

KOCAELİ ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
PROGRAMLAMA LABORATUVARI I
PROJE 3

Proje Teslim Tarihi: 29.12.2023

X (Twitter) Veri Analizi ve Kullanıcı İlişkilendirmesi



Proje Hedefi: Bu proje, Twitter API aracılığıyla kullanıcı verilerini çekerek bu verileri analiz etmeyi ve kullanıcılar arasında benzer ilgi alanlarına göre eşleştirmeler yapmayı hedeflemektedir. Bununla birlikte kullanıcı verilerini, hash tablolarıyla organize ederek ve arama algoritmalarını kullanarak verimli bir şekilde ilgi alanlarına göre eşleştirme yapmayı amaçlamaktadır.

Kazanım: Bu proje, öğrencilere veri yapıları, graf teorisi, hash tabloları, arama algoritmaları, API entegrasyonu, veri analizi ve ilişki yönetimi gibi konularda deneyim kazanma fırsatı sunacaktır. Hem veri yapıları ve algoritmaların hem de gerçek veri analizi ve ilişki çıkarımların pratiğini yaparak problem çözme ve analitik düşünme becerilerini geliştirmeye fayda sağlayacaktır.

Proje istekleri:

❖ Veri Kaynağı

- Twitter API, JSON dosyaları veya benzer bir platform aracılığıyla kullanıcı verileri çekilecektir.
- Girdi ve çıktı işlemleri için veritabanı kullanılmayacaktır. Bu işlemler düzenli bir formatta dosyalardan okunmalı, aynı şekilde çıktılar da dosyalara kaydedilmelidir.

❖ Kullanıcı

- Her kullanıcı aşağıdaki bilgileri içeren bir nesne olarak tanımlanmalıdır.
- Kullanıcı Adı, Ad-Soyad, Takipçi Sayısı, Takip Edilen Sayısı, Dil ve Bölge.
- Tweet İçerikleri: Kullanıcının paylaştığı tweetler.
- Takip Edilenler ve Takipçiler: Kullanıcının takip ettiği diğer kullanıcılar ve kendisini takip edenler.

❖ Veri Çekme ve Nesne Oluşturma

- Kullanıcılara ait veriler, Twitter API kullanılarak çekilmelidir.

- Kullanıcıların her biri için bir nesne oluşturulmalı ve bu kullanıcı nesneleri Hash tabloları ile organize edilmelidir.
- Tüm işlemler nesneler üzerinden yapılmalıdır. Nesneler başka nesneler içermelidir.
- Veri yapılarıyla ilgili tüm işlemlere ait algoritmaların kodu öğrenci tarafından yazılmalıdır. Veri yapıları için hazır yapıların kullanımı yasaktır.

❖ Kullanıcı İlişkilerinin Graf Olarak Modellenmesi

- Kullanıcı nesneleri arasındaki takipçi-takip edilen ilişkilerini kullanarak bir graf oluşturulmalıdır.
- Her kullanıcı bir düğüm (node) ve takipçi-takip edilen ilişkileri de kenarlar (edge) olarak temsil edilmelidir.

