

Oppgaver

INF1820

22/02/2017

Ordklassetagging

Under følger en liste av setninger henta fra NoWaC (Norwegian Web as Corpus). Bestem et tagsett, feks basert på ordklassene fra forelesningsfoilene, og utfør manuell ordklassetagging av orda.

1. Både Orangeriet og Ostehuset har etablert seg i østre bydel
2. Enten vi tror det eller ei , har det alltid vært en del av vår kultur
3. Vi har i lengre tid sett etter pent utemøblemang til hagen vår
4. Den svenske kirken går mye lenger enn den norske
5. Jeg har fått inn sykt mange videoer , så mailen min er stapp full nå
6. For jeg vet ikke hva jeg skulle gjort uten dere
7. Selv bryr jeg meg fint lite om hva andre tjener
8. For det andre får vi tro at Gud også vil ta seg av det som er vondest og mest uforståelig i livet vårt
9. Høst er tid for inneaktiviteter som strikking og hygge
10. Dersom akselbruddet hadde skjedd da toget kjørte i toppfart , kunne det gått forferdelig galt

NLTK

Kapittel 2 i NLTK-boka omhandler tekstkopora. Gjør oppgave 15, 17 (her kan det være smart å ta i bruk Stopwordskorpus), 18, 21 og 22.