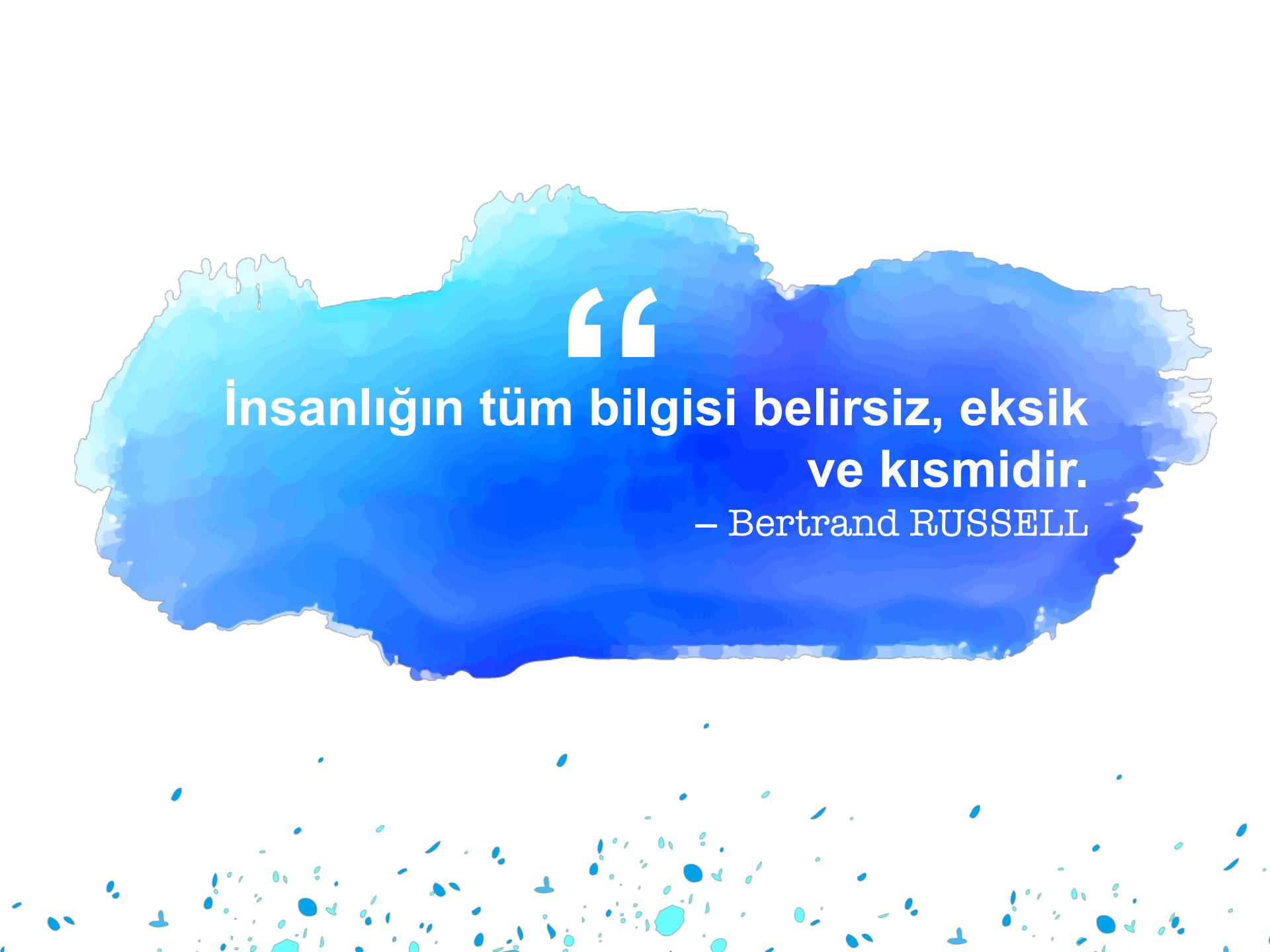




Bilgisayar Yardımıyla Çoktan Seçmeli Soru Geliştirme ve Açık Uçlu Soru Puanlama Uygulamaları

Okan Bulut

Centre for Research in Applied Measurement and Evaluation
University of Alberta



“

**İnsanlığın tüm bilgisi belirsiz, eksik
ve kısmıdır.**

– Bertrand RUSSELL

Ölçme: Örtük becerileri keşfetme bilimi



Ölçme: Gözle görülemeyenleri keşfetme bilimi



Teori → Ölçme → Değerlendirme

1	1	0	1	0	1	1	0	1	0	1	1	0	1	0	1	1	0	1	1	0
1	1	0	1	0	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1	0
0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0



-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	0	0	0	0	1	0	1	0	0	1	0	1	0	1	1	0	0	1	0	1
1	0	0	1	1	1	0	0	1	1	1	0	0	1	0	1	1	1	1	0	0

Ölçmenin kalitesini etkileyen faktörler

- Soru ve sınavların eğitim programı ile örtüşmesi
- Soru sayısı ve kalitesi
- Sınavların güvenliği ve lojistiği
- Sınav sonuçlarının elde edilmesi
- Sınav sonuçlarının değerlendirilmesi

Soru geliştirme



Soru yazmak yaratıcı bir sanattır. ([Rodriguez, 2005](#))



Soru yazmak bir bilimdir. ([Gierl & Haladyna, 2012](#))

Geleneksel soru geliştirme süreci



Geleneksel soru geliştirme süreci

- Kuzey Amerika'dan bir örnek
 - 5 ya da 6 konu/alan uzmanı
 - 1. gün: \$400
 - 2. gün: \$350
 - Ek ücret + yol masrafları: \$1500
 - Her uzman ortalama → 10-15 soru (toplam 60 soru)
 - Soru başı maliyet → $\$13,500/60 = \225
- Video ya da simülasyon içeren sorular → \$4000

Otomatik (özdevinimli) soru geliştirme

Otomatik soru geliştirme (OSG) bilgisayar yardımıyla soru modelleri üzerinden çok sayıda soru üretmeye dayalı bir yöntemdir.

Alan uzmanları soruları çözmede gerekli olan bilgi ve becerileri belirler. Bu aşama uzmanlar bilgi, deneyim, ve yaratıcılıklarını kullanmaktadır.



YARATICILIK

Her bir soru modelinde yer alan birçok bileşenin bilgisayar teknolojileri kullanılarak bir araya getirilmesi sürecidir.

Genellikle lineer programlama kullanılmaktadır.



BİLİM

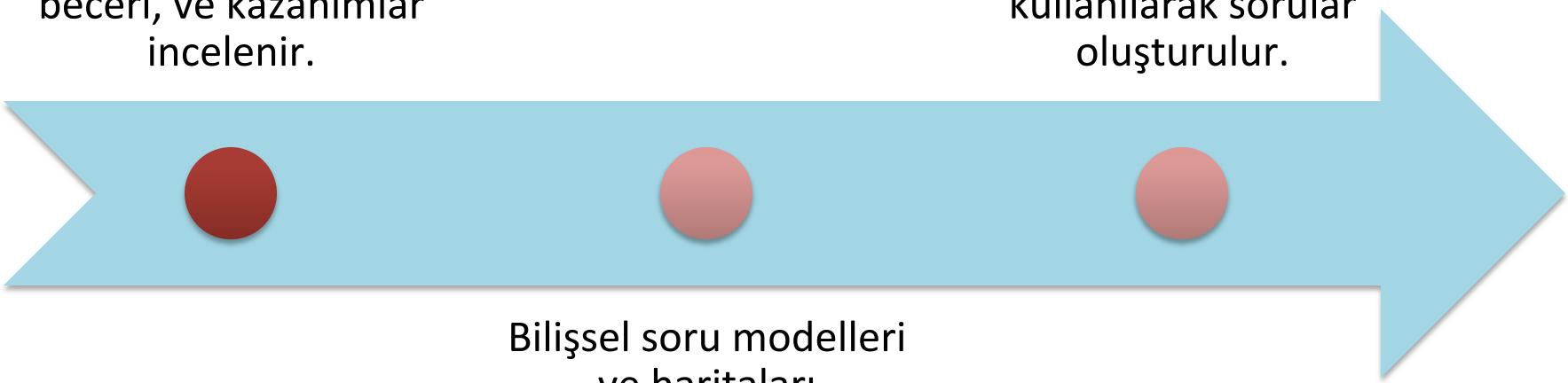
Sorular için yaratıcılık kullanılarak işlenen içerik ve diğer bileşenlerin bilgisayar teknolojileri yardımıyla çoktan seçmeli ya da açık uçlu sorulara dönüştürülmesi süreci.

