

Text Analytics and Natural Language Processing (NLP) with R Eğitimi

Eğitim Hakkında

Text Analytics and Natural Language Processing (NLP) with R eğitimi, katılımcılara metin verilerini keşfetmek ve analiz etmek için R programlamayı nasıl kullanacaklarını öğretir.

Bu ders, düz metin dosyaları, pdf'ler veya web gibi çeşitli kaynaklardan metin verilerini alma ve ardından bu verileri en son doğal dil işleme ve derin öğrenme tekniklerini kullanarak işleme yöntemlerini kapsamlı bir şekilde kapsar.

Neler Öğreneceksiniz

- Metin verilerini çeşitli kaynak biçimlerinden içe aktarmayı,
- Metin verilerini anlamlı birimlere şifrelemeyi,
- Belirli metin işlevlerini kullanarak metin verilerini karıştırmayı,
- Jetonlaştırılmış veriler üzerinde toplama önlemlerini hesaplamayı,
- Metin veri formatları arasında çeviri yapmayı,
- Belge sınıflandırması gerçekleştirmeyi,
- Konu modelleme gerçekleştirmeyi,
- NLP modellemeye uygun basit bir sinir ağı oluşturmayı.

Eğitim İçeriği

Yapılandırılmamış metin verileriyle çalışma

dize yöntemleri
normal ifade
metin dosyalarında okuma
baz incelemesi (R / Python)

İçe aktarma

bir metin dosyasındaki verileri ayrıştırma
onu düzenli bir yapıya aktarmak
bir pdf'den verileri ayrıştırma
Bir "pdf yığınınından"

web'den veri almak
Diğer yöntemlerin tartışılması
OCR
Elyazısı tanıma

Metin Verilerini Yönetme 1

düzenli bir metin biçimi
Metin veri formatlarına genel bakış
düzenli metin
simge listesi
belge terim matrisi veya belge frekans matrisi (dfm / dt)
docvars

ilişkili formatlar

kelimeleri durdur
lexica
kelime vektörleri / modelleri

Metin Verilerini Yönetme 2

metin belirtme
Lemma
n-gram
cümleler
Tweetler
Tf-idf
Günlük oranlar (tidylo)

Text Analizi

Text lexica
İnner_join ile duygu analizi
Diğer birimler tarafından analiz
Değer değişimi
VADER
Belge Sınıflandırması
Metin benzerliği - stringiest
Kosinüs
Mesafeyi düzenle
Belge sınıflandırması için Makine Öğrenimi
Bayes modeli

Konu Modelleme / Belge Kümeleme

LDA
stm

Metin ve Derin Öğrenme

Derin öğrenmeye giriş
Tensorflow + keras

Kelime vektörleri

word2vec

Text2vec
Eldiven
Spacy

Derin Öğrenme ve NLP'yi Birleştirme

CNN
RNN
LSTM

Adlandırılmış Varlık Tanıma (NER)

Konuşma etiketlemesinin (POS) parçası
Bağımlılık Ayrıştırma