

Doğal Dil İşleme Nedir?

Mehmet Fatih AMASYALI



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Doğal Dil İşleme

- Tanım: İnsanların iletişim için kullandıkları dillerin çeşitli amaçlar için bilgisayarla işlenmesi



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Dersin Amacı

- Doğal Dil işlemenin uygulama alanları ve yöntemleri hakkında genel bilgi vermek
- Belli alanlarda uygulama yapmak



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Uygulama alanları

- Sesten yazıya çeviri
- Komut anlama
- E-mail filtreleme / sınıflandırma
- Farklı diller arası çeviri
- Duygu durum analizi
- Benzer anlamda metin üretme
- İmla denetimi / düzeltme
- Bilgiye erişim (arama motorları)
- Bilgi çıkarımı (metinden veri tabanına)
- Soru cevaplama
- Doğal dille veri tabanı (yapısal, resim, video) sorgulama
- Metin özetleme
- Sanal müşteri temsilcileri
- Sohbet robotları
- Aşırma tespiti
- Hikaye, şiir üretimi
- ...

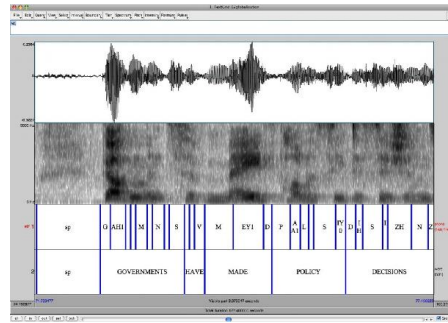


BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Sesten yazıya çeviri

- Giriş: ses sinyali , Çıkış: metin
- Kullanım alanları: sesle not alma, filmler için altyazı üretimi vb.
- $P(\text{soğan yedim}) > P(\text{sol an yedim})$
- $P(\text{Fatih Sultan Mehmet}) > P(\text{Fatih Tultan Mehmet})$
- Praat



Kemik

Komut anlama

- Giriş: text, Çıkış: komut ve parametreleri
- Sabit bir komut kümesine eşleme
- Yarın hava nasıl olacak ? →
Hava_durumu_bul(mevcut tarih+1gün,
mevcut konum)
- Ör: Siri, Cortana, Alexa, Arçelik



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

E-mail filtreleme / sınıflandırma

- Spam filter
- Gelen müşteri email lerini ilgili bölüme yönlendirme



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Farklı diller arası çeviri

- Türkçe → İngilizce
- Ne tür metinlerde daha başarılı?
- Dil modelleme gerekli. Kaynak dildeki bir kelimenin hedef dilde birden fazla karşılığı olabilir.
- $P(\text{taze balık aldım}) > P(\text{yeni balık aldım})$

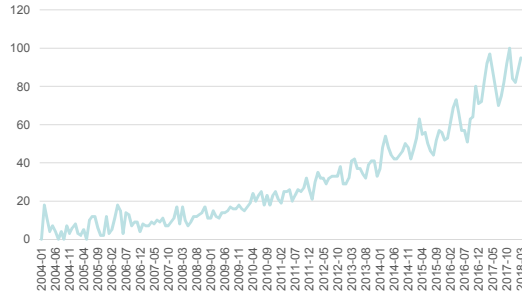


BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Duygu durum Analizi

- Bir firma, kişi, parti, kurum, ürün, olay vb. hakkında kamuoyu algısını ölçmek
- Popülerliği artan bir konu



- Algı operatörü bot'lara dikkat (tahminen %20)
- Eğitim kümesi genelde eskir. Güncelleme gerekir.

Yemeğimi 5 dakikada getirdiniz çok teşekkürler.

Yemeğimi **55** dakikada getirdiniz çok teşekkürler



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Benzer anlamda metin üretme

- Pastırma üreticileri artan maliyetler sebebiyle zam kararı aldı.
- Maliyetlerdeki artış yüzünden pastırma üreticileri fiyatları arttıracaklarını açıkladı.
- Metin benzerliğini ölçme
- Haber metni içerik değiştirme
- Web sayfası içerik değiştirme
- Twitter bot % 20



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

İmla denetimi / düzeltme

- Denetim: Morfolojik analiz gerekir ama yetmez.
- Dil modeli gerektirir
 - Maliye akanı yaptığı açıklamada
- Düzeltme: edit distance kullanılır. Dil modelleme gerekir.
 - Ali soan yedi
 - $P(\text{ali soğan yedi}) > P(\text{ali sokan yedi})$



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Bilgiye erişim (arama motorları)

- Kullanıcısı sorgusu ile ilgili sayfaları / belgeleri bulmak
- Sorgu genişletme / çoğaltma
- Cevapları sıralama (pagerank)



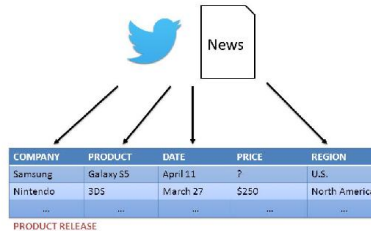
BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Bilgi çıkarımı

- Yapısal olmayan kaynağı, yapısal bir kaynağa dönüştürme
- Metin → veri tabanı kayıtları
- Yöntemler: Şablonlar, varlık ismi tanıma

Example: Information Extraction from Twitter



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Soru cevaplama

- factoid

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

Question

Which governing bodies have veto power?