OSG

Otomatik soru geliştirme süreci

Yeterli sayıda alandan uzmanlar ile bilgi, beceri, ve kazanımlar incelenir.

Şablon sorular hazırlanır ve bilgisayar yardımıyla bilişsel soru modelleri kullanılarak sorular oluşturulur.



Bilişsel soru modelleri ve haritaları oluşturulur. Anahtar özellikler belirlenir.

Otomatik soru geliştirme süreci

- ❑ Anahtar özellik sayısı arttıkça geliştirilebilecek soru sayısı da artar.
- ❑ Kapsamlı bir soru modelinde tüm kombinasyonlar uygulandığında yüzlerce soru tek bir tuşla üretilebilir.
- ❑ Üretilen soruların birbirine benzerliği kosinüs benzerlik indeksi ile ölçülebilir.
 - ❑ İki soru arasında kaç kelime ortak olarak kullanıldı?
 - ❑ 0 (sıfır benzerlik) ile 1 (tam benzerlik) arasında

Üretilen soruların geçerliğinin incelenmesi

1. Soru modellerinin geçerliği

- ❑ Soru modellerinin oluşturulmasında yer alan alan uzmanları modelleri tekrar baştan inceleyerek modeldeki bileşenlerin doğruluğunu kontrol eder.

2. Soruların geçerliği

- ❑ Soru modellerinin oluşturulmasında **yer almayan** alan uzmanları üretilen soruları inceler ve yeterliliklerine dair görüş bildirir.

Soru geliştirmek için kullanılan programlar

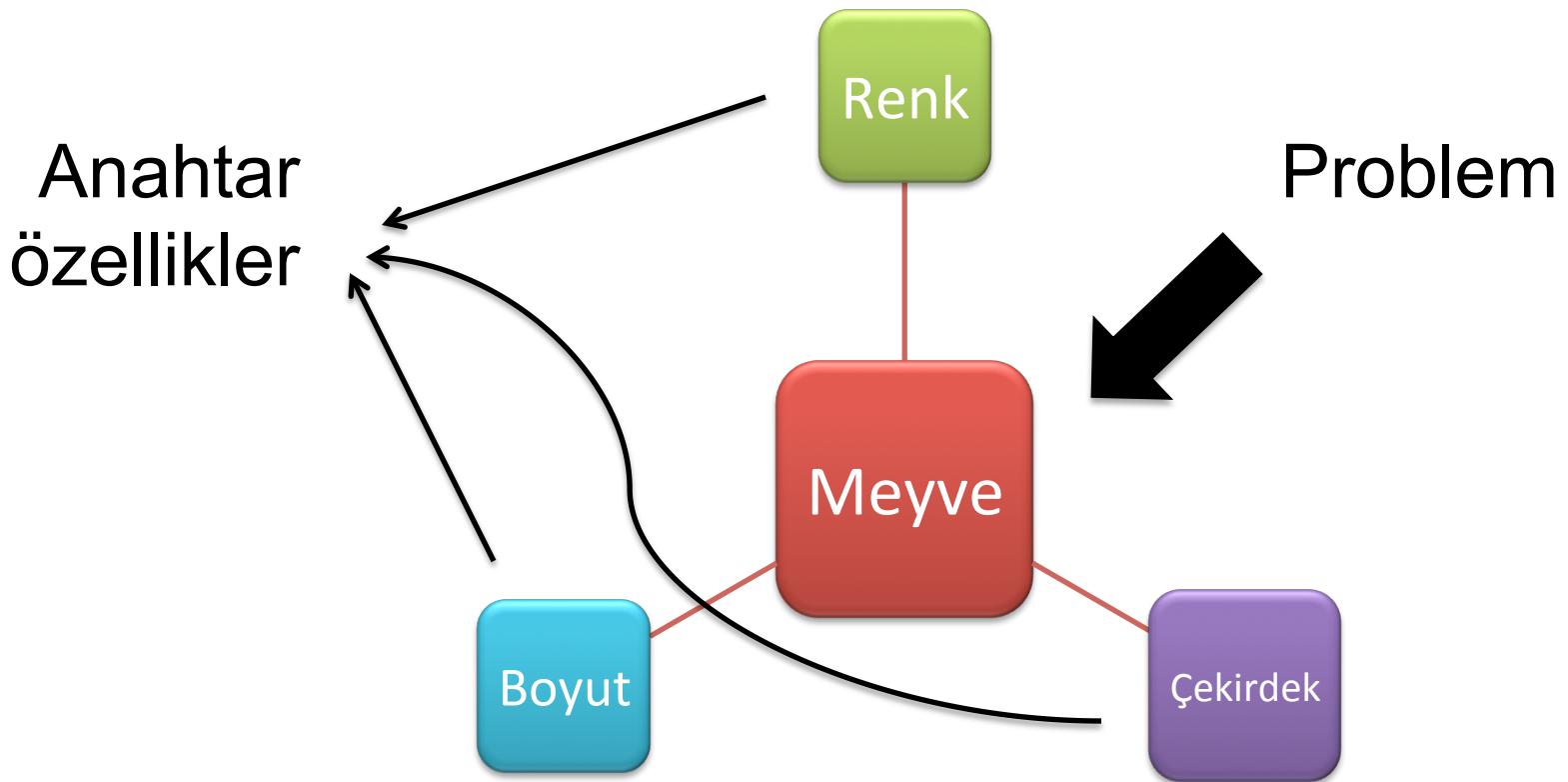
- [Item Distiller](#)
- [ModelCreator](#)
- [Automatic Question Paper Generator](#)
- [Itematic](#)
- [AIG](#)

Diğer otomatik soru geliştirme türleri

- Çok katmanlı soru modelleri ([Gierl & Lai, 2013](#))
 - Soru kökü
 - Verilen bir senaryoya göre bir durumu yorumlama
 - Verilen sonuçları yorumlama
 - Doğrudan soru cümlesi ile problemi ifade etmek
 - Anahtar özellikler
- Şablonla bağlı kalmadan doğal dil işleme (natural language processing) teknikleriyle soru geliştirme ([Aggrey et al, 2018](#); [Baghaee, 2017](#))

Otomatik soru geliştirme – Örnek (1)

Elimizdeki bir meyvenin rengi [renk], boyutu yaklaşık olarak [boyut], ve içindeki çekirdeği [çekirdek] ise bu meyve aşağıdakilerden hangisi olabilir?



Otomatik soru geliştirme – Örnek (1)

Elimizdeki bir meyvenin rengi [renk], boyutu yaklaşık olarak [boyut], ve içindeki çekirdeği [çekirdek] ise bu meyve aşağıdakilerden hangisi olabilir?