❖ İlgi Alanına Göre Eşleme

- Kullanıcı ilgi alanlarını hash tablolarında tutulmalıdır. Her bir ilgi alanı için bir hash tablosu oluşturulmalı ve ilgi alanını anahtar olarak kullanarak bu ilgi alanını paylaşan kullanıcıların listesini değer olarak saklanmalıdır.
- Kullanıcıların takip ettikleri, takipçileri ve tweet içerikleri arasında benzerlikleri tespit ederek, ortak ilgi alanlarına sahip kullanıcıları eşleştirilmelidir.
- Verilen bir ilgi alanı arama algoritmaları (hashing vb.) kullanılarak bulunmalıdır. Ayrıca bu ilgi alanını paylaşan kullanıcıların listesi çıkarılmalıdır.
- Kullanıcıların ortak ilgi alanlarını belirlemek için hash tabloları ve karşılaştırma algoritmaları (arama, sıralama algoritmaları vb.) kullanılmalıdır.
- Hash tabloları ve arama algoritmaları kullanılarak elde edilen benzer ilgi alanlarına sahip kullanıcılar arasındaki ilişkiler gösterilmelidir.
- Kullanıcıların ilgi alanlarına dayalı detaylı analiz raporları oluşturularak metin dosyasına kaydedilmelidir.
- Belirli bölge (konum) ve dil için trend olan hashtagler veya konular listelenmelidir.
- Kullanıcıların tweet içeriklerinde belirli anahtar kelimeleri ve hashtagleri içeren tweetler DFS algoritması kullanarak listelenmelidir.
- Belirli iki kullanıcının takipçilerinden ilgi alanına göre filtreleme yapılarak ortak ilgi alanına sahip kullanıcılar listelenmelidir:
 - Belirli iki kullanıcının takipçi listeleri kullanarak ilgi alanlarına göre bağlantılar kurulmalıdır.
 - BFS algoritması kullanarak, ilgi alanlarının benzerliği temel alınarak kullanıcılar arasında bir bağlantı olup olmadığı belirlenmelidir.
 - Minimum Spanning Tree algoritması kullanarak, benzer ilgi alanlarına sahip kullanıcılar arasında en az bağlantı ile ağaç oluşturulmalıdır.
 - Oluşturulan ağaç, ortak ilgi alanlarına sahip kullanıcıları ve bu kullanıcılar arasındaki ilişkileri temsil etmelidir.

❖ Graf Üzerinde Analiz

- Graf algoritmaları kullanarak kullanıcılar arasındaki bağlantıları analiz edin.
- Belirli bir kullanıcıdan başlayarak, o kullanıcıyla aynı seviyede (mesela belirli bir takipçi sayısına sahip) olan diğer kullanıcıları tespit etmek için BFS algoritmasını kullanabilirsiniz.
- Graf analizleri sonucunda elde edilen kullanıcı toplulukları ve bu toplulukların özellikleri çıktı olarak gösterilmelidir.

Ödev Teslimi:

- ❖ Ödevin raporu **LaTeX** veya **Word** kullanılarak yazılmalıdır.
- ❖ Proje grupları, her öğretim sadece kendi içerisinde olacak şekilde en fazla 2 kişilik oluşturulabilir.
- ❖ Proje raporunda aşağıdakiler dikkate alınmalıdır:
 - Doküman IEEE formatında en az 4 sayfa, Times New Roman yazı fontu ile, iki yana yaslı yazım formatında hazırlanmalıdır.
 - Özet, giriş, yöntem, deneysel sonuçlar, sonuç, yazar katkıları ve kaynakça bölümlerinden oluşmalıdır.
 - Problemin tanımı, problemin çözümü için yapılan araştırmalar, projenin kodlanması için kullanılacak kod yapısı ve geliştirme aşamaları anlatılmalıdır.
 - Yazar katkıları bölümünde her bir grup üyesinin projeye katkıları detaylı bir şekilde belirtilmelidir.
 - **Akış şeması ve UML diyagramı bulunmalıdır.**
- ❖ Dersin takibi projenin teslimi dahil edestek2.kocaeli.edu.tr sistemi üzerinden yapılacaktır. edestek2.kocaeli.edu.tr sitesinde belirtilen tarihten sonra getirilen projeler kabul edilmeyecektir.
- ❖ Proje ile ilgili sorular edestek2.kocaeli.edu.tr sitesindeki forum üzerinden **Arş. Gör. Dilara Görmez Açık** veya **Arş.Gör. Kübra Erat**'a sorulabilir.
- ❖ Sunum tarihleri daha sonra duyurulacaktır.
- ❖ Sunum sırasında kullandığınız herhangi bir satır kodu açıklamanız istenebilir.
- ❖ Göndermiş olduğunuz kodlar benzerlik testinden geçirilecektir. Benzerlik çıkması halinde aşağıdaki puanlama uygulanacaktır.
 - Gruplar arası benzer çıkma (kod parçası alma) projeden -(eksi)50 puan
 - Para karşılığı başka birine yaptırma projeden -(eksi)100 puan
 - İnternette kodu (kod parçası) alma projeden -(eksi)50 puan