemlerini içeren karmaşık bir işittir. Problemi basitleştirmek için araştırmalar soyut yöntemlerden ziyade metin üzerinden madencilik tekniklerinin kullanımına odaklanmıştır [2]. Bu teknik basitçe orijinal metnin alt cümleleri kullanması ile gerçekleştirilir. Bu yaklaşımla çıkarılan özetler parçanın ana temasının bulunmasını garanti etmeyebilir, ancak ana konuya uygunluk kararının verilmesi için metin içeriğini yaklaşık olarak rahatlıkla temsil ederler.

Metin özeti iki farklı yaklaşımla üretilebilir: Seçim (İng. Extractive) yolu, orijinal dokümanın bazı parçalarına referans verecek şekilde veya soyutlayarak bilgi verme (İng. Abstractive) yolu, metnin tamamıyla ilgili malumatı kapsayan bir çıkarım yapma şeklinde. Her iki durumda da özet kullanımının en büyük avantajı okuma zamanının azaltılmasıdır. Otomatik metotlar ile özet üretiminin başka avantajları da vardır: (i) özeti boyutu kullanıcı kontrolindedir; (ii) özet içeriği deterministiktir, metni belirleyicidir; (iii) özet içindeki metin parçaları ile orijinal metindeki pozisyonları arasında kolayca bağlantı kurulabilir. Otomatik metin özetleme araştırmalarında pek çok yöntem önerilmiştir. Çizge tabanlı yöntemler, öznitelik vektörü temelli yöntemler, kümeleme ve evrimsel yöntemler bunlar arasında gösterilebilir. Çalışmamızda global optimizasyon problemlerini çözmeye oldukça etkili olan Genetik Algoritmalar ve makine öğrenmesi algoritmaları belirli aşamalarda kullanılarak tek belge üzerinden metin özeti çıkarma yöntemi önerilmektedir. Evrimsel Algoritma işlemleri, bireylerin üretilmesi ve seçilmesi, mutasyonlar ve tekrar birey üretimi gibi simülasyon içinde yer alan ardışık veya iç içe yer alabilen olasılık temelli evreleri içermektedir.

Dokümanların ve özetlerin yer aldığı bir koleksiyon üzerinde klasik (eğitilebilir) bir makine öğrenme algoritmasının uygulanmasıyla eğitilebilir bir özetleyici elde edilebilir [3] Bu yaklaşımda dokümandaki her cümle metin öznitelik vektörleri olarak modellenir. Özetleme işlemi cümlelerin çıkarılan referans özetine ait olduğunu veya olmadığını etiketlediği iki

farklı sınıf olarak görülebilir. Eğitilebilir özetleyicinin bu sınıflara olan bağlantılı (korelasyon) kalıplara göre öğrenmesi beklenmektedir. Eğitim işleminden sonra yeni bir doküman sisteme verildiğinde, özete ait olan cümleler sınıflandırılacak ve özet çıkarılmış olacaktır. Burada önemli olana nokta öznitelik vektörlerinin nasıl elde edileceğidir.

II. İLGİLİ ÇALIŞMALAR

Silla ve arkadaşlarının yapmış oldukları bir çalışmada [4] otomatik metin özetleme, bir sınıflandırma (denetimli öğrenme) problemi olarak ele alınmıştır. Böylece makine öğrenmesine yönelik sınıflandırma yöntemleri bu belgeleri tanımlayan bir dizi niteliğe dayanan belgelerin özetlerini çıkarmak için kullanılmıştır. Çalışmada, otomatik metin özetleme görevini çözen sınıflandırma algoritmalarının performansını arttırmada Genetik Algoritma (GA) özelliği öznitelik seçiminin etkinliği araştırılmıştır. Ayrıca diğer çalışmalardan farklı olarak bu çalışma ClassSumm'un performansının bir ön işleme adımında karmaşık özellik seçim yöntemleri kullanılarak geliştirilip geliştirilemeyeceğini araştırmıştır. Metin özetleme literatüründe bir kriter olarak kullanılan TIPSTER koleksiyonunun The Wall Street Journal'dan çıkarılan haberlerden oluşan bir belge tabanı ile yapılan deneylerin hesaplama sonuçları rapor edilmiştir.