- ❖ **Renk:** Siyah, kırmızı, yeşil
- ❖ **Boyut:** 2 cm, 3 cm, 7 cm, 8 cm
- ❖ **Çekirdek:** Sert, çoklu ufak
- ❖ **Seçenekler:** Kiraz, kayısı, erik, domates, limon, avokado

Otomatik soru geliştirme – Örnek (1)

Elimizdeki bir meyvenin rengi **siyah**, boyutu yaklaşık olarak **2 cm**, ve içindeki çekirdeği **sert** ise bu meyve aşağıdakilerden hangisi olabilir?

- a) Erik
- b) Kiraz
- c) Kayısı
- d) Domates

Otomatik soru geliştirme – Örnek (1)

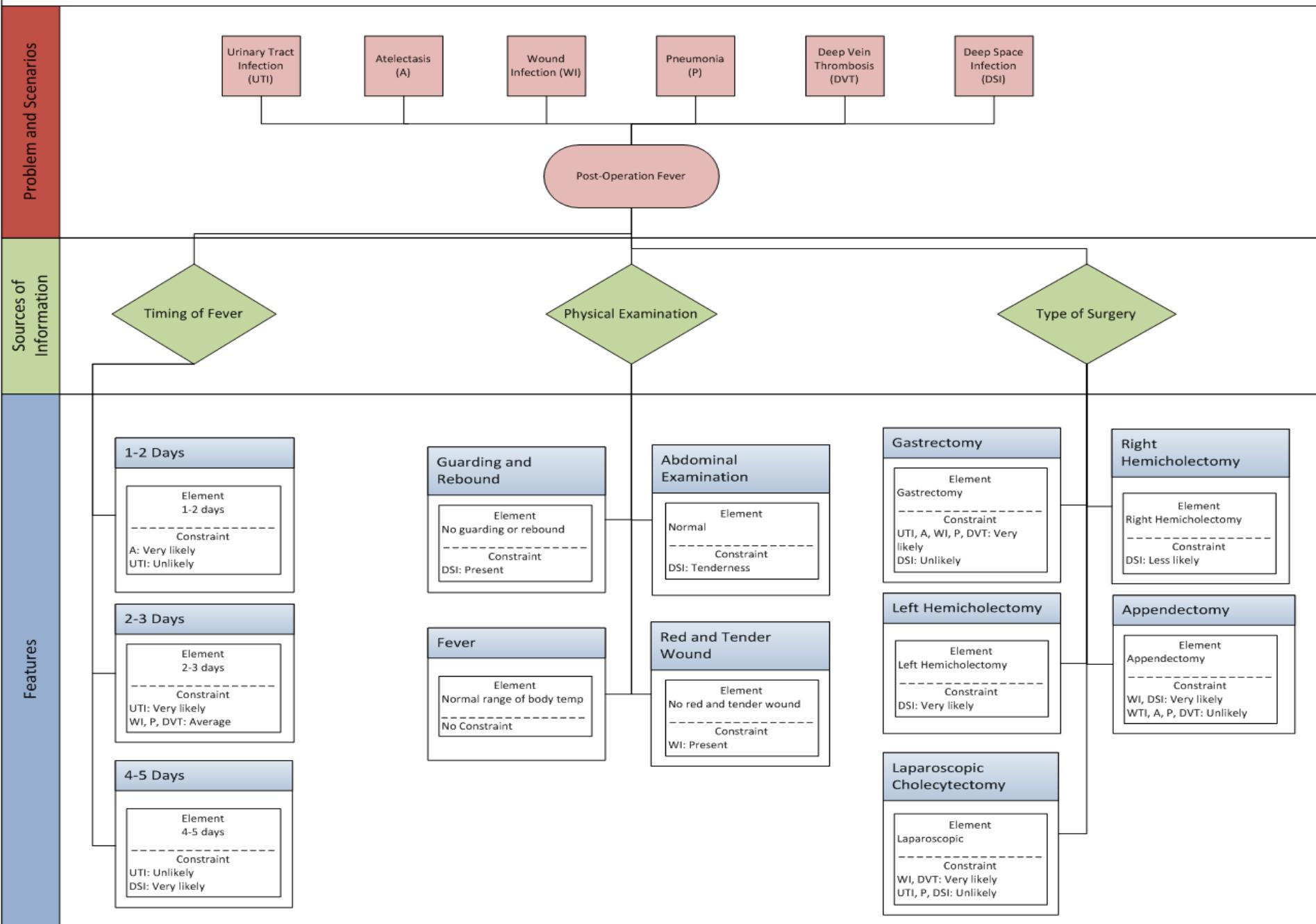
Elimizdeki bir meyvenin rengi **siyah**, boyutu yaklaşık olarak **2 cm**, ve içindeki çekirdeği **sert** ise bu meyve aşağıdakilerden hangisi olabilir?

- a) Erik → *Sert çekirdekli, siyah, ama daha büyük*
- b) Kiraz*
- c) Kayısı → *Siyah değil ve daha büyük*
- d) Domates → *Çoklu ufak çekirdekli ve daha büyük*

Otomatik soru geliştirme – Örnek (2)

- Bu bilişsel bir soru modelinde tıbbi bir tanı koymada doktor adaylarının sahip olması gereken bilgi, beceri, ve konu uzmanlığını gösterilmektedir.
- Modelde üç tane çıktı yer almaktadır:
 1. **Problemin** belirlenmesi (*ameliyat sonrası ateşlenme*)
 2. Tıbbi tanı koymada kullanılacak **bilgi kaynağı** (*örneğin, ameliyat türü*)
 3. Problemi farklılaştırmak için kullanılacak **anahtar özellikler** (*örneğin, hastanın o anki ateşi*)

A structure of problem solving knowledge for issues related to post-operation fevers



Otomatik soru geliştirme – Örnek (2)

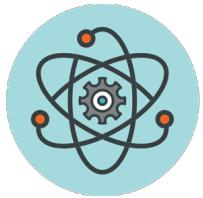
Bilişsel model kullanılarak şablon sorular oluşturulur ve modeldeki bileşenler kullanılarak farklı sorular elde edilir.

54 yaşındaki bayan bir hasta [...] ameliyatı olmuştur. Ameliyattan [...] gün sonra hastanın ateşi 38.5 derece olarak ölçülmüştür. Fiziksel muayene sonucunda [...] hastada belirtileri tespit edilmiştir. Aşağıdakilerden hangisi bu aşamada yapılabilecek en ideal uygulamayı göstermektedir?

