

< Doğal dil işleme
senaryosu >

TÜRKÇE DOĞAL DİL İŞLEME

8 - 9 AĞUSTOS 2024

< MINERVA'S LEGACY >

<EKİBİMİZ>

DR. AYŞEN KÜTAHYALIOĞLU (Danışman)

Takımımızın danışman öğretmenidir. Matematik alanında uzman bir akademisyendir ve projemizin her aşamasında bize rehberlik etmektedir.

MUSTAFA AK

Yazım ve dilbilgisi hatalarının tespiti ve düzeltilmesiyle ilgilenir. Türk dili ve edebiyatı mezunudur.

ECE AVCI

Anlam ve sentiment analizlerini yapar. Veri Analizlerinden sorumludur.

ELFİN EKİCİ

Kullanıcı Arayüzünün Tasarımını yapar yaratıcı ve kullanıcı deneyimi konusunda sorumludur.

ENES KAAN (Yazılım Geliştirici)

API ve kullanıcı arayüzünü geliştirir. Python, Java ve C dillerinde deneyimlidir. Small basic HTML ve C++ yazılım dillerinde kapsamlı bilgi ve deneyime sahiptir

BERRA EKİCİ (Kaptan)

Veri toplama ve temizleme süreçlerinden sorumlu. Detay odaklıdır. Liderlik yetenekleriyle ekibimizi yönlendirmektedir. Proje boyunca stratejik kararlar alarak, ekip uyumunu ve verimliliğini sağlamaktadır.

<PROJENİN TANIMI>

Bu proje, Türkçemizin doğru ve etkin kullanımını teşvik etmek amacıyla geliştirilmiştir. Sosyal medyanın ve dijital iletişim araçlarının yaygınlaşması, dilimize birçok yabancı kelimenin girmesine ve bu kelimelerin günlük dilde sıkça kullanılmasına yol açmıştır. Bu durum, Türkçenin saflığını ve zenginliğini tehdit etmekte, aynı zamanda dil bilincini zayıflatmaktadır. Dilimizin bu şekilde bozulması, kültürel kimliğimizin de zarar görmesine sebep olmaktadır. Proje, bu soruna çözüm bulmak amacıyla, yabancı kelimelerin Türkçe karşılıklarının sunulmasını ve bu kelimelerin metin içerisinde renkli gösterilmesini hedeflemektedir. Bu sayede, kullanıcıların yabancı kelimeleri fark etmesi ve yerine Türkçe karşılıklarını kullanmaları teşvik edilmektedir.

<PROJENİN SAĞLADIĞI ÇÖZÜM>

Projemiz, kullanıcıların yazdıkları metinlerdeki yabancı kelimeleri tespit eden ve bu kelimelerin Türkçe karşılıklarını öneren bir sistem geliştirmiştir. Bu sistem, kullanıcıların yazım sürecinde yabancı kelimeleri daha kolay fark etmelerini ve Türkçe karşılıklarını kullanmalarını sağlamaktadır. Örneğin, „trend“ kelimesi kullanıldığında, sistem bu kelimeyi tespit ederek kullanıcıya „eğilim“ kelimesini önermektedir.

Bu çözümden faydalanan kitle oldukça geniştir. Özellikle sosyal medya kullanıcıları, yazarlar, öğretmenler ve öğrenciler bu sistemden büyük ölçüde faydalanabilirler. Ayrıca, iş dünyasında resmi yazışmalar yapan kişiler, dil hassasiyeti olan bireyler ve dil eğitimi veren kurumlar da bu sistemden yararlanabilir.

<PROJE İŞ AKIŞI>

1. Proje Adımları:

- Veri Toplama ve Temizleme
- Model Eğitimi
- Yazım ve Dilbilgisi Hatalarının Tespiti ve Düzeltilmesi
- Anlam Analizi ve Sentiment Analizi
- API ve Kullanıcı Arayüzü Geliştirme
- Test ve Değerlendirme

2. Veritabanı ve Model Eğitimi (Opsiyonel): Bu aşamada, geniş bir veri seti toplayarak ve temizleyerek, NLP modellerini eğitebiliriz. Bu adım, daha gelişmiş bir sistem geliştirmek için gereklidir.

3. Test ve Değerlendirme: API'yi test etmek için bir HTTP istemcisi (örneğin Postman) veya curl komutunu kullanarak aşağıdaki şekilde test edebiliriz.

