

TÜRKÇE DESTEKTLİ ANLAMSAL ÇÖZÜMLEME VE BELGE SINIFLANDIRMA İÇİN ENDÜSTRİYEL BİR UYGULAMA

Projenin Tanımı

Önemi bilinmeyen veya bilindiği halde elindeki metinleri okumaktan başka bir çözüm yolu bulamayan veri sahiplerine metin madenciliği teknikleri kullanılarak hızlı ve verimli analiz yapılmıştır. Kullanıcı veri içerisinde daha önceden keşfedilmesi zor olan anlamsal öbeklemeleri, verinin hangi konseptler altında toplandığını, verinin toplandığı kitlenin sorunlarına veya ihtiyaçlarına ulaşmıştır. Bu bilgiler kullanıcının önüne doğrudan serilmemiştir, anlaşılabilirliği arttırmak açısından görselleştirilmiştir.

Videa Kimdir?

2017 yılında İzmir Dokuz Eylül Üniversitesi Tınaztepe kampüsü içerisindeki DEPART Beta binası bünyesinde Alper Çalapkulu tarafından kurulmuş ve varlığını devam ettiren genç bir AR-GE şirkettir. Metin madenciliği üzerine çalışmalar yapmak üzere kurulmuştur. Videa'nın müşterileri sosyal sorumluluk projeleri, belediyeler, şirketler olabileceği gibi toplumun her kesiminden birey olabilir.

AKADEMİK DANIŞMAN:

Dr. Öğr. Üyesi Mete EMİNAĞAOĞLU

PROJE EKİBİ:

Emre YILDIRIM, Eray İÇÖZ

Türkçe kaynaklı toplanan belgelerin anlamsal incelenmesi ve analiz edilmesi sonrasında da sınıflandırılması amaçlanmıştır. Kullanıcılardan elde edilen metin verisi (fikir) analiz edilip sınıflandırmaya uygun hale getirilecektir. Kümeleme algoritmalarıyla kümelenecek, konsept ataması yapılacaktır. Konseptte uygun şekilde veri görselleştirme teknikleriyle görselleştirilecektir.

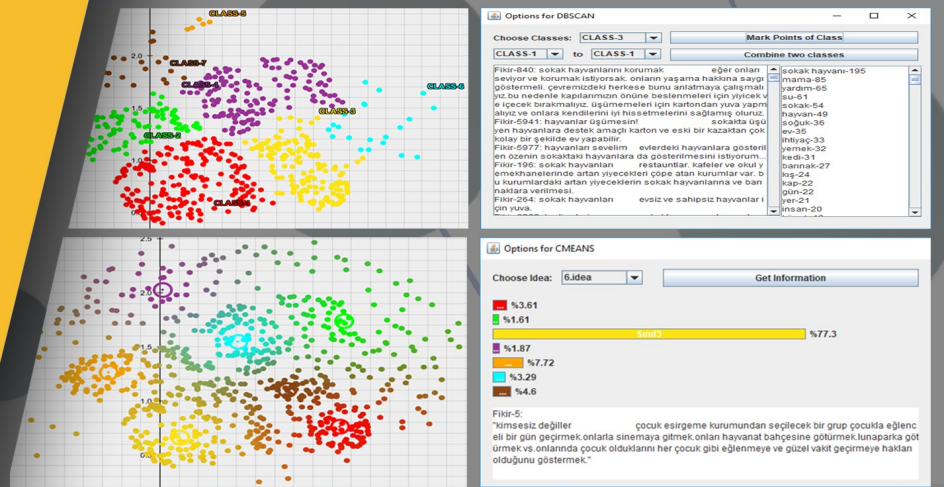
ANALİZ

Videa'nın mobil uygulaması içerisinde sahip olduğu kullanıcı popülasyonundan metin tabanlı veri alınmıştır. Alınan veri temizlenmiş ve analiz işlemlerinden (gövde, kelime tipi) geçirilmiştir. Veri iki boyutlu uzayda temsil edilebilmesi için Çok Boyutlu Ölçekleme tekniği (MDS) kullanılmıştır. DBSCAN ve fuzzy c-means kümeleme algoritmalarıyla kümeleme yapılmaya çalışılmıştır. Derlemin (corpus) içerisinde barındırdığı anahtar kelimeler tespit edilmiş bu anahtar kelimelerden konsept yapısı elde edilmiştir.

SONUÇ VE DEĞERLENDİRME

Proje sonunda, saf verinin analiz sonucu, kümeleme algoritmalarının paket hali, kümeleme algoritmalarından elde edilen sınıflandırılmış metinler, sınıflandırma hiyerarşisinin görselleştirildiği baloncuk grafiği (bubble chart), kelimelerin frekansını gösteren kelime bulutu (word cloud) ve kelimelerin birliktelik ilişkisini gösteren ağ analizi (network analysis) kullanıcıya sunulmuştur.

TASARIM VE UYGULAMA



HAYNAKÇA

Turgay Tugay BİLGİN, Yılmaz ÇAMURCU Maltepe Üniversitesi, Marmara Üniversitesi Dbscan, Optics Ve K-Means Kümeleme Algoritmalarının Uygulamalı Karşılaştırılması (2005). Ümit Can, Bilal Alatas Munzur Üniversitesi, Fırat Üniversitesi Duygu Analizi ve Fikir Madenciliği Algoritmaların İncelenmesi (2017). Bora ENGİN Microsoft Clustering White Paper, Univera Bilgisayar Sistemleri. (2009). Meltem IŞIK, A. Yılmaz ÇAMURCU, K-Means, K-Medoids Ve