Meena ve Gopalanı yapmış oldukları çalışmada [5] otomatik metin özetlemede Genetik Algoritmalar kullanılarak kullanılan önemli teknikler ve metodolojiler hakkında bir araştırma yapılmıştır. Bu çalışma, Evrimsel Algoritma tekniklerini uygulayarak otomatik metin özetleme tekniklerindeki büyüme ve iyileşme hakkında bir inceleme sunmuştur. Uygunluk fonksiyonundaki ek özellikleri dikkate alan geniş bir öznitelik çıkarma seti önerilmiştir. Algoritmalarını değerlendirmek için DUC 2002 veri setinden 10 belge kullanılmıştır. Sonuçların değerlendirilmesi amacıyla Rouge-1 metriği kullanılmıştır. Rouge-q, her bir kelimenin tek kelimeyle çıkarılan özetle bulunan sistem tarafından oluşturulan özetle olup olmadığını kontrol eder. Bu çalışma, Genetik Algoritmalar vb. Evrimsel Algoritma tekniklerini entegre ederek, metin özetleme sonucunun nasıl geliştirilebileceğine genel bir bakış sunmuştur. Birkaç yinleme için elde edilen sonuçlar, Genetik Algoritmaların uygun ağırlıkları bulma gücünün olduğunu göstermiştir.

Neto ve arkadaşlarının yapmış oldukları bir çalışmada [6] doğrudan orijinal metinden çıkarılan bir dizi öznitelikli kullanan eğitilebilir makine öğrenmesi algoritmalarının uygulanmasına dayanan bir özetleme sistemi sunulmuştur. Metindeki bazı öğelerin sıklığına dayalı istatistiksel öznitelik ve metnin basitleştirilmiş yapısından çıkarılan dilbilimsel öznitelik olmak üzere iki çeşit öznitelik sunulmuştur. Suanmalı ve arkadaşlarının yapmış oldukları çalışmada [7] Genetik Algoritma (GA) ile seçilen özniteliklerin etkinliği araştırılmıştır. GA, bir uygunluk fonksiyonu için Rouge tarafından üretilen anma (recall) ölçütü kullanılarak değerlendirilen her bir öznitelikliğin ağırlığını öğrenmek üzere DUC 2002 veri setinde 100 belgenin eğitimi için kullanılmıştır. GA tarafından elde edilen ağırlıklar, önemli öznitelik puanını ayarlamak için kullanılmıştır. Sonuçlar Microsoft Word 2007 özetleyicisi ve Copernic özetleyicisi ile 100 belge ve 62 görünmeyen belge için karşılaştırılmıştır. Sonuçlar, özetler için

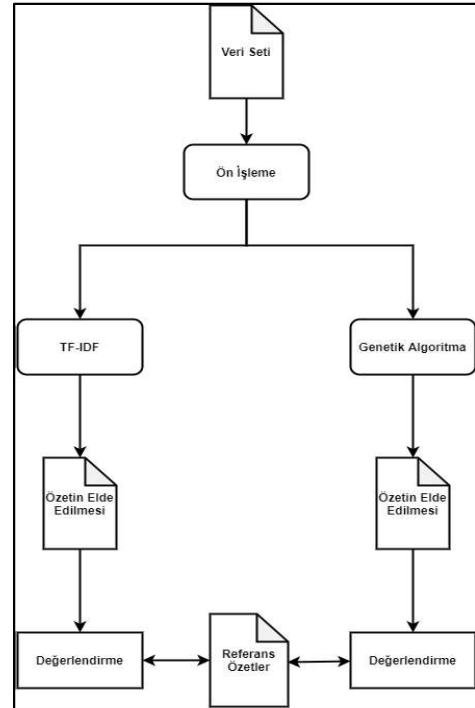
en iyi ortalama kesinlik (precision), anma (recall) ve f- skoru ölçütünün GA tarafından elde edildiği tespit edilmiştir.

