

```
In [ ]: import nltk
```

1. nltk.corpus.PlaintextCorpusReader

Para qué sirve: Esta función permite leer archivos de texto plano y acceder a ellos como un corpus.

Sintaxis: PlaintextCorpusReader(root, fileids, word_tokenizer=nltk.word_tokenize)

Ejemplo:

```
In [ ]: from nltk.corpus import PlaintextCorpusReader

corpus_root = '/ruta/a/los/archivos/txt'
wordlists = PlaintextCorpusReader(corpus_root, '.*\.txt')
```

2. nltk.tokenize.sent_tokenize

Para qué sirve: Divide un texto en oraciones utilizando un algoritmo de tokenización de oraciones.

Sintaxis: sent_tokenize(text)

Ejemplo:

```
In [ ]: from nltk.tokenize import sent_tokenize

text = "Este es un ejemplo de texto. ¡Espero que te sea útil!"
sentences = sent_tokenize(text)
```

3. nltk.tokenize.word_tokenize

Para qué sirve: Divide un texto en palabras utilizando un algoritmo de tokenización de palabras.

Sintaxis: word_tokenize(text)

Ejemplo:

```
In [ ]: from nltk.tokenize import word_tokenize

text = "Este es un ejemplo de texto."
words = word_tokenize(text)
```

4. nltk.corpus.gutenberg

Para qué sirve: Ofrece acceso al corpus de textos de Project Gutenberg.

Sintaxis: `nltk.corpus.gutenberg.fileids()`

Ejemplo:

```
In [ ]: from nltk.corpus import gutenberg  
  
file_ids = gutenberg.fileids()
```

5. nltk.corpus.WebText

Para qué sirve: Proporciona acceso al corpus de textos web, que contiene mensajes de chats y foros.

Sintaxis: `nltk.corpus.webtext.fileids()`

Ejemplo:

```
In [ ]: from nltk.corpus import webtext  
  
file_ids = webtext.fileids()
```

6. nltk.corpus.treebank

Para qué sirve: Ofrece acceso al corpus de Treebank, que contiene textos etiquetados con partes del discurso y árboles de sintaxis.

Sintaxis: `nltk.corpus.treebank.fileids()`

Ejemplo:

```
In [ ]: from nltk.corpus import treebank  
  
file_ids = treebank.fileids()
```

7. nltk.corpus.ConllCorpusReader

Para qué sirve: Lee archivos en formato CoNLL (formato tabular utilizado en el procesamiento del lenguaje natural) y los convierte en corpus.

Sintaxis: `ConllCorpusReader(root, file_pattern, columntypes, separator='\t')`

Ejemplo:

```
In [ ]: from nltk.corpus import ConllCorpusReader
```

```
corpus_root = '/ruta/a/los/archivos/conll'  
conll_reader = ConllCorpusReader(corpus_root, '.*\.conll', columntypes=('words', 'p
```

8. nltk.corpus.BracketParseCorpusReader

Para qué sirve: Lee archivos en formato de árboles de sintaxis bracketed y los convierte en corpus.

Sintaxis: BracketParseCorpusReader(root, file_pattern)

Ejemplo:

```
In [ ]: from nltk.corpus import BracketParseCorpusReader  
  
corpus_root = '/ruta/a/los/archivos/parse'  
parse_reader = BracketParseCorpusReader(corpus_root, '.*\.txt')
```

9. nltk.corpus.XMLCorpusReader

Para qué sirve: Lee archivos XML y los convierte en corpus.

Sintaxis: XMLCorpusReader(root, fileids, wrap_etree=False)

Ejemplo:

```
In [ ]: from nltk.corpus import XMLCorpusReader  
  
corpus_root = '/ruta/a/los/archivos/xml'  
xml_reader = XMLCorpusReader(corpus_root, '.*\.xml')
```

10. nltk.corpus.CategorizedPlaintextCorpusReader

Para qué sirve: Lee archivos de texto y los organiza en categorías.

Sintaxis: CategorizedPlaintextCorpusReader(root, fileids, cat_pattern, encoding='utf8')

Ejemplo:

```
In [ ]: from nltk.corpus import CategorizedPlaintextCorpusReader  
  
corpus_root = '/ruta/a/los/archivos/categorizados'  
reader = CategorizedPlaintextCorpusReader(corpus_root, r'(?!\.)*\.txt', cat_patter
```

11. nltk.probability.FreqDist

Para qué sirve: Calcula la frecuencia de cada elemento en una lista.

Sintaxis: FreqDist(tokens)

Ejemplo:

```
In [ ]: from nltk.probability import FreqDist

tokens = ['apple', 'banana', 'apple', 'orange', 'banana']
freq_dist = FreqDist(tokens)
```

12. nltk.collocations.BigramCollocationFinder

Para qué sirve: Encuentra las colocaciones de bigramas en un texto.

Sintaxis: BigramCollocationFinder.from_words(tokens)

Ejemplo:

```
In [ ]: from nltk.collocations import BigramCollocationFinder

tokens = ['this', 'is', 'a', 'test', 'sentence']
bigram_finder = BigramCollocationFinder.from_words(tokens)
```

13. nltk.translate.bleu_score.corpus_bleu

Para qué sirve: Calcula el puntaje BLEU para una traducción automática en un corpus.

Sintaxis: corpus_bleu(list_of_references, hypotheses)

Ejemplo:

```
In [ ]: from nltk.translate.bleu_score import corpus_bleu

references = [['this', 'is', 'a', 'test'], ['another', 'example']]
hypotheses = [['this', 'is', 'a', 'test'], ['example', 'of', 'another']]
score = corpus_bleu(references, hypotheses)
```