



DERLEM

Doç. Dr. Pınar CİHAN

Derlem

- Bir dilde yazılmış ve söylenmiş olanları içeren sözcük kümesine derlem denir.
- Bir dil hakkında varsayım oluşturmak veya dilin dil bilimsel tanımına başlamak için kullanılan yazılı metin veya kayıtlı konuşmalardan oluşan bilgi topluluğudur.
- Dil bilimciler, dildeki değişim ve gelişmeleri incelemek için kullanırlar.
- Bilişimciler dilin modelini çıkarmak, biçim bilimi ve söz dizimi çalışmalarında yararlanmak için kullanırlar.
- Bir doğal dilin özelliklerinin ortaya çıkarılması; dil eğitimi, sözlük hazırlama, dil çözümleme, anlam çıkarma, konuşma tanıma, optik karakter okuma, diller arası çeviri, veri şifreleme/çözme, heceleme, engellilere yardımcı olma gibi konularda temel yapıyı oluşturur.
- Dil özellikleri bilindiğinde;
 - harfler arası ilişkiler, sözcükler arası bağıntılar ve metnin dili belirlenebilir.

Derlem

- ▶ Doğal dillerin özellikleri;
 - ▶ Yapısal ve olasılıksal olarak 2 temel yöntemle incelenir.
- ▶ Yapısal inceleme; sözcük türleri, kök, gövde ve ekler.
- ▶ **Olasılık** temelli incelemeler;
 - ▶ **Harf** bazında; Bir harften sonra diğerinin gele sıklığı, ünlü ve ünsüz harfler
 - ▶ **Kelime** bazında; Kelimelerdeki harf sayıları, birbirinin arkasından gelme durumları, cümledeki kelime dizilimleri.
- ▶ Güncel dilden veya eski kitap ve belgelerden oluşturulabilir.
 - / ▶ Eski dildeki belgeler için optik okuyucularla tarama yapılarak bilişim ortamına aktarılmalıdır.
- ▶ Derlemi oluştururken, birleşik kelime, çoğul kelime gibi aynı kökten türeyen ancak farklı anlamlar içerebilen kelimelerin ne şekilde değerlendirileceği belirlenmelidir.
 - ▶ Bu çözümleme algoritmalarının dikkate alacağı durumları belirlemek için gereklidir.
- ▶ Her dil için hazırlanacak derlemde, kullanılacak etiketler o dil için gerekli olmalı.

Derlem

- ▶ DDİ için oluşturulan bir derlem;
 - ▶ Sözcükler nitelikleri ile etiketlenmiş olabilir,
 - ▶ Kelimelerin kökleri, aralarındaki bağlar ve anlamları belirtilmiş olabilir.
- ▶ Kapsamlı ve nitelikli bir derlem hazırlayabilmek yoğun insan gücü gerektirir.
 - ▶ Kelimelerin kök, bağ ve anlamlarının etiketlenmesi gibi işlemler için
- ▶ Kapsamlı bir derlemin etiketleme veya kategorize etme işlemi, büyük miktarda veri işleme gerektirdiği için insanlar tarafından yapılması oldukça zahmetli ve zaman alıcı olabilir. Bu nedenle, makine öğrenimi ve yapay zeka teknolojileri kullanılarak otomatik etiketleme yöntemleri geliştirilmiştir.
 - ▶ Bu yöntemler arasında, kelimeleri etiketlemek için kullanılan en yaygın yaklaşım, "üst denetimli öğrenme" olarak adlandırılan bir yöntemdir. Bu yöntemde, bir etiketli veri kümesi kullanılarak bir makine öğrenimi modeli eğitilir ve daha sonra bu model, yeni verileri otomatik olarak etiketlemek için kullanılır.
 - ▶ Diğer etiketleme yöntemleri arasında, "unsupervised learning" ve "semi-supervised learning" gibi yöntemler de bulunmaktadır. Bu yöntemler, daha az etiketli veri kullanarak etiketleme yapma işlemi gerçekleştirir.

