

# Python og NLTK i INF2820, V2017

---

## Bakgrunnskunnskaper fra INF1820

INF2820 bygger på en del av [INF1820 - Introduksjon til språk- og kommunikasjonsteknologi](#). Vi vil bl.a. bruke noen av de samme redskapene i de to emnene: NLTK og Python. Studenter som ikke er kjent med disse redskapene, bør sette seg inn i dem så raskt som mulig.

Siden <http://www.uio.no/studier/emner/matnat/ifi/INF1820/v16/pensumliste/index.html> viser hvilke deler av NLTK som var pensum i INF1820 i fjor. Du finner også flere detaljer om hva som ble dekket i INF1820 på <http://www.uio.no/studier/emner/matnat/ifi/INF1820/v16/timeplan/index.html> under Gruppeundervisning. Vi vil bygge på deler av kapittel 1 og 2 i NLTK-boka (men ikke kap. 5). Du anbefales å arbeide deg gjennom følgende

- Kapittel 1 fra NLTK-boka, særlig seksjon 1.2 og 1.4.
- Gjøre oppgavene fra seksjon 1.8: 1, 2, 3, 8, 9, 10, 13, 16
- Seksjon 2.3

Når du arbeider med NLTK bør du sitte ved skjermen med et Python-vindu og selv utføre det som står. Det er det du lærer av. Du vil også tjene på å arbeide mer med programmering i Python, seksjonene 4.1-4.4 i NLTK-boka. På siden med INF1820-pensum finner du også peker til mere stoff om Python,

- [How to think like a computer scientist: Learning with Python](#) (anbefalt)
- The official [Python documentation](#)

## Valg av Python-oppsett

### Python 2.x vs python 3.x.

Python finnes i to varianter python 2.x og python 3.x. De er ikke helt kompatible, dvs ikke all kode fra python 2.x vil kjøre i python 3.x. Derfor er det en del som holder fast ved python 2.7 (siste versjon av 2.x). Vi vil bruke python 3.x i INF2820 dette semesteret. Det gjør vi fordi INF1001 bruker python 3.x, og fordi NLTKs siste versjon er lagt til rette for python 3.x. Siste versjon av NLTK virker også med python 2.x, men dokumentasjonen er utformet for python 3.x.

På IFIs linux-maskiner vil kommandoen *python* gi deg python 2.7, mens kommandoen *python3* gir deg python 3.

### NLTK

IFIs python3 er satt opp med NLTK, slik at kommandoen  
import NLTK  
skal virke.

Når du arbeider på IFIs Linux-maskiner trenger du ikke laste ned NLTK-dataene til ditt eget område. Dataene ligger på /projects/nlp/nltk\_data. Hvis du legger følgende linje til .bashrc-filen din og så logger ut og logger på igjen, vil du ha permanent tilgang.  
export NLTK\_DATA=/projects/nlp/nltk\_data

### Python/Ipypthon/Idle

Når du sitter på en av IFIs Linux-maskiner kan du gi kommandoen *python3*. Du kommer da rett inn i et Python-shell hvor du kan arbeide interaktivt. Python er forskjellig fra Java. I Java skriver vi et program, deretter kompilerer vi det, og så kan vi kjøre det. Python kompilerer ikke programmene, men interpreterer dem og legger opp til interaktivt arbeid i utviklingsfasen. Det gjør også NLTK(-dokumentasjonen), og vi vil bruke denne arbeidsformen i INF2820.

Alternativt til kommandoen *python* kan du gi en av kommandoene *idle* eller *ipython*. Disse gir deg også et Python-shell, men de gir også en del tilleggsfunksjonalitet som automatisk indentering (der du trenger 4 blanke), og de kan fullføre kommandonavn for deg når du har skrevet første delen. Denne funksjonen er uvurderlig når vi skal arbeide med NLTK. Foreløpig fører disse til python 2 på IFIs maskiner, men vi arbeider med å få enkelt oppsett for å bruke dem med python 3.

### Python på hjemmemaskin

For python på egen maskin er det en del anbefalinger på

<http://www.uio.no/studier/emner/matnat/ifi/INF1100/h16/ressurser/installering.html>

Men merk at de bruker python 2.7, så du må gjøre tilpasninger. Vi er også interessert i andre pakker enn dem, først og fremst NLTK. For et system som gjør det lett å installere nye pakker, anbefales Anaconda.