

WEB TRAFİK LOGLARINA DAYALI YAPAY ZEKA DESTEKLI SORU-CEVAP SISTEMI

SILA AKKAYA

Ağustos ————— 2024

PROJE AMACI

Bu projede, bir web sitesinin trafik loglarını kullanarak bir yapay zeka destekli soru-cevap (Q&A) sistemi geliştirilmiştir. Sistem, kullanıcılardan gelen doğal dildeki soruları alarak, ilgili log verilerini analiz eder ve bu verilerle uygun yanıtları oluşturur. Temel olarak Retrieval-Augmented Generation (RAG) modeli kullanılmıştır.

KULLANILAN YÖNTEMLER

- Veri Hazırlığı: Web trafik loglarının işlenmesi ve vektörleştirilmesi.
- Bilgi Alma: FAISS vektör veri tabanı kullanılarak bilgi retrieval işlemleri.
- Jeneratif Model: T5 modeli kullanılarak kullanıcı sorularına yanıt üretimi.

• VERİ HAZIRLIĞI VE ÖN İŞLEME

Log Dosyasının Analizi

Web trafik logları, IP adresleri, zaman damgaları, HTTP yöntemleri, erişilen sayfalar, durum kodları ve yanıt boyutları gibi bilgileri içermektedir. Bu veriler, log dosyasından ayrıştırılmış ve işlenmiştir.

Veri Temizleme ve Yapılandırma

Log dosyasındaki gereksiz veya bozuk kayıtlar temizlenmiş ve veri yapısı düzenlenmiştir. IP adresleri, sayfalar, zaman damgaları ve diğer bilgiler uygun formatta saklanmıştır.

Vektörleştirme

Log verileri, sentence-transformers kullanılarak vektörlere dönüştürülmüş ve FAISS vektör veri tabanına eklenmiştir.

• RAG MODELİNİN KURULUMU

Bilgi Alma (Retrieval)

FAISS kullanılarak kullanıcı sorusuna en uygun log kayıtları bulunur. Bu işlem, kullanıcı sorusunun vektörünün, vektör veri tabanındaki en yakın vektörlerle karşılaştırılması yoluyla gerçekleştirilir.

Jeneratif Model (Generation)

T5 modeli kullanılarak, bilgi alma aşamasında elde edilen log kayıtlarıyla birlikte kullanıcı sorusuna uygun yanıtlar üretilir.

• SİSTEM ENTEGRASYONU VE TEST

Sistem Entegrasyonu

Bilgi alma ve jeneratif model bileşenleri entegre edilmiştir. Kullanıcıdan gelen bir soru, önce FAISS ile bilgi alma aşamasına gönderilir, ardından bu bilgilerle T5 modelinden yanıt üretilir.

Test Senaryoları

Test senaryoları, sistemin doğruluğunu ve performansını değerlendirmek için oluşturulmuştur.

KARŞILAŞILAN SORUNLAR VE ÇÖZÜMLER

Sorunlar

- Veri Temizleme: Log verilerinin temizlenmesi ve yapılandırılması sırasında karşılaşılan zorluklar.
- Model Performansı: T5 modelinin yanıt üretme sürecindeki zorluklar.

Çözümler

- Veri Temizleme: Veri temizleme yöntemlerinin iyileştirilmesi.
- Model Performansı: Model hiperparametrelerinin ayarlanması ve iyileştirmeler yapılması.