Bhat ve arkadaşlarının yapmış oldukları bir çalışmada [8] özet çıkarma ve özet soyutlama gibi her iki yaklaşımın melez yöntemini sunmuşlardır. Özet çıkarma yönteminde anlamsal ilişki kullanılmıştır ve duyguların cümleler arasında önemli bir rol oynadığı varsayılmıştır. Üretilen özet çıkarma yöntemi daha sonra özetin soyutlamayı soyut özete dönüştürmek için bir hibrit özetleyici olarak sonuçlanan WordNet, Lesk algoritması ve POS'un bir birleşimi olan yeni dil üreticisiyle elde edilmiştir. Özetleme yöntemi DUC 2007 veri seti kullanılarak değerlendirilmiştir ve MS Word ile karşılaştırıldığında önemli sonuçlar elde edildiği gösterilmiştir.

Sadiq ve arkadaşlarının yapmış oldukları bir çalışmada [9], istatistik ve sezgisel yöntemleri birleştiren otomatik metin özetlemesi için hibrit bir sistem olan yeni yaklaşım sunulmuştur. Sistem, tek başına istatistik veya istatistik ile sezgisel yöntemlerden oluşmaktadır. İstatistiksel yöntemde özet, bazı istatistik özelliklerine göre çıkarılmışken, istatistik + sezgisel yöntemde özet, birleştirilmiş istatistik özelliklere ve kelime sıklığı, konum, cümle uzunluğu ve belge başlığı ile benzerlik gibi sezgisel özelliklere dayanarak çıkarılmıştır.

III. ANALİZ VE TASARIM

Şekil 1'de gösterildiği gibi veri setini ilk önce ön işlemeden geçirip daha sonra elde edilen cümleler üzerinde genetik algoritma ve tf-idf tabanlı istatistiksel yöntemler uygulanarak otomatik özet çıkarma gerçekleştirilmiştir. Ön işleme otomatik metin özetlemede önemli bir basamaktır, çünkü üretilen özetin kalitesi metin gösteriminin etkinliğine bağlıdır. Ön işleme 5 ana adımdan oluşmaktadır. Bunlar cümle segmentasyonu, ayrıştırma işlemi, durak kelimelerinin atılması, kelimeleri küçük harflere çevirme ve kök alma işlemidir.



Şekil 1 Sistem Modeli

Cümle segmentasyonunda, ön işleme adımında ilk önce dosyadan girdiler okunur, daha sonra düzenli ifadeler kullanılarak cümle içinde geçen noktalama işaretleri atılır. *Ayrıştırma*, doğal dil metnindeki karakter dizileri birbirinden farklı, anlamlı parçalara (token) bölünmesidir [12]. *Durak Kelimelerinin Atılması*, ön işlemede bir sonraki adım, belgelerin uygunluğu açısından faydalı bilgilere nadiren katkıda bulunan ve belgede sıkça görülen ancak belgenin önemli içeriğinin tanımlanmasında daha az anlam sağlayan sözcüklerin kaldırılmasıdır. *Küçük harfe çevirme* ve *kök alma*, ön işlemenin son aşaması kelimeleri küçük harflere çevirdikten sonra bunların köklerinin alınması işlemidir. Kök almadaki amaç verinin boyutunu azaltılması ve çekim eki almış farklı yazılan kelimelerin ortak paydada ifade edilmesidir [13].

Ön işleme adımları gerçekleştirildikten sonra girdi olarak verilen kelimelerin sözlük yapısı kullanılarak frekanslarının çıkarılması işlemi gerçekleştirilmiştir. Frekanslar hesaplanırken her bir kelimenin ayrı ayrı frekansı ve kelime çiftlerinin frekansları ayrı ayrı hesaplanmıştır. Python programlama dilinde kullanılan *DEAP* kütüphanesiyle genetik algoritma adımları gerçekleştirilmiştir. Genetik algoritmada uygunluk fonksiyonu başta bireylerin frekansları temel alınarak hesaplanmıştır. Temel bileşenleri:

- Bireylerin karakter dizisi formatında temsil edilmesi (kromozomlar),
- Uygunluk fonksiyonun bulunması,
- Popülasyonun oluşturulması ve seçilmesi,
- Ebeveyn bireylerinin seçim mekanizması,
- Tekrar kombinasyon oluşturma, çaprazlama ve mutasyon gibi farklı operatörler,
- Bireylerin tekrardan popülasyon içine dağıtılması olarak sayabiliriz

