X (Twitter) Veri Analizi ve Kullanıcı İlişkilendirmesi

Berat Ölmez, Sait Ömer San 210201074, 220201001 Bilgisayar Mühendisliği Bölümü Kocaeli Üniversitesi Kocaeli/Türkiye beratolmez123@hotmail.com, sainsaitömer@gmail.com

Özetçe—Bu proje, Twitter kullanıcılarının verilerini analiz etmeyi ve kullanıcılar arasında benzer ilgi alanlarına göre eşleştirme yapmayı amaçlamaktadır. Kullanıcı verilerinin hash tabloları ile organize edilmesi, arama algoritmalarının kullanımı, gerçek veri analizi ve ilişkisel çıkarımlar, JSON dosyasını okuma ve organize etme gibi birçok kavramın kazanımını ve proje üzerinde ilişkilendirilmesini konu alır.

Anahtar Kelimeler — Veri Yapıları, Graf Teorisi, Hash Tabloları, Arama Algoritması, API Entegrasyonu, Veri Analizi, Node (Düğüm), Edge (Kenar), Hashing, BFS, Minimum Spanning Tree

I. Giris

Bu proje, Twitter API aracılığıyla kullanıcı verilerini çekmeyi ve elde edilen verileri analiz ederek benzer ilgi alanlarına sahip kullanıcıları eşleştirmeyi amaçlamaktadır. Kullanıcı verileri, hash tabloları ve etkili arama algoritmaları kullanılarak düzenlenerek, verimli bir şekilde ilgi alanlarına göre eşleştirmeler yapılması hedeflenmektedir. Bu sayede projenin temel amacı, Twitter kullanıcılarını benzer ilgi alanlarına sahip gruplara yönlendirmek ve etkilesimlerini artırmaktır.

II. YÖNTEM

A. Veri Toplama ve İşleme

Projemizde kullanıcıların sosyal medya platformlarında paylaştıkları tweetler üzerinde analiz yapmak üzere veri toplama ve işleme adımları gerçekleştirilmiştir. Kullanıcı verileri, belirli bir json uzantılı dosyadan çekilmiştir.

1. Tweet Verilerinin Toplanması

Kullanıcı verileri her bir kullanıcının sosyal medya hesaplarından paylaştığı tweetler alınmıştır. Bu tweetler, kullanıcıların profil bilgileriyle ilişkilendirilmiştir.

2. Tweet İçeriklerinin Analizi

Her bir tweetin içeriği, doğal dil işleme (NLP) teknikleri kullanılarak analiz edilmiştir. Analiz sürecinde, tweetlerde geçen kelimeler ve hashtag'ler belirlenmiş, ilgi alanlarını belirlemek üzere kullanılmıştır.

B. İlgi Alanı Analizi

Projemizde kullanıcıların ilgi alanlarını belirlemek amacıyla tweet içerikleri üzerinde bir ilgi analizi gerçekleştirilmiştir. İlgi analizi, belirli anahtar kelimelerin ve hashtag'lerin tweet içeriklerindeki sıklığını inceleyerek kullanıcıların ilgi alanlarını belirlemektedir.

1. Anahtar Kelimelerin ve Hashtag'lerin Belirlenmesi

Projede kullanıcıların ilgi alanlarını belirlemek üzere belirli anahtar kelimeler ve hashtag'ler tanımlanmıştır. Bu tanımlanan kelimeler ve hashtag'ler, ilgi alanı analizinde kullanılmıştır.

2. Ilgi Analizi Raporlarının Oluşturulması

Kullanıcıların ilgi alanlarına dayalı detaylı analiz raporları, belirlenen anahtar kelimeler ve hashtag'ler üzerinden oluşturulmuş ve metin dosyalarına kaydedilmiştir.

C. Kullanıcı İlişki Analizi

Kullanıcı ilişki analizi, belirli iki kullanıcı arasındaki ortak ilgi alanlarına ve takipçi ilişkilerine dayanmaktadır. Bu analiz, DFS (Depth-First Search) ve BFS (Breadth-First Search) algoritmalarını içermektedir.

1. Ortak Ilgi Alanına Sahip Kullanıcılar

Belirli iki kullanıcı arasında ortak ilgi alanına sahip kullanıcılar belirlenmiştir. DFS algoritması, kullanıcıların tweet içeriklerindeki anahtar kelimeler ve hashtag'ler üzerinden bu ilişkiyi tespit etmek üzere kullanılmıştır.

2. Kullanıcılar Arasındaki Bağlantıların Kurulması

BFS algoritması kullanılarak, belirli iki kullanıcının takipçi listeleri üzerinden ilgi alanlarına göre bağlantılar kurulmuştur. Minimum Spanning Tree (MST) algoritması kullanılarak, benzer ilgi alanlarına sahip kullanıcılar arasında en az bağlantı ile ağaç oluşturulmuştur.