Derlem

- Otomatik etiketleme yöntemleri, insanlar tarafından yapılan işi azaltarak, büyük veri kümeleri üzerinde hızlı ve verimli bir şekilde etiketleme yapmamıza olanak tanır. Ancak, bu yöntemlerin doğruluğu ve güvenilirliği, veri setinin kalitesi, model eğitim süreci ve diğer faktörlere bağlı olarak değişebilir.

Derlem Türleri

Amacına göre

- Dil Bilimi deremi
 - Dil bilimi çalışmalarına yöneliktir.
- NLP derlemi
 - Doğal dil işleme çalışmalarına yöneliktir.

Yapısına göre

- Dengeli derlem
- Dengesiz derlem

Dengesiz Derlem

- farklı kaynaklardan ve farklı uzunluklarda metin örnekleri alarak oluşturulan bir metin koleksiyonudur.
- Örneğin, bir dengesiz derlem, bir dildeki gazete makaleleri, kitap bölümleri, sosyal medya mesajları, akademik makaleler ve diğer farklı kaynaklardan alınan metinleri içerebilir. Bu metinler, belirli bir dilbilgisi yapısına uygun olarak düzenlenmez, ancak daha doğal bir şekilde bir araya getirilir.
- Dengesiz derleme, doğal dil işleme uygulamaları için önemlidir, çünkü bir dilden farklı kaynaklardan elde edilen farklı uzunluklarda metinler, gerçek hayatta karşılaşılan dil kullanımına daha yakın bir örnek sağlar. Ancak, farklı kaynaklardan ve uzunluklardan örnekler alınması nedeniyle, bazı bölümlerin daha fazla temsil edilmesine ve diğerlerinin daha az temsil edilmesine neden olabilir. Bu nedenle, doğal dil işleme modellerinin eğitimi için dengeli ve dengesiz derlemelerin her ikisinin de kullanılması yaygındır.

Dengeli Derlem

- bir doğal dil işleme çalışması için kullanılan bir metin koleksiyonudur.
- Dengeli derlem, belirli bir dilin kullanımı hakkında kapsamlı bir bilgi sağlamak için önceden belirlenmiş kriterlere göre seçilmiş ve özenle hazırlanmış bir metin kümesidir. Dengeli derlemede, belirli bir dilbilgisi yapısı ve dil kullanımı özellikleri dikkate alınarak, belirli bir türde metinlerden eşit miktarda örnekler alınır. Bu örnekler, farklı kaynaklardan seçilir ve belirli bir dilbilgisi yapısına uygun olarak düzenlenir. Bu, doğal dil işleme modelinin birden fazla alanda çalışabilmesi ve daha genel bir anlayışa sahip olabilmesi için gereklidir.

Dengeli Derlem

- ▶ İlk olarak derlemin dili ve kapsayacağı zamana karar verilmelidir.
- ▶ Ağırlıklı olarak dil bilimcilerin kullanma isteyeceği derlem türüdür.
 - ▶ Dilbilimciler; bir kelimenin doğduğu zamanı belirlemek, kullanım sıklığını incelemek, dildeki yerli ve yabancı kelimelerin değişimlerini incelemek gibi amaçlar için kullanır.
- ▶ İlgili dilde, belirlenmiş zaman aralığında yayımlanmış metinlerden belirli büyüklükte örnekler alınarak oluşturulur.
- ▶ Metinler değişik alanlardan seçilerek dilin değişik alanlardaki kullanımı örneklenir.
 - ▶ Her konu başlığı altında alt başlık oluşturulur.
 - ▶ Konu ve alt konuların dil temsilindeki ağırlığı hesaplanır.
 - ▶ Bu ağırlık toplumun konuya verdiği öneme bağlıdır.
- ▶ Dengeli derlemde; derlem boyutunun 2.500.000 kelime ve örnek metin boyutunun 5000 kelime olması beklenir.