Basit hikayelerde çıkarım:

- 1 Mary banyoya gitti.
- 2 Sandra bahçeye gitti.
- 3 Daniel bahçeye gitti.
- 4 Daniel ofise gitti.
- 5 Sandra oradaki sütü aldı.
- 6 Sandra sütü oraya bıraktı.
- 7 Süt nerede? bahçe 6 2



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Doğal dille veri tabanı (yapısal, metin, resim, video) sorgulama

- text2SQL (komut anlama)
- Parkta oynayan çocuk resimleri
- Komik kedi videoları
- Klasik yaklaşım: web sayfalarında resim / videolar için yazılan tanımları kullanma
- Otomatik resim ve video işleme / etiketleme geliyor (image captioning)



A person riding a motorcycle on a dirt road.



Two dogs play in the grass.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.

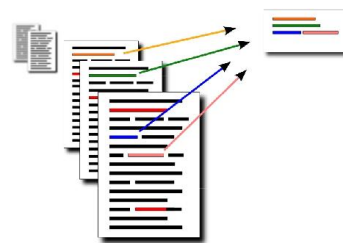


BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Metin özetleme

- Cümle seçme (kısmen kolay)
- Textrank: en çok cümlelerin benzediği cümleleri seç, seçilen cümleler de birbirinden farklı olsun
- Üretici modeller (zor)

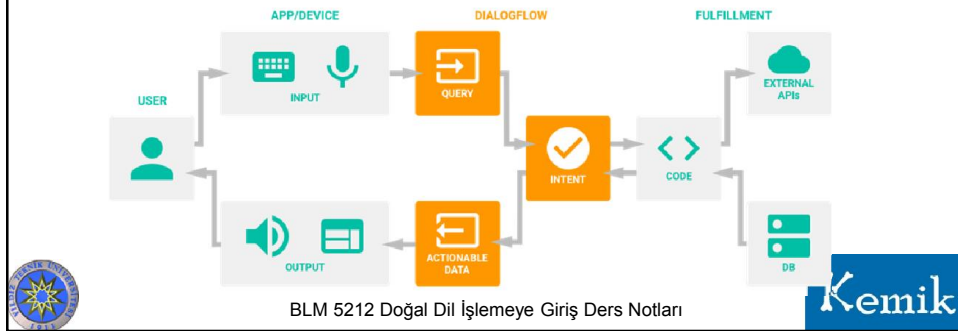


BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Sanal müşteri temsilcileri

- Sitelerde rutin sorulara cevap veren / rutin işlemleri yapan sanal asistanlar
- Cevaplayamayacağını anladığında gerçek müşteri temsilcilerine yönlendirir
- Dialogflow



Sohbet robotları

- Konu kısıtı olmadan sizinle sohbet edebilen sistemler
- Eğlencelik
- Gerçekten yapılması çok zor
- Eliza → psikoterapist
- Alice (AIML)
- Loebner prize
- 2017 kazananı

<https://www.pandorabots.com/mitsuku/>



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Aşıırma (İntihal) tespiti

- Çoğu intihal sentaktik benzerlik taşır.
- Bir çoğu da çeviri tabanlı
- Çeviri de otomatik yapılmışsa yakalamak kolay
- Semantik benzerlik



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Hikaye, şiir üretimi

- Üretilmiş bir şiir (eğlencelik)

Ben kimim içimde bir de gör.

Sen bir yana bağrına bakarsak
Bir damla dağların bir çocuk kanıyorum

Durdum korkunun kalın yağmurlarında

Ben bende bu yolda yürekler
Bir gün karanlığın beni
Görmeyen bir gün kalmadım

Doğmaya bakarsan bir zaman

- <https://www.plot-generator.org.uk/>



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Tarihçe

- Klasik yaklaşım:
 - kural tabanlı
 - anlaşılabilirlik ön planda
- Popüler yaklaşım:
 - büyük veriler üzerinde istatistiksel öğrenme tabanlı
 - fonksiyonellik ön planda
 - Turing öne geçmiş durumda



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik

Yöntemler

- Metin temsilleri
 - Şekilsel
 - Anlamsal
- Morfolojik analiz (Zemberek)
- Bağımlılık çözümleme (İTÜ NLP pipeline)
- Zamir çözümleme
- Değiş tespiti
- İstatistiksel dil modelleme
- Makine öğrenmesi



BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

Kemik