Kodun Çalışma Mantığı

Programda bulunan okujson adındaki python dosyasının içinde bulunan dosya_oku_ve_yazdir fonksiyonu yardımıyla json dosyasından kullanıcıların özellikleri çekilmektedir. Bu özellikler liste_oluştur fonksiyonu yardımıyla User sınıfından oluşturulan kullanıcı nesnelerinde tutulmaktadır. Programda liste_olustur fonksiyonu BagliListe ismindeki linked list yapısına kullanıcıları liste olarak eklemektedir. liste_olustur fonksiyonunda oluşturulan HashTable sayesinde kullanıcılara proje isterlerinde bulunan özellikler verilmektedir. Projede

hazır olmayan BagliListe ve HashTable sınıfları sık sık kullanılmıştır. Kullanıcıların okunma işlemini sağlayan yapılar tamamlandıktan sonra dosyasından okunan kullanıcıları eklemek için BagliListe isimli sınıf kullanılarak graf isimli sınıf oluşturulmuştur. Her kullanıcı graf isimli sınıfa node olarak eklenmistir. Bu sınıfta bulunan kullanicilariiliskilendir fonksiyonu kullanılarak takip-takipci ilişkisi bulunan kullanıcılar arası grafa edge eklenmiştir. Bu fonksiyonlar kullanılarak hazır graf kullanılmadan kullanıcılar grafa eklenmiştir. Proje isterlerinde bulunan DFS(Depth First Search) ve BFS(Breadth Firtst Search) algoritmalarının BagliListe hazırlanması için programa kullanılarak Stack ve Queue sınıfları eklenmiştir. Stack veri yapısı DFS algoritmasında kullanılmıştır. Bu algoritma graf üzerinde bulunan düğümleri derinlemesine gezmektedir. Queue veri yapısı ise BFS algoritmasında kullanılmıştır. BFS algoritması kullandığı kuyruk yapısı sayesinde graf üzerinde bulunan komşuları önce gezerek tüm düğümleri gezmektedir. Programda takipci sayısi ortak olan kullanıcılar ortak takipci sayisi bfs fonksiyonuyla metin dosyasına yazdırılmaktadır. Bu fonksiyon parametre olarak graf ve başlangıç düğümü almaktadır. Başlangıç düğümünden başlayıp bütün düğümleri BFS algoritmasıyla gezerek başlangıç düğümünün takipçi sayısıyla aynı olan düğümleri listelemekte ve bu düğümlerdeki kullanıcıların kullanıcı isimlerini metin dosyasına yazmaktadır. Programda ortak dil dfs isimli fonksiyon DFS algoritmasını kullanarak graf üzerinde bulunan ortak dile sahip kullanıcıları metin dosyasına yazmaktadır. bulma asamasında alanını kullanılmıştır: Birinci yöntemde belirli ilgi alanlarını kapsayan kelimelerin kullanıcı tweet'leri içinde olup olmadığı kontrol edilerek ilgi alanı bulunmuştur. Bu yöntem AnalyzeTweets içerisindeki alanlyze tweets fonksiyonu kullanılarak uygulanmıştır. analyze_tweets fonksiyonu NTLK kütüphanesini kullanarak tweet içerisindeki anlamı olan kelimeleri köklerine göre ayırmaktadır ve bu ayırdığı kelimelerde ilgi alanını aramaktadır. İkinci yöntemde ise verilen ilgi alanini bul fonksiyonu yardımıyla arama yoluyla ilgi alanı bulunmuştur. alanları metin dosyasına ilgi alani txt fonksiyonu kullanılarak yazdırılmıştır. Programda olusturulmus olan User sınıfında bulunan dfs tweet ilgialani fonksiyonu DFS algoritmasını kullanarak kullanıcıların ilgi alanının belirlenmesini

sağlayan tweet gönderimlerini metin dosyasına yazdırmaktadır. Kod çıktı olarak 8 adet txt dosyasına proje isterlerinde bulunan analizleri yazdırmaktadır. Ayrıca networkx kütüphanesini kullanarak seçilen kullanıcının takip takipçi ilişkilerine göre graf çizmektedir.

III. DENEYSEL SONUÇLAR

1. İşlemin süreci

| The properties | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Property | Proper

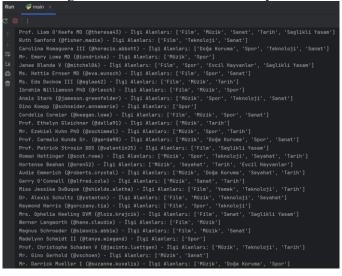
Takipçi-Takip ilişkilerinin dosyaya yazdırılması
 takipcitakipiliskileri - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

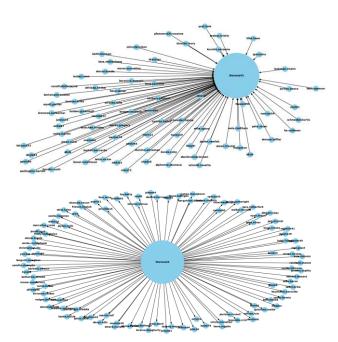
Kullanıcı: theresa43 Takip Ettikleri:

- hattie.hickle
- calista40
- jreichel
- layne.luettgen
- mheller
- nyah.braun
- alvina.russel
- orpha.ziemann
- alfreda38
- hane.rogelio
- orrin24
- toy.betty
- nader.fermin
- fahey.aditya
- jayme.mills
- estelle62