Dengeli Derlem

- Dengeli bir derlemin ana amacı, farklı konu alanlarında, farklı kaynaklardan alınan metinlerin dengeli bir şekilde temsil edilmesidir. Bu, derlemin genel kullanımına uygun olmasını sağlar ve örnekleme dayalı çalışmalarda doğru sonuçlar elde etmek için önemlidir.
- Derlem boyutunun 2.500.000 kelime olması, oldukça büyük bir veri kümesi anlamına gelir ve geniş bir yelpazede kullanım için uygun olabilir. Bununla birlikte, derlemin kullanım amacına ve hedefine bağlı olarak daha küçük veya daha büyük boyutlarda da olabilir.
- Örnek metin boyutunun 5000 kelime olması, dengeli bir derlemin standart boyutu olarak kabul edilebilir. Bu boyut, farklı metin kaynaklarının eşit temsili için yeterli olabilir ve örnekleme dayalı çalışmalarda da kullanıma uygun olabilir. Ancak, kullanım amacına bağlı olarak daha küçük veya daha büyük örnek metin boyutları da olabilir.
- Özetle, dengeli bir derlemin boyutu ve örnek metin boyutu, derlemin kullanım amacına, hedefine ve veri kaynaklarına bağlı olarak değişebilir. Ancak, genel olarak, derlemin dengeli olması, farklı konu alanlarını temsil etmesi ve örnekleme dayalı çalışmalarda doğru sonuçlar elde etmek için önemlidir.

Dengeli Derlem

Örneğin

Metin Türü	Etki Oranı (%)	Metin Sayısı	Etki Oranı (%)
Basından yazılar			18
Haberler	5	25	
• Siyasi	2	10	
• Spor	2	10	
• Ekonomi	1	5	
Eğitsel Yayınlar			37
Bilimsel Yayınlar	7	35	
• Fen Bilimleri	5	25	
• Matematik	2	10	
Edebi yazılar		65	28
• Roman	78	40	
• Hikaye	5	25	

Dengeli Derlem

- Dengeli derlemin en önemli özelliklerinden biri, belirli bir dilbilgisi yapısı ve dil kullanımı özelliklerine uygun olarak seçilmiş eşit sayıda örneklerden oluşmasıdır. Metin sayısı, dengeli derlemin boyutunu belirler ve dilin farklı alanlarını kapsayan örneklerin sayısını belirler. Genellikle dengeli derlem oluştururken, belirli bir dilin farklı alanlarından ve kaynaklarından belirli sayıda metin örneği toplanır.
- Etki oranı ise, dengeli derlemin belirli bir alan veya kaynak grubundaki metinlerin toplam sayısı içindeki payını ifade eder. Örneğin, bir dengeli derlemin 1000 metin örneği içerdiğini varsayalım. Bu derlemin belirli bir kaynak grubundan (örneğin gazete makaleleri) toplanmış 200 metin örneği varsa, etki oranı $200/1000 = 0.2$ (veya %20) olacaktır.
- Dengeli derlem oluştururken, farklı alan ve kaynak gruplarından eşit sayıda örnek almak, bir dilin farklı kullanım şekillerinin örneklerini içermesini sağlar. Bu, doğal dil işleme modellerinin daha geniş bir dil anlayışına sahip olmasını sağlar ve daha iyi sonuçlar elde etmelerine yardımcı olabilir.

Dengesiz Derlem

- ▶ Daha çok bilişimcilerin kullandığı derlem türüdür.
- ▶ Güncel metin örnekleri ile oluşturulması yeterlidir.
- ▶ Çok sayıda metin içerir.
- ▶ Derlemdeki metinler farklı alanlarda seçilmelidir ancak dengelenmeleri çok önemli değildir.
- ▶ Bazı sıradışı sözcükler için daha kullanışlıdır.
- ▶ Bilişimciler;
 - ▶ Dilin ses özelliklerinin, biçim bilimsel özelliklerinin çıkarılması, kelimelerin niteliklerinin ve anlamlarının belirlenmesi gibi birçok alanda kullanırlar.
- ▶ Derlemin hazırlanmasında; internetteki yayınlardan, kitap ve belgelerden faydalanılabilir.

