

1906003132015

Doğal Dil İşleme

BAİBÜ Bilgisayar Müh.

Dr. Araş. Gör. İsmail Hakkı Parlak

ismail.parlak@ibu.edu.tr

Oda: 329

Dersin Gereksinimleri

- Temel düzeyde
 - kalkulus
 - lineer cebir
 - olasılık & istatistik
- Python programlama

Dönem Boyunca...

- 3 adet programlama ödevi (%30)
- 1 Vize (%30)
- 1 Final (%40)
- Ödevler bireyseldir ve kod paylaşımı yapılamaz
- Fikirler tartışılıp paylaşılabılır ancak kod paylaşılamaz
- Kod paylaşımı durumunda «kopya çekme» olarak değerlendirilir ve cezası uygulanır

Kaynaklar

- Speech and Language Processing (3rd ed.), Dan Jurafsky & James H. Martin
- Natural Language Processing, Jacob Eisenstein
- Ders notları, sunumlar, lab çalışmaları...
- Dersin ibuzem/übys sayfası

Doğal Dil İşleme (Natural Language Processing - NLP)

- Doğal dil: Türkçe, İngilizce, Farsça, ...
- Bilinçli bir planlama veya önceden düşünme olmaksızın kullanım ve tekrarlama yoluyla insanlarda doğal olarak gelişen herhangi bir dildir.
- Doğal diller, konuşma veya işaret dili gibi farklı biçimler alabilir. Bilgisayarları programlamak için kullanılanlar gibi yapılandırılmış dillerden farklıdırlar.

Doğal Dil İşleme (Natural Language Processing - NLP)

- Doğal dil işleme, insan dilini bilgisayarlara erişilebilir kılmak için kullanılan yöntemler kümesidir.
- Otomatik makine çevirisi
- Metin sınıflandırması
- Spam filtreleri
- Arama motorları
- Siri, Alexa, Google Assistant
- Chat botları
- Çağrı merkezleri
- GitHub Copilot
- Yapay yazarlar

Makine Öğrenmesi (Machine Learning - ML)

- Makine öğrenimi (ML), 'öğrenen' yöntemleri, yani bazı görevlerde başarıyı artırmak için verilerden yararlanan yöntemleri anlamaya ve oluşturmaya adanmış bir araştırma alanıdır.
- Yapay zekanın (Artificial Intelligence - AI) bir parçasıdır.
- ML algoritmaları, açıkça programlanmadan tahminlerde bulunmak veya kararlar almak için eğitim verileri olarak bilinen örnek verilere dayalı bir **model** oluşturur.
- Neden klasik algoritmalar yerine ML?

Temel Tanımlar

- Cümle: Kendi başına anlamlı kelimeler dizisi. Çoğunlukla özne, tümleç ve yüklemden meydana gelir. Büyük harfle başlar; noktalama işareti ile biter.
- Simgе (Token): Kelime, noktalama işaretleri, heceler, karakterler...
- Harfler: Türkçe için [a-zA-Z], karakter, char.
- Söz varlığı (Vocabulary): Kelime hazinesi. Kaç kelime kullanalım?
- Bütünce / Derlem (Corpus): Külliyyat, dilbilim analizlerinde kullanılmak üzere bir araya getirilmiş yazılar topluluğu.

Temel Tanımlar

- Söz Dizimi (Syntax): Form, biçim. Bir cümle doğru şekilde nasıl oluşturulur.
- Anlambilim (Semantics): Anlam. Dilin yapısı, düşünce ve anlam arasındaki karşılıklı bağlantı.

Doğal dillerde belirsizlik (muallaklık) bulunur.

- Yeni eve geldim.
- Oh ne güzel!

Temel Tanımlar

- N-gram: art arda gelen N tane simge (token)
- 1-gram (unigram): çiçek
- 2-gram (bigram): sarı çiçek
- 3-gram (trigram): küçük sarı çiçek
- Vektör: büyüklüğü ve yönü olan nesne.