Uygunluk fonksiyonu, en iyi bireyleri, yani en iyi sonuçları bulmada belirgin bir rol oynamaktadır. Doküman özetini optimize etmeye yardımcı olan bir görev fonksiyonu olarak tanımlanabilir. Problem uzayı 0 ve 1'ler şeklinde kodlanmış karakter dizisi formundaki bireyler olarak temsil edilir; Permutasyon veya Değer gibi farklı kodlama tipleri de kullanılabilir. Genetik Algoritma yaklaşımında birey uzayı rastgele üretilir ve yukarıda bahsedilen operatörler kullanılarak bir sonraki adım için yeni nesil oluşturularak ilerletilir. Algoritma belirli bir tekrarlanma sayısına ulaşınca veya iyi bir uygunluk fonksiyonu sonucu seviyesine gelen bireyler elde edilince sonlanır. Yapılan çalışmada uygunluk fonksiyonu ve tekrarlanma sayısına göre oluşturulmuş en uygun 10 cümle seçim ile çıkarılarak listelenmiştir. Ayrıca tf-idf (terim frekansı-ters doküman frekansı) yöntemi ile kelimelerin cümlede geçme sıklıklarına göre özet cümleler çıkarılmıştır. En sonunda farklı yöntemlerle elde edilen özet cümlelerle, insanlar tarafından oluşturulmuş referans özet cümleler karşılaştırılarak performans değerleri hesaplanmıştır.

IV. VERİ SETİ VE SONUÇLAR

Geliştirilen yöntem pek çok farklı veri setleri üzerinde çalıştırılabilir. Üretilen özet metninin, doğrulama özetleri ile karşılaştırabilmek için, karşılaştırma özetlerinin ana

metinde geçen cümlelerin bir alt kümesi olmasına dikkat edilmelidir. BBC News Summary [10] veri seti 2004 – 2005 yılları arasında yayımlanan ekonomi, eğlence, politika, spor ve teknoloji konu başlıkları altında ayrı ayrı 200'den fazla toplam 2225 haber içeriği ve bunların seçim yaklaşımıyla üretilmiş özetlerini içermektedir. Deneysel çalışmada BBC News Summary veri seti kullanılmıştır. Değerlendirme kriterleri olarak da anma (İng. recall), duyarlılık (İng. precision), f-skoru ve ROUGE metrikleri kullanılmıştır. Bu hesaplarda TP, FP, TN, FN gibi nitelikler kullanılmıştır.

TP (İng. True Positive): Referans özette ve çıkarılan özetle de bulunan cümleler. *FP* (İng. False Positive): Referans özetle bulunmayan ama çıkarılan özetle bulunan cümleler. *TN* (İng. True Negative): Referans özetle bulunmayan ve çıkarılan özetle de yer almayan cümleler. *FN* (İng. False Negative): Referans özetle yer alan ancak çıkarılan özetle bulunmayan cümleler.

Kullanılan performans ölçütleri aşağıda açıklanmıştır.

Anma: Çıkarılan özetteki cümleleri doğru tespit etme oranı.

$$recall = \frac{TP}{TP + FN} \quad (1)$$

Duyarlılık: Çıkarılan özetteki cümleler hangi oranda doğrudur.

$$precision = \frac{TP}{TP + FP} \quad (2)$$

Görüldüğü üzere anma ve duyarlılık iki önemli metrik ve aralarında bir denge söz konusudur. Birini arttırmak için yapılan işlemler diğerini azaltacaktır. Bununla baş edebilmek için dengersiz dağılımlı veri setleri için F1-skoru kullanılmaktadır. Ekstra durumların ortaya çıkmasını önlemek için aritmetik ortalama yerine harmonik ortalama kullanılır.

$$F_{Score} = 2 * \frac{precision * recall}{precision + recall} \quad (3)$$

Tablo I'de genetik algoritma yöntemi kullanılarak üretilen özetler ve veri setinde bulunan referans özet dosyaları karşılaştırılarak anma, duyarlılık ve f-skoru değerleri hesaplanmıştır. Veri setinde her makale için ayrı ayrı özet dosyaları olması ile her bir test durumu için hesaplanan değerler toplamı, en son aşamada test uzayı sayısına bölünüp Tablo-I'de gösterildiği gibi ortalama değerler hesaplanmıştır.