Örnek

- ▶ "Öğrenciler, öğretmenleri tarafından verilen ödevleri yapmak zorundadırlar."
- ▶ Yukarıdaki metne eğitim etiketi atayalım.

Örnek

► # Kütüphaneleri yüklüyoruz

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

- Bu kodda, NLTK (Natural Language Toolkit) kütüphanesi içindeki bazı modüllerin (nltk.tokenize ve nltk.corpus) kullanımına yönelik import işlemleri gerçekleştiriliyor. İlk satırda nltk kütüphanesi yükleniyor ve sonraki iki satırda word_tokenize ve stopwords modülleri yükleniyor.
- word_tokenize: metinleri parçalamak (tokenize) için kullanılır. Bu modül, bir cümleyi veya bir paragrafı kelimelere ayırır.
- stopwords: stop words (durak kelimeler) listesine erişim sağlar. Bu modül, metindeki yaygın ve anlamsız kelimeleri filtrelemek için kullanılır.

Örnek

- ▶ # Örnek metnimizi tanımlıyoruz

```
ornek_metin = "Öğrenciler, öğretmenleri tarafından verilen ödevleri yapmak zorundadırlar."
```

- ▶ # Metni küçük harfe dönüştürüyoruz

```
ornek_metin = ornek_metin.lower()
```

```
ornek_metin
```

- ▶ Metin işleme (NLP) uygulamalarında büyük/küçük harf farklılıkları göz önünde bulundurulmaz. Bu nedenle, metni küçük harflere dönüştürmek, etiketleme yaparken tutarlılık sağlar ve işlem kolaylığı sağlar. Örneğin, "Öğrenciler" ve "öğrenciler" aynı kelime olarak kabul edilir ve aynı etiketi alır. Bu, programın daha doğru sonuçlar vermesine ve daha az hata yapmasına olanak tanır.

Örnek

- ▶ # Metni kelimelere ayırıyoruz

```
import nltk

nltk.download('punkt')

# Metni kelimelere ayırıyoruz
kelimeler = word_tokenize(ornek_metin)
kelimeler
```

- ▶ Bu kod, NLTK kütüphanesinden "punkt" modülünü indirir. Bu modül, metinleri kelimelere veya cümlelere ayırmak için kullanılan bir araçtır.
- ▶ Ardından, `ornek_metin` değişkenindeki `metin` `word_tokenize()` fonksiyonu kullanılarak kelimelere ayrılır. Bu, metnin her bir kelimesinin ayrı ayrı işleme tabi tutulabilmesini sağlar.

Örnek

- ▶ # Stop words (durak kelimeler) listesini yüklüyoruz

```
nltk.download('stopwords')

# Stop words (durak kelimeler) listesini yüklüyoruz
durak_kelimeler = stopwords.words('turkish')
durak_kelimeler
```

- ▶ Türkçe dilinde yaygın olarak kullanılan stop words (durak kelimeler) listesini yükledik. Stop words listesi, dil işleme uygulamalarında genellikle işe yaramayan yaygın kelimelerdir. Örneğin, "bir", "ve", "ya", "şu", "bu" gibi kelimeler stop words olarak kabul edilir ve genellikle göz ardı edilir.
- ▶ stopwords.words('turkish') ifadesi, NLTK kütüphanesinde yer alan stopwords modülünden, Türkçe dilindeki stop words listesini çağırır ve durak_kelimeler değişkenine atar. Böylece, stop words listesi etiketleme işlemi için kullanılabilir.