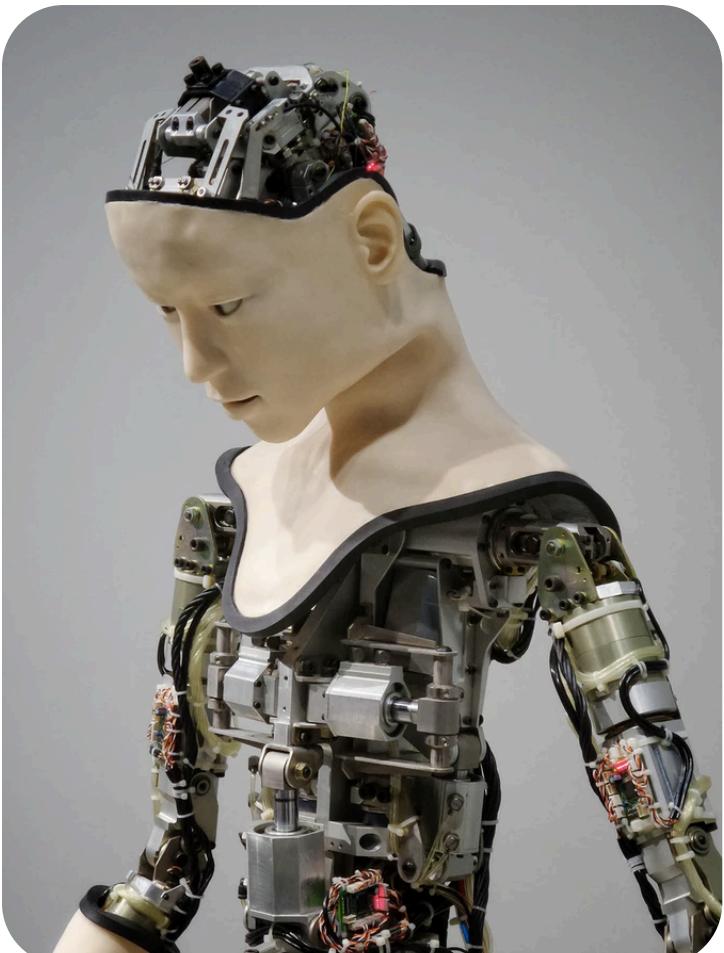
[...] Mide ameliyatı; kalın bağırsak ameliyatı; kolon ameliyatı; apandisit ameliyatı; safra kesesi ameliyatı

[...] 1 ile 6 gün arası

[...] Kızarmış ve hassas bir yara bölgesi; karın kısmında ağrı; baldırda ağrı; karında kasılma



Yapay zeka ve dil



Google AI

<https://ai.google/>

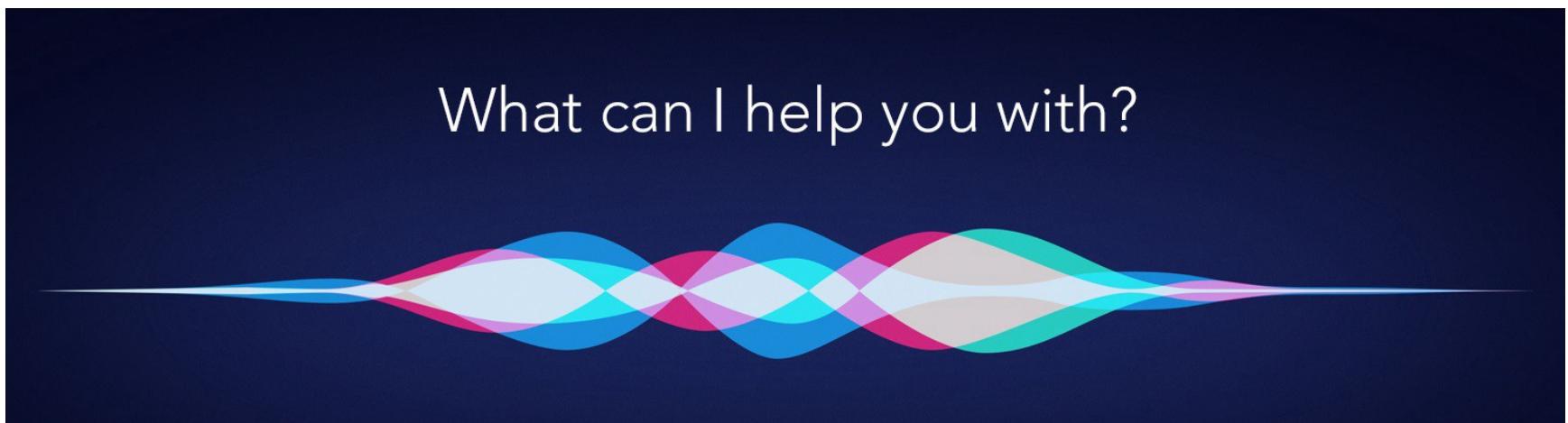
<https://cloud.google.com/natural-language/>

“Dil her şeyin anahtarıdır”, diyor Kurzweil.

Bir makale yazdığınızda sadece bir dizi ilginç kelimeyi bir araya getirmiyorsunuz. Makalenizdeki mesaj aslında bir bilgi ve bilgisayarlar henüz bu bilgiyi tam olarak algılayamıyor. Bu nedenle bilgisayarların yazıları okuyabilmesini ve kullanıcıların sorularını cevaplamak için mantıklı bir diyaloga girebilmelerini istiyoruz.” -- Ray Kurzweil (Google'da mühendislikten sorumlu direktör)

Kaynak: [The Guardian](#)

Bilgisayar dili anlayabilir mi?

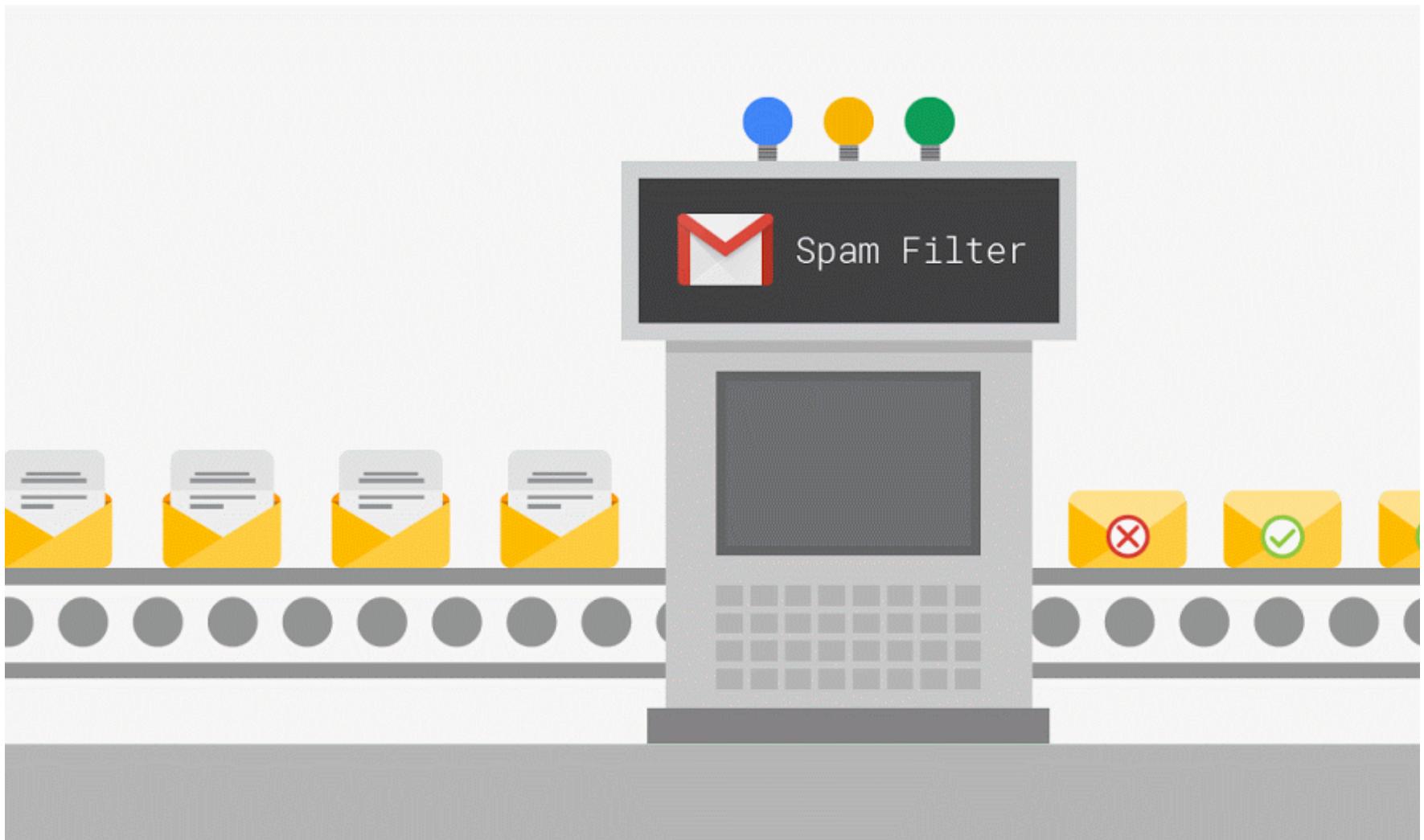


Kaynak: <https://www.analyticsindiamag.com/behind-hello-siri-how-apples-ai-powered-personal-assistant-uses-dnn/>



Kaynak: https://blogs.technet.microsoft.com/microsoft_presse/auf-diesen-4-saeulen-basiert-cortanas-persoenlichkeit/

Bilgisayar dili anlayabilir mi?