TABLO I. GENETİK ALGORİTMA İLE ÜRETELEN ÖZETLERİN ANMA, DUYARLILIK VE F-SKORU DEĞERLERİ

Anma (%)	Duyarlılık (%)	F-Skor (%)
67.35	41.25	51.16
31.03	11.25	16.51
31.17	30.00	30.57
59.57	35.00	44.09
79.55	43.75	56.45
57.89	27.50	37.29
54.84	21.25	30.63
46.30	31.25	37.31
60.71	42.50	50.00
46.51	25.00	32.52
Ortalama Anma (%)	Ortalama Duyarlılık (%)	Ortalama F-Skor (%)
53,49	30,87	39,15

Tablo-II’de ise tf-idf yöntemi kullanılarak üretilen özetler ve veri setinde bulunan referans özet dosyaları karşılaştırılarak anma, duyarlılık ve f-skoru değerleri hesaplanmıştır. Veri setinde her makale için ayrı ayrı özet dosyaları olması ile her bir test durumu için hesaplanan değerler toplamı, en son aşamada test uzayı sayısına bölünüp Tablo-II’de gösterildiği gibi ortalama değerler hesaplanmıştır.

TABLO II. TF-IDF İLE ÜRETELEN ÖZETLERİN ANMA, DUYARLILIK VE F-SKORU DEĞERLERİ

Anma (%)	Duyarlılık (%)	F-Skor (%)
44.90	22.68	30.14
41.38	18.75	25.81
90.91	32.56	47.95
34.04	19.05	24.43
27.27	15.79	20.00
13.16	8.33	10.20
54.84	34.00	41.98
38.89	28.00	32.56
80.36	43.69	56.60
72.09	34.44	46.62
Ortalama Anma (%)	Ortalama Duyarlılık (%)	Ortalama F-Skor (%)
49.78	25.72	33.92

Metin özetleme çalışmalarında kullanılan bir diğer önemli ölçüt kriteri rouge metriğidir. Tablo III ve IV’de genetik algoritma ve tf-idf yöntemleri kullanılarak üretilen özetler ve veri setinde bulunan referans özet dosyaları karşılaştırılarak Rouge değerleri hesaplanmıştır. Rouge (İng. Recall – Oriented Understudy for Gisting Evaluation) otomatik üretilen özet ve metinlerin referans özetleri arasında karşılaştırma yapmak için kullanılır. Rouge-N formatında Rouge-1, Rouge-2, Rouge-3... gibi kelimelerin birlikte görünme frekansları hesaplamalarda değerlendirilir. Örneğin; Rouge-1 üretilen özet ve referans özet arasındaki unigram’ların örtüşmesine göre hesaplanır. Rouge-2 ise, üretilen ve referans özetleri arasındaki bigram’ların eşleşmesi ile hesaplanır. Bunlara ek olarak Rouge-L LCS ile en uzun eşleşen kelime dizinin ölçülmesini hesaplar. Rouge-W ise Rouge-L yönteminin ağırlıklandırılmış LCS’nin hesaplanması ile bulunur.

TABLO III. GENETİK ALGORİTMA İLE ROUGE 1-2-3-L-W SKORLARI

	Anma (%)	Duyarlılık (%)	F-Skor (%)
Rouge -1	70.00	40.00	50.91
Rouge -2	26.67	13.33	17.78
Rouge -3	0.00	0.00	0.00
Rouge -L	74.17	46.53	57.18
Rouge -W	47.72	35.98	41.03

TABLO IV. TF-IDF İLE ROUGE 1-2-3-L-W SKORLARI

	Anma (%)	Duyarlılık (%)	F-Skor (%)
Rouge-1	25.00	33.33	28.57
Rouge-2	0.00	0.00	0.00
Rouge-3	0.00	0.00	0.00
Rouge-L	31.50	40.03	35.26
Rouge-W	18.95	33.33	24.16

Tablo I ve Tablo III’deki genetik algoritma yöntemi kullanılarak oluşturulan özetlerde, Tablo II ve Tablo IV’deki tf-idf tabanlı oluşturulan özetlerden daha yüksek başarı sonuçları elde edilmiştir. Bunun nedeni, genetik algoritmada bireyler arasından en iyilerinin Seçilim (Elitizm) ile bir sonraki nesle

aktarılmasıdır. Ayrıca tf-idf tabanlı yöntemde sadece kelime frekansı temel alınırken, genetik algoritma yaklaşımında kelimelerin n-gram şeklinde görülme olasılıkları kullanıldığından başarıya etkisi gözlemlenmiştir.