Örnek

- ▶ # Durak kelimeleri metinden çıkarıyoruz

```
kelimeler = [kelime for kelime in kelimeler if not kelime in durak_kelimeler]  
kelimeler
```

- ▶ Bu işlem, metindeki stop words (durak kelimeler) olarak adlandırılan yaygın kullanılan kelimeleri (örneğin: "ve", "ya", "ama", "gibi" vb.) çıkararak metnin işlenmesini kolaylaştırmak ve daha anlamlı sonuçlar elde etmek için yapılır.
- ▶ Bu kelimeler, bir metnin anlamını etkilemeden sık sık kullanıldıkları için, özellikle doğal dil işleme işlemlerinde genellikle çıkarılırlar. Bu şekilde, metindeki daha az yaygın ve anlamlı kelimelerin belirlenmesi daha kolay hale gelir.

Örnek

- ▶ # Metnin etiketini belirliyoruz

```
konu_alani_etiketi = "Eğitim Alanı"
```

- ▶ # Etiketle birlikte metni yazdırıyoruz

```
print("{}: {}".format(konu_alani_etiketi, " ".join(kelimeler)))
```

Eğitim Alanı: öğrenciler , öğretmenleri tarafından verilen ödevleri yapmak zorundadırlar .

```
# Kütüphaneleri yüklüyoruz
import nltk

from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

# Örnek metnimizi tanımlıyoruz
ornek_metin = "Öğrenciler, öğretmenleri tarafından verilen ödevleri yapmak zorundadırlar."

# Metni küçük harfe dönüştürüyoruz
ornek_metin = ornek_metin.lower()
ornek_metin

import nltk
nltk.download('punkt')

# Metni kelimelere ayırıyoruz
kelimeler = word_tokenize(ornek_metin)
kelimeler

nltk.download('stopwords')

# Stop words (durak kelimeler) listesini yüklüyoruz
durak_kelimeler = stopwords.words('turkish')
durak_kelimeler

# Durak kelimeleri metinden çıkarıyoruz
```

Hazır Derlem Örnekleri

- ▶ Penn Treebank
- ▶ Brown derlemi
- ▶ Santral derlemi
- ▶ ODTÜ Türkçe derlemi
- ▶ Türkçe Ulusal Derlemi



Derlem Örnekleri: Penn Treebank

- Bu derlem, ABD İngilizcesi için geliştirilmiş ve yaygın olarak kullanılan bir derlemidir. 4 milyondan fazla kelime içerir ve gazete makaleleri, dergiler ve edebi eserler gibi farklı kaynaklardan oluşur.

Derlem Örnekleri: Brown Derlemi

- Brown derlemi, İngilizce için bir başka popüler derlemdir. 15 milyon kelime içerir ve 500'den fazla kaynaktan oluşur. Bu derlemin özelliği, farklı kaynaklardan oluşması ve bu nedenle çeşitli İngilizce varyasyonlarını içermesidir.

Derlem Örnekleri

Brown Derlemi;

Metin türü	Metin sayısı	Oran %	Metin türü	metin sayısı	Oran %
<i>Haberler</i>	44	8,8	<i>Güncel töre</i>	48	9,6
• Siyasi			• Kitap		
• Spor			• Dergi		
• Toplumsal			<i>Anı</i>	75	15
• Spor haberleri			• Kitap		
• Ekonomi			• Dergi		
• Kültürel			<i>Çeşitli</i>	30	6
<i>Köşe yazıları</i>	27	5,4	• Kamu belgeleri		
• Köşe yazıları			• Kurumsal raporlar		
• Günlük yorumlar			• Sanayi raporları		
• Başyazı			• Üniversite katalogları		
<i>Yorumlar</i>	17	3,4	• Sanayi yayınları		
• Tiyatro			<i>Eğitim</i>	80	16
• Kitap			• Doğa bilimleri		
• Müzik			• Eczacılık		
• Dans			• Matematik		
<i>Dini yayınlar</i>	17	3,4	• <i>Sosyal bilimler</i>		
• Kitaplar			• Siyasal bilgiler		
• Dergiler			• Uygarlık		
• Risaleler			• Teknoloji		
<i>Beceri ve uğraşlar</i>	36	7,2	<i>Kurgu</i>	29	5,6
• Kitaplar			• Roman		
• Dergiler			• Hikâye		