3. İlgi alanlarının belirlenmesi ve ekrana yansıtılması



4. İlişkilerin graf olarak gösterilmesi



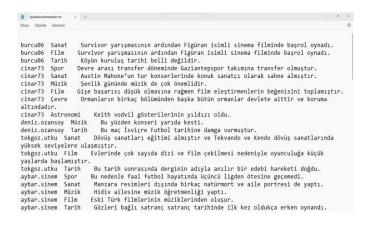
5. Bölgeye göre arama sonuçlarının dosyaya yazdırılması



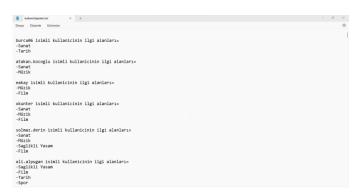
Dosya Düzen Biçim Görünüm Yardım

Aranan Bölge=TR rlesch'nin ortak ilgi alanları:

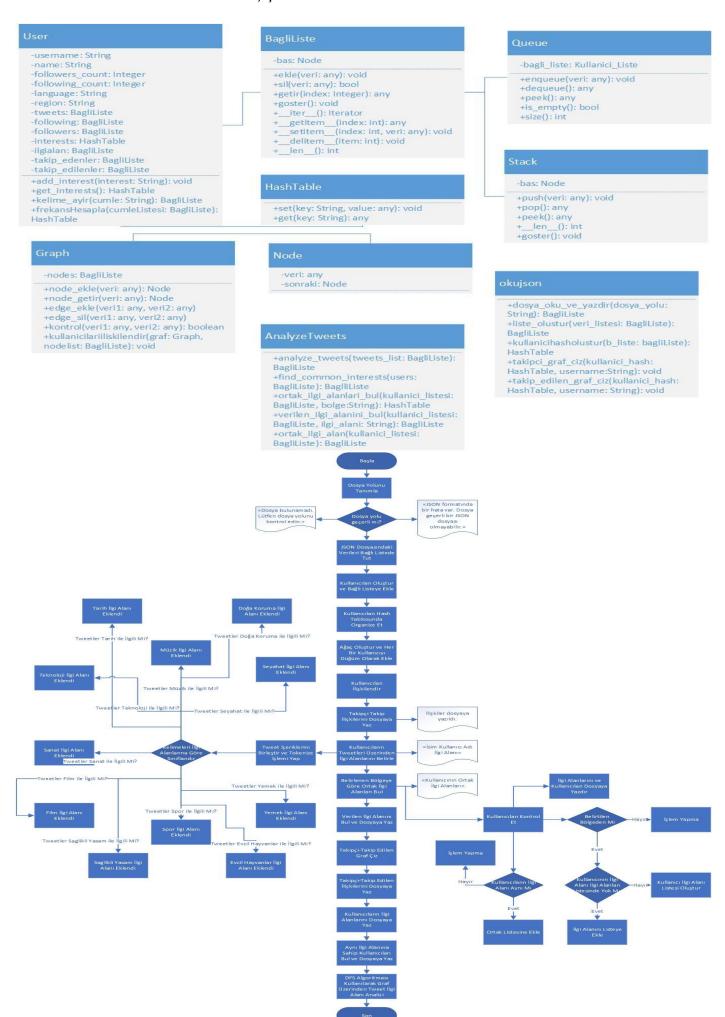
- Müzik
- Film
- Spor
 - 6. İlgi alanı seçilen tweet gönderimleri



7. Kullanıcıların ilgi alanları



IV. UML DİYAGRAMI VE AKIŞ ŞEMASI



V. KATKILAR

- Sait Ömer San:
- JSON dosyası okuma ve verileri nesne yapılarında tutma
- Hash table, linked list ve stack yapılarını tanımlama ve verileri bu yapılarda organize etme
- User yapısını düzenleme
- Graf oluşturma ve çizdirme
- Graf analizi, analiz için BFS algoritması kullanma ve analiz sonucu çıktı oluşturma
- Belirli bölge (konum) için kullanıcıların ilişkilendirilmesi
- Berat Ölmez:
- User yapısını düzenleme
- Doğal dil işleme kütüphaneleri kullanarak tweet analizi
- Analiz sonucu ilgi alanı bulma ve organize etme
- Analiz sonucu verileri eşleştirme

VI. SONUÇ

Bu çalışma, farklı tipteki verilerin analiz ve organizasyonlarının etkili ve verimli kullanılması adına örnek teşkil etmektedir. Kullanıcı verilerinin birbiri ile etkileşimini ve eşleştirmesini sağlamak ve çeşitli algoritmalar ile desteklenerek veriler üzerinde çalışmanın bir çok yöntemi öğrenilmiş, metotları uygulanmıştır.

KAYNAKLAR

- [1] www.medium.com
- [2] https://bilisim.io
- [3] dergipark.org.tr
- [4] https://www.simplilearn.com/tutorials
- [5] https://www.sadikturan.com
- [6] https://github.com