Kaynak: <https://i.gifer.com/Ou1t.gif>

Doğal dil işleme teknikleri



- Doğal dil işleme ([natural language processing](#))
 - [Yapay zeka + bilgisayar bilimi + dilbilim] → Metin ve konuşmaları anlamak
 - Makine öğrenme (machine learning) ve derin öğrenme (deep learning) modellerinin yazılı ve sözlü metinlere uygulanması
- Örnek: Akıllı telefonlarda yazıkça yeni kelimelerin önerilmesi, kelimelerin otomatik tamamlanması

Doğal dil işleme teknikleri

- Bilgisayar insanlar gibi dili sezgisel olarak **anlayamıyor**.

“LeBron James dün akşamki maçta adeta alev almıştı. Rakip takımı tek başına dağıtı.”



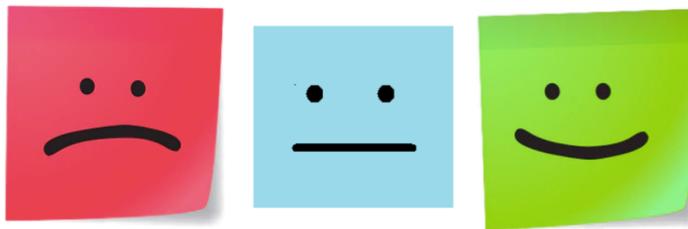
Doğal dil işleme teknikleri

- Doğal dil işleme → diziler ya da dizgiler (token)
 - Cümle dizileri
 - Kelime dizileri
 - Kapantılı (bitiş) kelimeler (ve, çünkü, fakat, ben, kendi, gibi)
 - Kurallı ifadeler
 - Kelimelerin vektörleştirilmesi

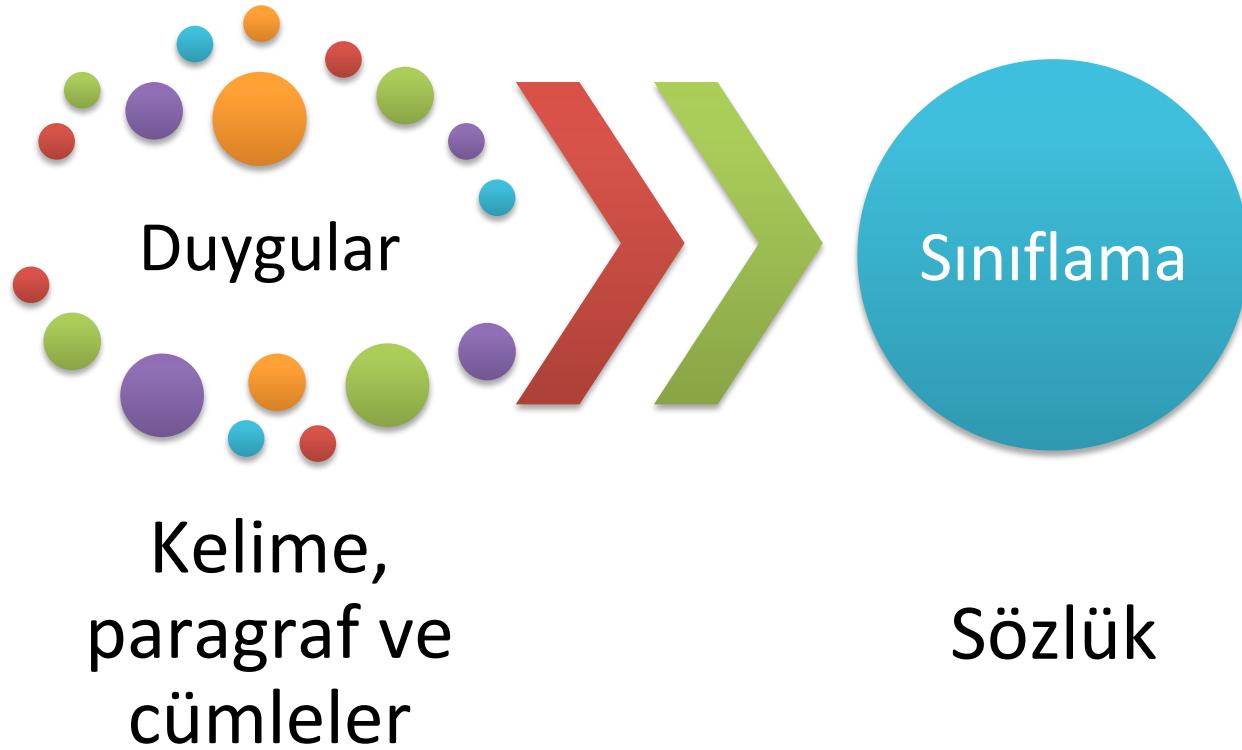


Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam gravida metus ornare arcu posuere, quis sollicitudin ex efficitur. Aliquam erat volutpat. Nunc id facilisis ex. Aliquam fringilla viverra arcu, nec rhoncus nunc dictum eget. Nulla interdum at elit in blandit. Vestibulum gravida, ligula et dignissim dictum, mauris neque maximus magna, quis ultrices purus elit eu erat. Pellentesque ex dui, viverra mollis nisi nec, pharetra bibendum ipsum. Donec sit amet placerat magna. Fusce rutrum eros in dictum vehicula. Sed vitae massa sapien. Donec id nunc sem. Aenean euismod dolor magna, sed dapibus lectus rutrum eget. Morbi nec porttitor diam.