Derlem Örnekleri: Santral Derlemi

- ▶ Amerikan İngilizcesi
- ▶ 1990 yılı
- ▶ Telefon santrali üzerinde geçen telefon konuşmalarından derlenmiş
- ▶ 6'şar dakikalık 2430 konuşma kaydının yazılı biçimi.
- ▶ Dilin sözel incelemeleri ve konuşmayı yazıya aktarma çalışmaları için kullanılmakta

Derlem Örnekleri



- ▶ ODTÜ Türkçe Derlemi
 - ▶ ODTÜ Enformatik Enstitüsü
 - ▶ 2 milyon kelime
 - ▶ Metinler 8 farklı konudan
 - ▶ Metin büyüklükleri 2000 kelime
 - ▶ Güncel dil

Metin türü	Metin sayısı	Oran %	Metin türü	metin sayısı	Oran %
Roman	123	15,63	Gezi	19	2,41
Öykü	114	14,49	Söyleşi	7	0,89
Araştırma inceleme	49	6,23	Anı-günce-hatıra	18	2,29
Makale	38	4,83	Haber	419	53,24

Derlem Örnekleri

- Türkçe Ulusal Derlemi;
 - Mersin Üniversitesi
 - 50 milyon sözcük
 - Dengeli derlem

Metin türü	Oran %	Metin türü	Oran %
Kurgusal düz yazı	19	Düşünce ve inanç	4
Toplum bilimleri	16	Dünya sorunları	20
Sanat	7	Uygulamalı bilimler	8
Ticaret ve finans	8	Doğa ve temel bilimler	1
		Diğer	14

Kullanılan Etiketler

Etiket		Etiket	Açıklama	Etiket	Açıklama
VB	Eylem	AJ	Önad	CJ	Bağlaç
NN	Ad	AV	Belirteç	ON	Yansıma
PN	Adıl	PP	İlgeç	NP	Özel ad
NB	Sayı	ITJ	Ünlem	AB	Kısaltma
				MI	Soru eki

Etiketler

- ▶ Doğal Dil işlemede metnin her kelimesine bir etiket vermenin "part-of-speech tag" (POS tag) olduğundan söz etmiştik.
- ▶ Bu etiketler, kelimenin dil bilgisi yapısını ve cümledeki rolünü belirlemek için kullanılır.
- ▶ Aşağıda en yaygın kullanılan etiketlerin listesi ve açıklamaları verilmiştir:

- CC: bağlaç (coordinating conjunction)
- CD: sayısal ifade (cardinal number)
- DT: tanımlayıcı belirteç (determiner)
- EX: "there" olarak kullanılan sözcükler (existential there)
- FW: yabancı sözcük (foreign word)
- IN: bağlaç, alt kategorileri şunlardır: alt bölümler: alt bölümleri şunlardır: alt bölümler: alacakaranlık (preposition/subordinating conjunction), sonuna kadar (postposition)

- JJ: sıfat (adjective)
- JJR: sıfat, üstünlük derecesi (adjective, comparative)
- JJS: sıfat, en üstün derecesi (adjective, superlative)
- LS: liste işareti (list item marker)
- MD: yardımcı fiil (modal)
- NN: isim (noun), tekil
- NNS: isim, çoğul (noun, plural)
- NNP: özel isim (proper noun), tekil
- NNPS: özel isim, çoğul (proper noun, plural)
- PDT: tanımlayıcı belirteç (predeterminer)