V. SONUÇLAR VE GELECEK ÇALIŞMALAR

Genetik algoritmalar kullanılarak otomatik metin özetlemenin gerçekleştirildiği bu çalışmada başarı oranlarının yüksek olduğu ve özet çıkarma çalışmalarında daha iyi bireyleri seçerek problemi çözen genetik algoritmaların kullanılabilirliği gösterilmiştir. Gelecekte yapmayı hedeflediğimiz çalışmalar;

- Genetik algoritmanın uyum fonksiyonunu makine öğrenmesi algoritmaları ile geliştirecek melez bir sistem tasarlamak
- Farklı yapay zekâ yöntemlerini bir araya getirerek sistemin performansını geliştirmek.
- Kümeleme ve sınıflandırma yöntemleri kullanılarak birbirinden farklı cümlelere özet içerisinde yer verebilmek.

KAYNAKLAR

- [1] Mani, I. Automatic Summarization. Benjamin’s Publ. Co. Amsterdam Philadelphia (2001)
- [2] Sparck-Jones, K. Automatic summarizing: factors and directions. In Mani, I.; Maybury, M. Advances in Automatic Text Summarization. The MIT Press (1999) 1-12
- [3] Neto, Joel Larocca, Alex A. Freitas, and Celso AA Kaestner. "Automatic text summarization using a machine learning approach." Brazilian Symposium on Artificial Intelligence. Springer, Berlin, Heidelberg, 2002.
- [4] Silla C.N., Pappa G.L., Freitas A.A., Kaestner C.A.A. (2004) Automatic Text Summarization with Genetic Algorithm-Based Attribute Selection. In: Lemaitre C., Reyes C.A., González J.A. (eds) Advances in Artificial Intelligence – IBERAMIA 2004. IBERAMIA 2004. Lecture Notes in Computer Science, vol 3315. Springer, Berlin, Heidelberg
- [5] Meena, Y.K., Gopalani, D.: Evolutionary algorithms for extractive automatic text summarization. Procedia Comput. Sci. 48, 244–249 (2015)
- [6] Neto J.L., Freitas A.A., Kaestner C.A.A. (2002) Automatic Text Summarization Using a Machine Learning Approach. In: Bittencourt G., Ramalho G.L. (eds) Advances in Artificial Intelligence. SBIA 2002. Lecture Notes in Computer Science, vol 2507. Springer, Berlin, Heidelberg
- [7] Suanmali, L., Salim, N., & Binwahlan, M. S. (2011). Genetic algorithm based sentence extraction for text summarization. International Journal of Innovative Computing, 1(1)
- [8] Bhat, I. K., Mohd, M., & Hashmy, R. (2018). SumItUp: A Hybrid Single Document Text Summarizer. In Soft Computing: Theories and Applications (pp. 619-634). Springer, Singapore
- [9] Sadiq, A. T., Chawishly, S. A., & Muhamad, K. S. (2009). Text Summarization Using Hybrid Methods. In First Conference of Iraqi IT Association.
- [10] <http://mlg.ucd.ie/datasets/bbc.html>
- [11] Meena, Y.K., Gopalani, D.: Evolutionary algorithms for extractive automatic text summarization. Procedia Comput. Sci. 48, 244–249 (2015)
- [12] Ammar, A., & AYDIN, T. A Comparison of Performance Metrics of Turkish Twitter Messages Using Text Representations
- [13] Karcioğlu, A. A., & Aydin, T. (2019, April). Sentiment Analysis of Turkish and English Twitter Feeds Using Word2Vec Model. In 2019 27th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE. doi: 10.1109/SIU.2019.8806295