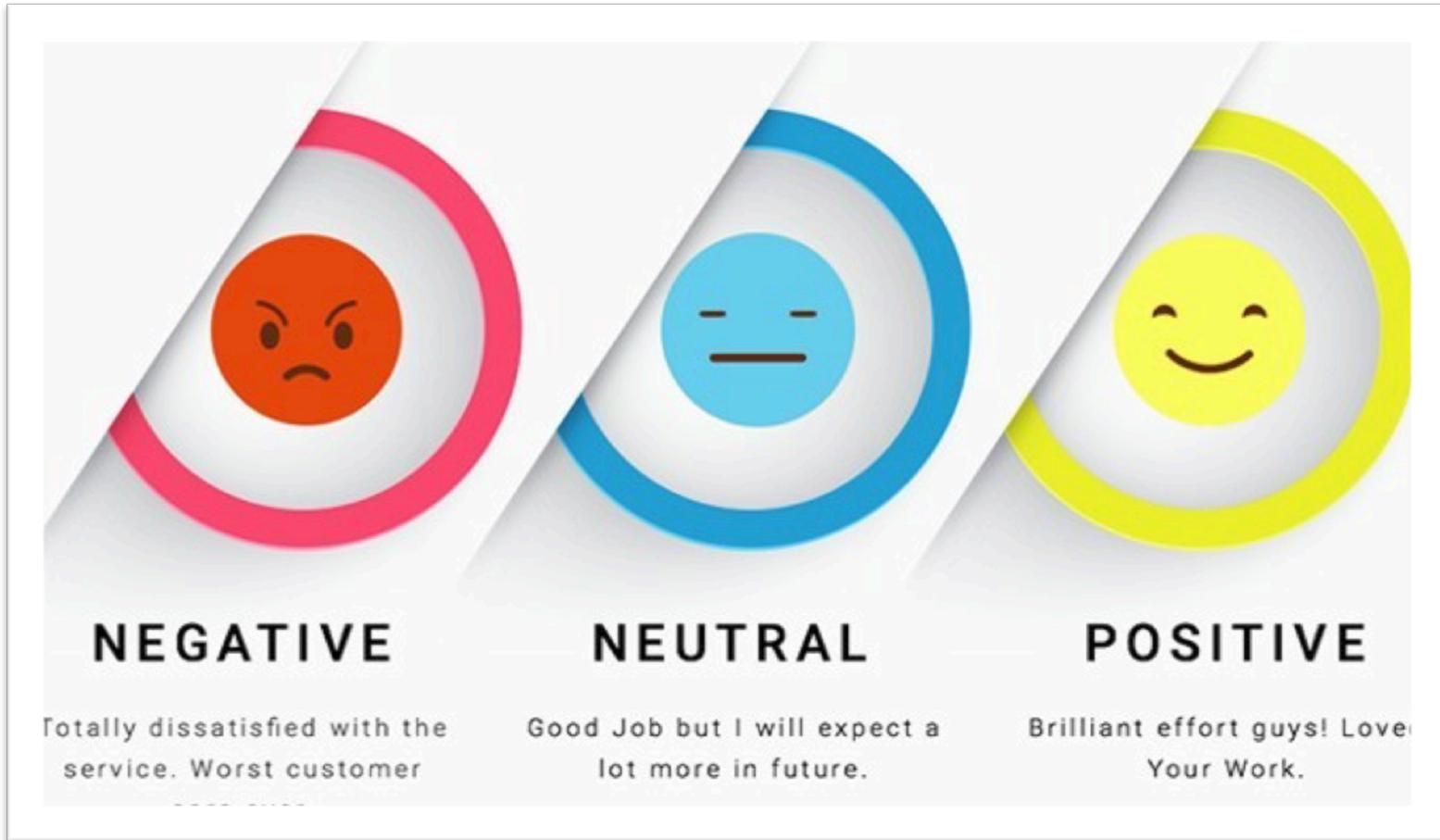
Doğal dil işleme teknikleri



- Kelime ve cümlelerde duygusal analizi



Doğal dil işleme teknikleri



Kaynak: <https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html>

Ölçme ve dil ilişkisi

Çoktan seçmeli soru



Açık uçlu soru



YAZILI SINAV YAPACAĞIM DEDİĞİMDE LİSANS
ÖĞRENCİLERİ (TEMSİLİ)

Ne zaman açık uçlu soru tercih edilmeli?



Açık uçlu



Çoktan
seçmeli

Çözüm: Doğal dil işleme teknikleri ile puanlama

- ❑ Automated essay scoring ismiyle de bilinen yöntem ile öğrencilerin sorulara verdikleri cevaplar dijital ortama aktarılıp analiz edilebilir ve puanlanabilir.
- ❑ Denetlenen öğrenme yöntemi (*supervised learning*)
 - ❑ [Keras](#)
 - ❑ [Tensorflow](#)
 - ❑ [R](#) ya da [Python](#)



Doğal dil işleme teknikleri ile puanlama süreci



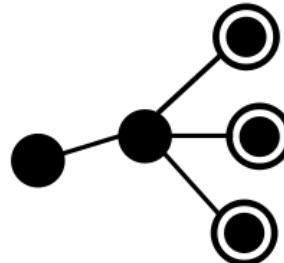
Öğrencilerin cevapları



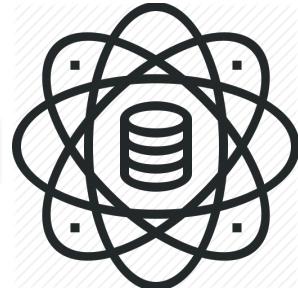
Öğretmenler cevapları değerlendirdir.



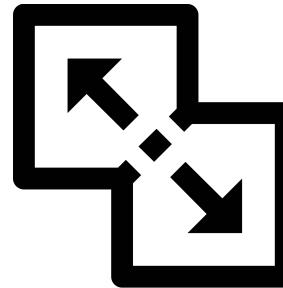
Cevaplar dijital ortama aktarılır.



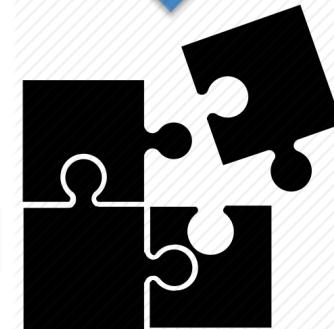
Set 2 ile model test edilir.



Set 1 ile model oluşturulur.



Veriler iki ayrı sete bölünür.



Cevaplar cümle ve kelimeleme ayrıılır.

Doğal dil işleme teknikleri ile puanlama - Örnek

- **Örneklem:** 2230 10. sınıf öğrencisi
- **Sınav:** 10 soruluk fen konularını kapsayan açık uçlu bir sınav. Her soru 0 puan (minimum) ve 3 puan (maksimum) aralığında değerlendirilmiştir.
- **Değerlendirme:** 2 öğretmen tarafından tüm sınavlar bağımsız bir şekilde puanlanmıştır. Korelasyon = 0.91
- **Analizler:** Örnekte sadece 1. soruya verilen cevaplar incelenecektir.

Doğal dil işleme teknikleri ile puanlama - Örnek

1. Soru: Bir grup öğrenci aşağıdaki adımları takip ederek bir deney gerçekleştirmiştir:

1. Verilen dört örnekteki maddelerin kütleleri hesaplayınız.
2. Her bir maddenin üzerine aynı boyutta bir kap kullanarak sirke dökünüz.
3. Maddeleri ayrı kaplara yerleştiriniz.
4. 24 saat sonunda maddeleri kaplardan çıkarıp damıtılmış su ile yıkayınız.
5. Maddelerin kuruması için 30 dk bekleyiniz.
6. Her bir maddenin kütlesini tekrar hesaplayınız.

Doğal dil işleme teknikleri ile puanlama - Örnek

Öğrencilerin bulguları aşağıdaki tabloda gösterilmektedir.

Örnek	Başlangıç kütlesi	Bitiş kütlesi	Fark
Mermer	9.8	9.4	-0.4
Kireç taşı	10.4	9.1	-1.3
Tahta	11.2	11.2	0.0
Plastik	7.2	7.1	-0.1

Yukarıda anlatılan deneyi tekrar etmek için hangi ek bilgilere (en az 3 tane) ihtiyaç olduğunu açıklayınız.

Doğal dil işleme teknikleri ile puanlama - Örnek

- ❑ Verileri iki sete böldük (Set 1: 1672; Set 2: 558)
- ❑ Cevaplarda en sık kullanılan 5000 kelimeyi tespit ettik.
- ❑ Her bir kelimeye birer indeks atadık (1 ile 5000 arasında).
- ❑ Her bir öğrencinin cevabı içerisindeki kelimeler 5000 kelimelik liste ile eşleştiğinde 1, diğer türlü 0 atayarak vektörler oluşturduk.