Etiketler

- RB: zarf (adverb)
- RBR: zarf, üstünlük derecesi (adverb, comparative)
- RBS: zarf, en üstün derecesi (adverb, superlative)
- RP: edat (particle)
- SYM: sembol (symbol)
- TO: "to" sözcüğü (to)
- UH: ünlemler (interjection)
- VB: fiil (verb), taban hali
- VBD: fiil, geçmiş zaman (verb, past tense)
- VBG: fiil, -ing hali (verb, gerund/present participle)
- VBN: fiil, edilgen çatı, geçmiş zaman (verb, past participle)
- VBP: fiil, üçüncü tekil şahıs, şimdiki zaman (verb, non-3rd person singular present)
- VBZ: fiil, üçüncü tekil şahıs, şimdiki zaman (verb, 3rd person singular present)
- WDT: tanımlayıcı belirteç (wh-determiner)

Örnek: Brown Corpus'undan ilk 50 kelimeyi alarak etiketleme yapalım

```
import nltk
nltk.download('brown') # Brown Corpus'u indirin

# makine öğrenimi tabanlı bir etiketleyici
nltk.download('averaged_perceptron_tagger')

from nltk.corpus import brown

# Brown Corpus'tan bir metin örneği alalım
text = brown.words(categories='news')[0:50]

# Metin örneğini etiketleyelim
tagged_text = nltk.pos_tag(text)

# Etiketli metin örneğini yazdıralım
print(tagged_text)
```

```
import nltk
nltk.download('brown') # Brown Corpus'u indirin
nltk.download('averaged_perceptron_tagger')

from nltk.corpus import brown

# Brown Corpus'tan bir metin örneği alalım
text = brown.words(categories='news')[0:50]
text

# Metin örneğini etiketleyelim
tagged_text = nltk.pos_tag(text)

# Etiketli metin örneğini yazdıralım
print(tagged_text)
```

Örnek: treebank veri kümesindeki cümle sayısını hesaplayıp ekrana yazdıralım

```
import nltk
nltk.download('treebank')

from nltk.corpus import treebank

num_sentences = len(treebank.sents())

print("Treebank veri kümesindeki cümle sayısı:", num_sentences)
```

```
import nltk

nltk.download('treebank')

from nltk.corpus import treebank

num_sentences = len(treebank.sents())

print("Treebank veri kümesindeki cümle sayısı:", num_sentences)
```


Örnek: treebank veri kümesindeki cümleleri ekrana yazdıralım

```
import nltk
nltk.download('treebank')

from nltk.corpus import treebank

for sentence in treebank.sents():
    print(sentence)
```

```
import nltk

nltk.download('treebank')

from nltk.corpus import treebank

for sentence in treebank.sents():
    print(sentence)
```

Örnek: treebank veri kümesindeki ilk cümleyi etiketleyelim

```
import nltk
nltk.download('treebank')

from nltk.corpus import treebank

# Treebank'taki ilk cümleyi al
sentence = treebank.sents()[0]

# Cümleyi etiketle
tagged_sentence = nltk.pos_tag(sentence)

print(tagged_sentence)
```

```
import nltk
nltk.download('treebank')

from nltk.corpus import treebank

# Treebank'taki ilk cümleyi al
sentence = treebank.sents()[0]

# Cümleyi etiketle
tagged_sentence = nltk.pos_tag(sentence)

print(tagged_sentence)
```

Örnek: Yazdığımız bir cümleyi etiketleyelim

```
import nltk

# örnek bir cümle
sentence = "Bu bir örnek cümledir."

# cümleyi kelime parçalarına ayırma (tokenization)
tokens = nltk.word_tokenize(sentence)

# cümle etiketleme (part-of-speech tagging)
pos_tags = nltk.pos_tag(tokens)

print(pos_tags)
```

```
import nltk

# örnek bir cümle
sentence = "Bu bir örnek cümledir."

# cümleyi kelime parçalarına ayırma (tokenization)
tokens = nltk.word_tokenize(sentence)

# cümle etiketleme (part-of-speech tagging)
pos_tags = nltk.pos_tag(tokens)

print(pos_tags)
```

Kaynaklar

- Eşref Adalı, Türkçe Doğal Dil İşleme, Akçağ Yayınları

