

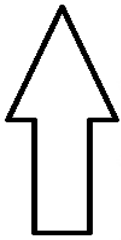
NLP (DDİ)

- Natural Language Processing (Natural Language Processing), Doğal Dil İşleme (Doğal Dil İşleme), Hesaplamalı Dilbilim (Computational Linguistic)
- Hedefler:
 - NLU (Natural language understanding) Yazılmış olan bir metni anlamak
 - NLG (natural language generation) Doğal dilde metin üretmek (otomatik çeviride hem de understanding hem de generation yapılıyor)
- Yaklaşımlar
 - Lingusitik (Dilbilim) Yaklaşımı insanlar dili nasıl anlıyor, bunu anlayarak makineye bunu öğretmek (cümleyi öğelerine ayırmak, cümleler arasındaki bağlantıyı bulmak) genelde yavaş, etkilidir
 - İstatistiksel Yaklaşımlar metin sınıflandırma, kelimelerin anlamını bilmeden kelimelerin kullanıldığı sıklığa vs bakarak sınıflandırma yapmak lingusitik'e göre daha hızlıdır
 - Hibrit Yaklaşımlar

NLP (DDİ) - Linguistik

- Natural Language Processing (NLP), Doğal Dil İşleme (DDİ)

- **Lingusitik (Dilbilim) Yaklaşımı**



- Pragmatics (kullanımbilim) kapıyı kapatabilir misin sorusuna evet kapatabilirim diye cevap verilmemesi
- Semantics (Anlambilim)
- Syntax (Sözdizim) morphology'de bütün ihtimalleri alıyoruz bu kısımda cümlenin neresinde kullanıldığına göre kesinleştiriyoruz değil kelimesi için güzel değil kötü değil gibi
- Morphology (Şekilbilim) bir kelimenin ekleri ve kökleri nelerdir, bu kelimenin farklı anlam ihtimalleri nelerdir
- İstatistiksel Yaklaşımlar
- Hibrit Yaklaşımlar

NLP (DDİ) - İstatistik

- Natural Language Processing (NLP), Doğal Dil İşleme (DDİ)
 - Lingüistik (Dilbilim) Yaklaşımı
 - İstatistiksel Yaklaşımlar
 - N-gram kelimeleri karakter karakter analiz etmek
(3 karakterin arka arkaya gelme frekansları)
 - TF-IDF
 - Word-Gram
 - BOW (Bag of Words, Kelime Çantaları) metin sınıflandırma, bir mesaja bakıp kelimeleri sınıflandırarak spam olup olmadığını anlamak
 - Hibrit Yaklaşımlar

NLP (DDİ) Örnek 1

- Natural Language Processing (NLP), Doğal Dil İşleme (DDİ)
- **Yazar Tanıma**
 - Lingusitik (Dilbilim) Yaklaşımı
 - POS-Tagging (Part of Speech Tagging, Metin Parçası Etiketleme)
cümledeki öğelerin sınıflandırılması(fil isim ..)
 - İstatistiksel Yaklaşımlar
 - İstatistiksel Dağılım sınıflandırılan öğelerin sayılarına bakılması
 - Hibrit Yaklaşımlar
 - Metindeki kelime tiplerinin dağılımı ve cümle yapıları.

Eğer yazarın ayırt edici bir özelliği var ise(çok fazla sıfat kullanması vs) bu yöntemler kullanılarak yazar bulunabilir

NLP (DDİ) Örnek 2

- Natural Language Processing (NLP), Doğal Dil İşleme (DDİ)
- **Metin Sınıflandırma (Örn. Çağrı Merkezi, Sosyal Ağlar, Mailler vs.)**
 - Lingusitik (Dilbilim) Yaklaşımı
 - Örüntü tanımı (isim tamlaması, sıfat tamlaması vs.)
 - İstatistiksel Yaklaşımlar
 - İstatistiksel Dağılım
 - Hibrit Yaklaşımlar
 - Örüntü dağılımı sınıflandırma

NLP (DDİ)

- Natural Language Processing (NLP), Doğal Dil İşleme (DDİ)
- **Anomali Yakalam (Saldırgan, veya Virüs, veya Tehdit)**
 - **Lingusitik (Dilbilim) Yaklaşımı**
 - X
 - İstatistiksel Yaklaşımlar
 - Anomali yakalama ve kelime dağılımları
 - Hibrit Yaklaşımlar
 - Anahtar kelime çıkarımı
- Gerçek uygulamalar
 - Duygu Analizi (Sentimental Analysis)
 - Metin Sınıflandırma (Text Categorization)
 - Metin Özetleme (Document Summarization)
 - Soru Cevaplama (Question Answering)
 - Etiket Bulutları ve Anahtar Kelime çıkarımı

Bazı DDİ Kütüphaneleri

- NLTK : nltk.org
- SpaCy : spacy.io
- Stanford NLP
- OpenNLP: Apache : opennlp.apache.org
- Rapid Automatic Keyword Extraction (RAKE)
- Amueller Word Cloud
- Tensor Flow : Word2Vec

Türkçe DDİ Kütüphaneleri

- Zemberek (<http://zembereknlp.blogspot.com>)
- İTÜ : <http://tools.nlp.itu.edu.tr>
- Tspell (<http://tspell.sourceforge.net/hakkinda.html>)
- Yıldız Teknik Üniversitesi: Kemik (<http://www.kemik.yildiz.edu.tr>)
- Wordnet (Balkanet)
- TrMorph (<http://coltekin.net/cagri/trmorph/>)
- TSCorpus (<https://tscorpus.com>)
- Metu- Sabancı Tree Bank ve ITU Doğrulama Kümesi (<https://web.itu.edu.tr/gulsenc/treebanks.html>)

NLP Adımları

kelimelerin boyutlarının aynı seviye
getirilmesi(FUTBOL = futbol makinenin bu
iki kelimeyi aynı algılaması lazım)

stop wordslerin çıkartılması (ve ...)

html etiketlerinin çıkartılması