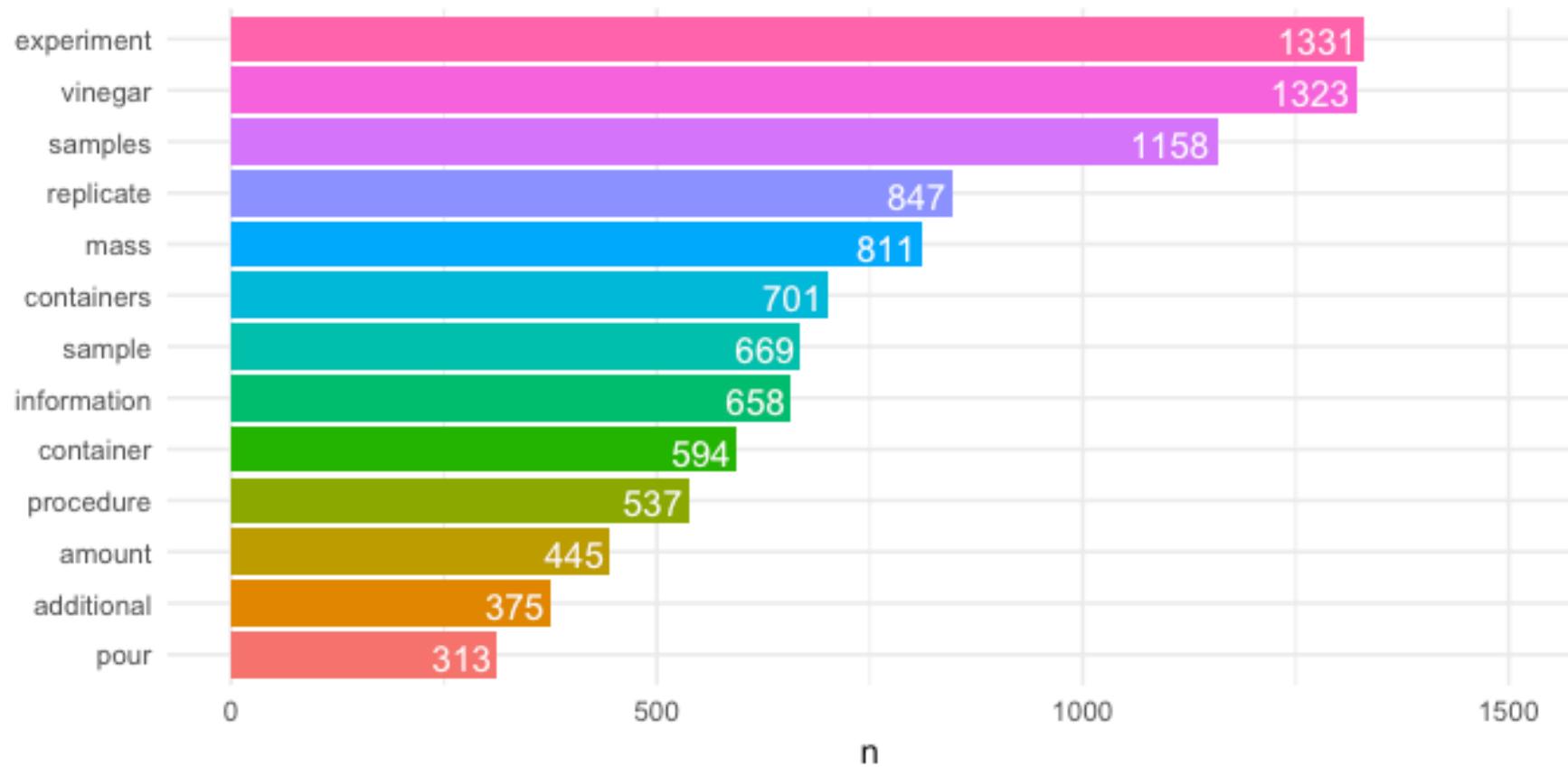
```
top_words = 5000 #top most-frequent words extracted from the dataset

tokenizer = Tokenizer(nb_words=top_words)
tokenizer.fit_on_texts(texts)
sequences = tokenizer.texts_to_sequences(texts)

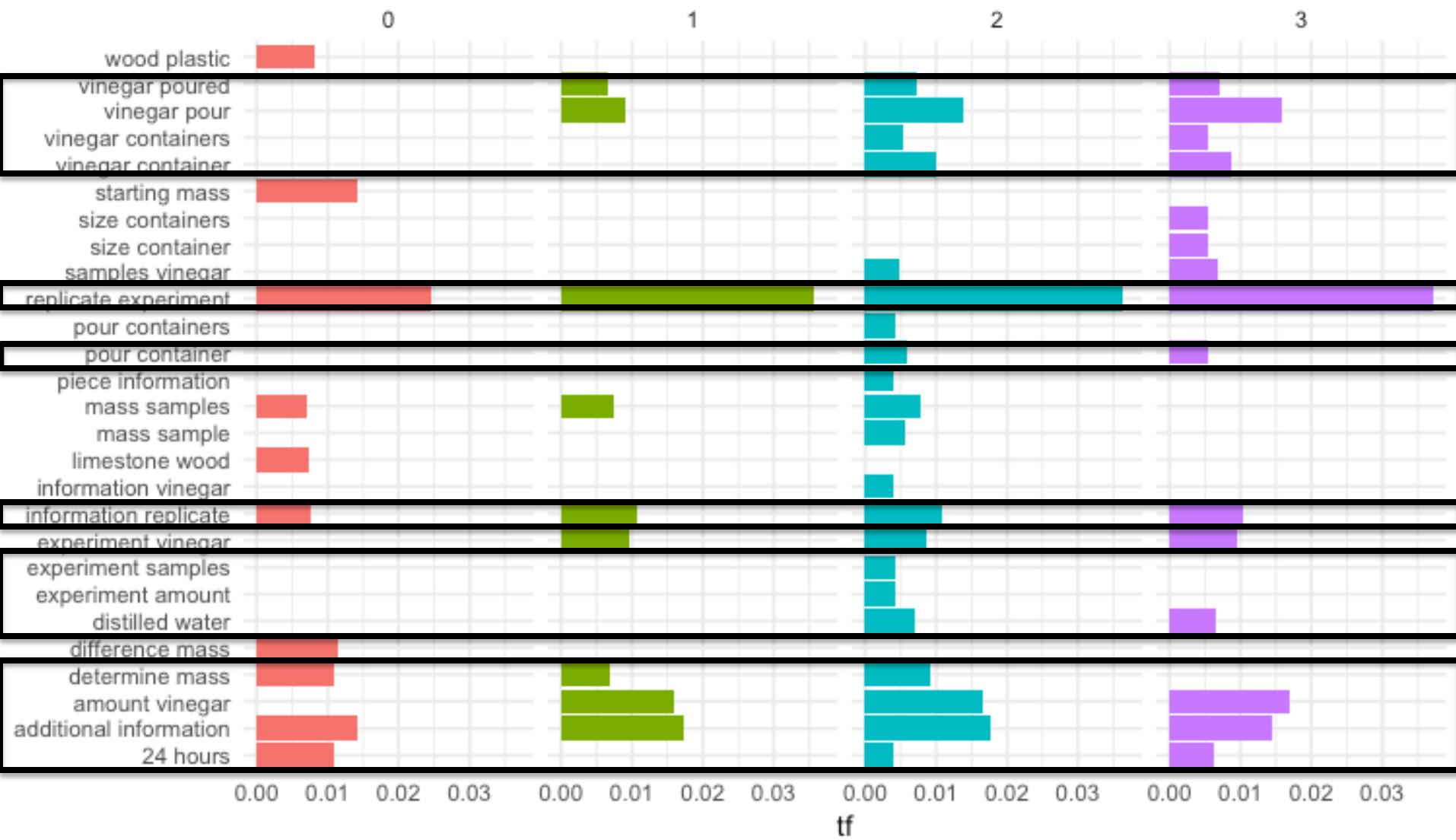
word_index = tokenizer.word_index
print('Found %s unique tokens.' % len(word_index))

max_response_length = 500
data = pad_sequences(sequences, maxlen=max_response_length)
```