<VERİ SETİ>

Projede kullanılan veri seti, kullanıcı katkılarından toplanmıştır. Bu veri seti, yabancı kelimelerin ve onların Türkçe karşılıklarının geniş bir yelpazede temsil edilmesini sağlamaktadır. Veri seti oluşturulurken şu adımlar izlenmiştir:

1. Veri Toplama: Çeşitli kaynaklardan yabancı kelimeler ve Türkçe karşılıkları toplanmıştır.
2. Veri Temizleme: Toplanan veriler, gereksiz ve hatalı bilgilerden arındırılarak temizlenmiştir.
3. Veri Etiketleme: Yabancı kelimeler ve Türkçe karşılıkları etiketlenmiştir.
4. Veri Çeşitlendirme: Farklı alanlardan (sosyal medya, haber, akademik metinler) veri toplanarak, kelime kullanımındaki çeşitlilik sağlanmıştır.

<YÖNTEM VE TEKNİKLER>

Projede kullanılan başlıca algoritmalar, modeller ve doğal dil işleme teknikleri şunlardır:

- nltk (Natural Language Toolkit): Doğal dil işleme için temel araçlar sunar ve metin işleme, sınıflandırma gibi işlemler için kullanılmıştır.
- spacy: Hızlı ve doğru dil işleme modelleri sağlar. Metin içerisindeki yabancı kelimeleri tespit etmek için kullanılmıştır.
- scikit-learn: Makine öğrenimi modelleri ve metin sınıflandırma için kullanılmıştır.
- textblob: Metin işleme ve duygu analizi için basit bir arayüz sunar.
- transformers: Gelişmiş dil modelleri (örneğin BERT) kullanarak metin işleme işlemlerinde yüksek doğruluk sağlar.
- flask: Web tabanlı uygulama geliştirmek için kullanılmış ve kullanıcı arayüzü oluşturulmuştur.

<MODEL EĞİTİMİ VE DEĞERLENDİRME>

Model eğitimi, geniş bir veri seti üzerinde gerçekleştirilmiştir. Eğitilen model, yabancı kelimeleri tespit ederek Türkçe karşılıklarını önermektedir. Model eğitimi sırasında şu adımlar izlenmiştir:

1. Veri İşleme: Toplanan veri seti, eğitime uygun hale getirilmiştir. Bu aşamada, veriler normalleştirilmiş, etiketlenmiş ve özellik çıkarımı yapılmıştır.
2. Model Seçimi ve Eğitimi: En uygun model belirlenmiş ve veri seti üzerinde eğitilmiştir. Çeşitli makine öğrenimi algoritmaları (örneğin, Naive Bayes, SVM) ve dil modelleri (BERT, GPT) kullanılmıştır.
3. Model Değerlendirme: Modelin performansı doğruluk, hassasiyet, geri çağırma ve F1 skoru gibi metriklerle değerlendirilmiştir. Çapraz doğrulama yöntemi kullanılarak modelin genel performansı analiz edilmiştir.
4. Model İyileştirme: Geri bildirimler ve değerlendirme sonuçları doğrultusunda model iyileştirilmiştir.

<SONUÇLAR>

Proje kapsamında elde edilen bulgular, sosyal medyada ve diğer yazılı platformlarda yabancı kelime kullanımının azaltılabileceğini ve Türkçenin daha doğru kullanılabileceğini göstermiştir. Kullanıcılar, bu sistem sayesinde yazdıkları metinlerdeki yabancı kelimeleri fark ederek Türkçe karşılıklarını öğrenmekte ve kullanmaktadır. Bu da dil bilincinin artmasına ve dilimizin korunmasına katkı sağlamaktadır.

<PROJE YOL HARİTASI>

Gelecekte projenin daha da geliştirilmesi için şu adımlar planlanmaktadır:

1. Dil Modeli ve Veri Setinin Genişletilmesi:Daha fazla veri toplanarak ve mevcut modelin daha geniş bir dil yelpazesinde eğitilmesi sağlanarak modelin doğruluğu artırılacaktır.
2. Mobil Uygulama Entegrasyonu Kullanıcıların mobil cihazları üzerinden de bu sistemi kullanabilmesi için mobil uygulama geliştirilecektir.
3. Kullanıcı Deneyimi Araştırmaları: Kullanıcı deneyimi araştırmaları yapılarak geri bildirimler doğrultusunda sistemin kullanım kolaylığı artırılacaktır.
4. Farklı Dil Kombinasyonları İçin Çoklu Dil Desteği: Sistem, yalnızca Türkçe değil, diğer dillerde de kullanılabilecek şekilde genişletilecektir.
5. Eğitim Kurumları ile İş Birliği:Eğitim kurumları ve dil kursları ile iş birliği yapılarak, dil eğitimi veren yerlerde sistemin kullanımı teşvik edilecektir.

<DEMO VIDEO>

<https://www.powtoon.com/s/eXySkJ5VkCj/1/m/s>

TEŞEKKURLER