**Proje Raporu**

**Projeyi teslim alma tarihi: 19.08.2024**

**Proje bitiş tarihi : 21.08.2024**

**Proje adı:** Web trafik loglarına dayalı yapay zeka destekli soru-cevap sistemi

**1. Giriş**

Bu proje kapsamında, web trafik loglarına dayalı bir yapay zeka destekli soru-cevap sistemi geliştirdim. Amaç, kullanıcılardan gelen doğal dildeki soruları analiz ederek uygun log kayıtlarını bularak ve bu verileri kullanarak anlamı olan yanıtlar oluşturmak. Projede Retrieval-Augmented Generation (RAG) modeli kullanıldı, log verilerinin vektör tabanlı bir veri tabanında saklanması ve bu verilere dayalı yanıtların oluşturuldu.

**2. Gelişmeler ve Süreç**

* **Veri Hazırlığı ve Ön İşleme (19.08.2024)**:
  + Log dosyası üretildi ve, gerekli alanlar (IP adresi, zaman damgası, HTTP yöntemi, URL, durum kodu) seçildi ve bu üretimin ardından hatlı veya yanlış veriler temizlendi.
  + Veriler TF-IDF kullanılarak vektörlere dönüştürüldü ve HNSWLib tabanında depolandı.
* **RAG Modelinin Kurulumu (19.08.2024 )**:
  + Bilgi alma (retrieval) kısmı TF-IDF ve HNSW kullanılarak gerçekleştirildi. Kullanıcı sorularına en uygun log kayıtları başarıyla bulunup geri döndürüldü.
  + TF-IDF ve HNSW kullanılma sebebi ise büyük bir data havuzundan en doğru sonuca ulaşmamı sağlaması adına en faydalı yollardan biri olması TF-IDF ve HNSW kullanmama yol açtı
  + Kullanıcının sorduğu sorular, dinamik olarak alınarak vektör tabanında arandı ve en yakın sonuçlar geri döndürüldü.
* **Almanya uçuşum ve ulaşım süreci (20.08.2024)**:
* **Sistem Entegrasyonu ve Test (21.08.2024)**:
  + Bilgi alma sistemi tamamen entegre edildi.
  + Sistem, yaptığım testlerden geçmiş ve hepsinde çok başarılı oldu
  + Belirli sorgular ("first" veya "last" gibi) için loglar tarih sıralamasına göre sıralanmış ve en eski veya en yeni 5 log, kullanıcıya doğru bir şekilde sunuldu.
  + Kullanıcıdan gelen sorulara verilen cevapların doğruluğu test edildi. Sorular aynı zamanda dinamik olarak alındı ve yanıtlar kriterlere göre döndürüldü.
  + Bu test ve optimizasyon işlemi en çok kafamı karıştıran ve işin içinden çıkmakta en çok zorlandığım kısımdı
* **Raporun hazırlanması (21.08.2024):**
  + Raporun hazırlanması oldukça kolaydı ve çok zaman almadı

**3. Karşılaşılan Zorluklar**

* **Vektörizasyon ve Performans**: Vektör veri tabanı ile çalışırken orta boyutlu verilerle başa çıkmak dahi zor oldu. Performans optimizasyonları için vektör veri tabanının ayarlanması gerçekten çok zaman aldı
* **RAG Modelinin Entegrasyonu**: Bilgi alma ve jeneratif modelin entegrasyonu sırasında, dil modeli tarafından anlamlı ve doğru yanıtlar üretilmesini sağlamak zordu. Dil modelinin ilgili log verilerini doğru yorumlaması ve kullanıcı sorularıyla ilişkilendirmesi beni en zorlayan ve araştırırken en çok zaman verdiğim bölümdü.

**4.** **Projede kullanılan tekniklerin listesi:**

* Data Generation - Random Data Generation
* Data Preprocessing - CSV File Handling, Text Concatenation
* Text Vectorization - TF-IDF (Term Frequency-Inverse Document Frequency)
* Query Expansion - Synonym Expansion using WordNet
* Stop Words Removal - Filtering out common stop words
* Similarity Search and Indexing - HNSW (Hierarchical Navigable Small World)
* Log Retrieval and Filtering - Threshold-based Filtering, Temporal Sorting (First/Last)
* Interactive Query System - Command-line Interface for User Queries
* Dimensionality Reduction - Reducing Vector Dimensions (if applied)
* Error Handling and Debugging - Interactive Debugging, Exception Handling in Code

**5. Sonuç**

Proje genel olarak başarılı bir şekilde tamamlandı ve sistem, kullanıcıdan gelen sorulara anlamlı ve doğru yanıtlar üretebilecek şekilde yapıldı. Ancak, sistemin doğruluğunu ve performansını daha da artırmak için ek optimizasyonlar ve iyileştirmeler yapılabilir.

Bu rapor, projenin temel süreçlerini ve karşılaşılan zorlukları özetlemektedir. Proje süresi toplamda 2 gün sürmüş olup, her aşama hızlı bir şekilde yazılmıştır bunun en büyük nedeni ise projeyi teslim aldığım günden 1 gün sonra Almanya uçuşum olması ve bu süreçte kod yazamamam. Değerlendirmemin buna göre de yapılmasını rica ederim.

Umut Uygur

Contact: [umuttuygurr@gmail.com](mailto:umuttuygurr@gmail.com)

https://www.linkedin.com/in/umut-uygur-24b920249/