Doğal dil işleme teknikleri ile puanlama - Örnek



Doğal dil işleme teknikleri ile puanlama - Örnek



Doğal dil işleme teknikleri ile puanlama - Örnek

- ❑ Cevaplarda yer alan ve, ama, çünkü gibi kelimeleri kaldırırdık.
- ❑ Her bir cevap içerisinde kullanılan toplam kelime ve cümle sayılarını hesapladık ve önceki aşamada hazırladığımız 0-1-0 gibi kodlanmış veriler ile birleştirdik.

```
from nltk.tokenize import word_tokenize, sent_tokenize
from nltk.corpus import stopwords

#Set of stopwords from nltk.corpus package
stop = set(stopwords.words('english'))

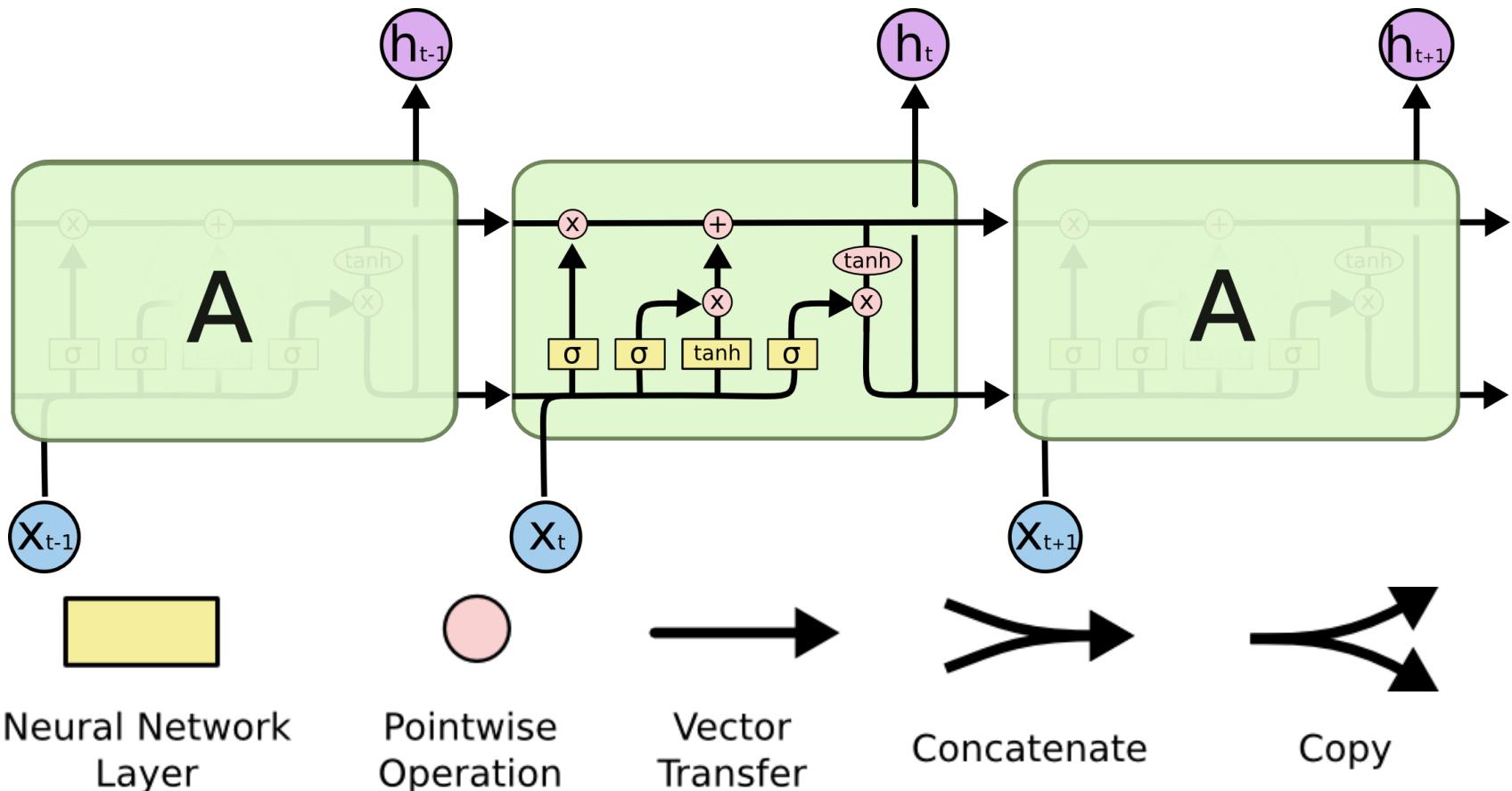
word_counts_X = [] # List of word Counts in each example response
for x in essay_text:
    word_counts_X.append(len([i for i in word_tokenize(x) if i not in stop]))

sent_counts = [] # Set of sentence counts in each example response
for x in essay_text:
    sent_counts.append(len(sent_tokenize(x)))

#Concatenate the three data-vectors into a single matrix or feature-set
features_train = np.column_stack((sent_count_train,word_count_train))
features_test = np.column_stack((sent_count_test,word_count_test))
```

Doğal dil işleme teknikleri ile puanlama - Örnek

- Eldeki verilere (Set 1) **Long Short-Term Memory Networks (LSTM)** modelini uyguladık.



Doğal dil işleme teknikleri ile puanlama - Örnek

- Set 1'den elde edilen model Set 2 üzerinde test edildiğinde doğru puanlama oranının yaklaşık olarak **%81** olduğu görülmüştür.

```
Train on 26200 samples, validate on 2912 samples
Epoch 1/20
26200/26200 [=====] - 663s - loss: 1.1203 - acc: 0.6412 - val_loss: 0.6958 - val_
acc: 0.6954
Epoch 2/20
26200/26200 [=====] - 662s - loss: 0.7272 - acc: 0.7366 - val_loss: 0.5990 - val_
acc: 0.7521
Epoch 3/20
26200/26200 [=====] - 662s - loss: 0.6306 - acc: 0.7829 - val_loss: 0.5512 - val_
acc: 0.7785
Epoch 4/20
26200/26200 [=====] - 663s - loss: 0.5670 - acc: 0.8151 - val_loss: 0.5434 - val_
acc: 0.7874
Epoch 5/20
26200/26200 [=====] - 663s - loss: 0.5175 - acc: 0.8406 - val_loss: 0.5189 - val_
acc: 0.8070
```



Doğal dil işleme teknikleri ile puanlama - Örnek

- Set 1'de geriye kalan 9 soru da modellemeye eklendiğinde öğrencilerin sınavda aldığı toplam puanın **%87** doğruluk oranıyla tahmin edilebildiği tespit edilmiştir.

Avantajlar:

- Güvenilir, tutarlı, ve hızlı puanlama
- Değerlendirme sürecinin maliyeti açısından tasarruflu

Sınırlılıklar:

- Modelin yeterli ölçüde doğrulanabilmesi için örneklem boyutu büyük olmalıdır.
- Öğretmenlerin puanlama esnasında tutarlı olmaları gereklidir.



UNIVERSITY OF
ALBERTA



Teşekkürler!!!

Görüş ve sorularınız için: bulut@ualberta.ca