

Installere NLTK

Vi installerer NLTK, slik som andre Python-pakker, med `pip`. Det spiller ingen rolle i hvilken mappe du "står" når du installerer Python-pakker.

Dersom du vanligvis kjører Python-kode med `python`, bruker du kun `pip` for å installere pakker, og dersom du vanligvis kjører Python-kode med `python3`, bruker du `pip3` for å installere pakker.

Dersom du er usikker kan du prøve deg fram med `python --version` og `python3 --version`. Bruk den som gir deg versjon 3!

```
~ python --version  
Python 2.7.16
```

```
~ python3 --version  
Python 3.9.7
```

På min maskin må jeg bruke `python3` for å kjøre Python versjon 3. Altså bruker jeg `pip3` videre.

Jeg installerer NLTK med `pip3 install nltk`.

```
~/gruppetime/in1140 > pip3 install nltk
DEPRECATION: Configuring installation scheme with distutils config files is deprecated and will no
in the near future. If you are using a Homebrew or Linuxbrew Python, please see discussion at https
m/Homebrew/homebrew-core/issues/76621
Collecting nltk
  Downloading nltk-3.6.3-py3-none-any.whl (1.5 MB)
    | ██████████ | 1.5 MB 5.3 MB/s
Requirement already satisfied: click in /opt/homebrew/lib/python3.9/site-packages (from nltk) (8.0.
Requirement already satisfied: regex in /opt/homebrew/lib/python3.9/site-packages (from nltk) (2021
Requirement already satisfied: tqdm in /opt/homebrew/lib/python3.9/site-packages (from nltk) (4.61.
Requirement already satisfied: joblib in /opt/homebrew/lib/python3.9/site-packages (from nltk) (1.0
Installing collected packages: nltk
DEPRECATION: Configuring installation scheme with distutils config files is deprecated and will no
in the near future. If you are using a Homebrew or Linuxbrew Python, please see discussion at https
m/Homebrew/homebrew-core/issues/76621
➔ Successfully installed nltk-3.6.3
```

Hvor mye tekst som skrives ut under installasjonen varierer litt, men det viktigste er at det står mot bunnen av utskriften at installasjonen var vellykket.

Gratulerer, du har nå installert NLTK! Men for å få ordentlig nytte av NLTK, må vi laste ned litt "NLTK-data".

For å gjøre det må vi først åpne Python-interpreten i terminalen. Dette gjør du med `python` eller `python3`, den som gir deg versjon 3. Du vet at Python-interpreten er åpen, når det står `>>>` på starten av hver linje.

```
~/gruppetime/in1140 > python3
Python 3.9.7 (default, Sep  3 2021, 04:31:11)
[Clang 12.0.5 (clang-1205.0.22.9)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

Inne i interpreten skriver vi først `import nltk`.

```
>>> import nltk
>>>
```

Dersom neste linje er blank og den ikke kommer med noen feilmelding, er NLTK installert riktig.

Når vi har importert NLTK, kan vi laste ned pakkene vi trenger. Skriv inn `nltk.download()`.

```
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

Dette vil åpne et nytt vindu. Det kan ta litt tid før dette vises. Du får ikke bruke Python-interpreten eller terminaløkta igjen før du lukker dette vinduet.

Velg samlingen
book og deretter
Download.

*Gå til siste side hvis du
bruker MacOS og får
en feilmelding her.*

Identifier	Name	Size	Status
all	All packages	n/a	not installed
all-corpora	All the corpora	n/a	not installed
all-nltk	All packages available on nltk_data gh-pages branch	n/a	not installed
book	Everything used in the NLTK Book	n/a	not installed
popular	Popular packages	n/a	not installed
tests	Packages for running tests	n/a	not installed
third-party	Third-party data packages	n/a	not installed

Download Refresh

Server Index:

Download Directory:

Etter vellykket nedlastning bør samlingen markeres med grønt.

Identifier	Name	Size	Status
all	All packages	n/a	partial
all-corpora	All the corpora	n/a	partial
all-nltk	All packages available on nltk_data gh-pages branch	n/a	partial
book	Everything used in the NLTK Book	n/a	installed
popular	Popular packages	n/a	partial
tests	Packages for running tests	n/a	partial
third-party	Third-party data packages	n/a	not installed

Download Refresh

Server Index:

Download Directory:

Finished downloading collection 'book'.

Når vi lukker vinduet får vi bruke Python-interpreten igjen. For å teste om dataen ble lastet ned skriver vi inn `from nltk.corpus import inaugural`. Dette er en korpuspakke som inneholder amerikanske presidenttaler.

```
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
True
>>> from nltk.corpus import inaugural
>>>
```

Dersom du ikke får noen feilmelding, kan du starte på ukesoppgavene. Lykke til!

For å lukke Python-interpreten, slik at vi kan bruke den "vanlige" terminaløkta igjen, skriver vi `exit()`.

Merk at vi kun trenger å bruke `nltk.download()` når vi skal laste ned ny data. Vi trenger ikke å skrive det i hvert program vi lager.

`import nltk` må du skrive i hvert program som du vil bruke NLTK i.

Løsning for MacOS:

1. Lukk vinduene som dukket opp og lukk terminalen du bruker for å installere NLTK.
2. Åpne Finder (filutforsker).
3. Søk etter `Install Certificates.command` (pass på at du søker på "denne Mac-en", ikke bare "nylige").
4. Dobbelklikk på fila som dukker opp. Dette er et skript inkludert i Python-installasjonen som fikser problemet.
5. Et terminalvindu åpnes. Når det står i bunnen at prosessen er utført, kan du lukke vinduet.
6. Start en ny terminaløkt og følg denne guiden fra side 3.