

spaCy for Norwegian Nynorsk

About spaCy

General Introduction

spaCy is a library for Natural Language Processing in Python and is widely used for language technological applications. Read more about spaCy [here](#).

Data

NorNE

[NorNE](#) adds named entity annotations on top of the [Norwegian Dependency Treebank](#) and was created as a collaboration between Schibsted Media Group, Språkbanken at the National Library of Norway and the Language Technology Group at the Department of Informatics, University of Oslo.

The NorNE corpus is published under the same [license](#) as the Norwegian Dependency Treebank.

NB-BERT

[NB-BERT-base](#) is a general BERT-base model built on the large digital collection at the National Library of Norway.

NB-BERT is based on the same structure as the [BERT Cased multilingual model](#), and is trained on a wide variety of Norwegian texts (both Bokmål and Nynorsk) from the last 200 years.

Licence

[The MIT License \(MIT\)](#)

Questions and feedback

Questions, comments and any kind of feedback are very welcome. We also welcome suggestions for improvement, corrections and modifications, and would like to be informed about derived resources (with an open license), which may be of interest to the speech technology community.

To get in touch with us, please use sprakbanken@nb.no.

[NORWEGIAN TEXT OVERLEAF]

spaCy for nynorsk

Om spaCy

Generell introduksjon

spaCy er eit bibliotek for NLP i Python, og er mykje brukt i språkteknologiske applikasjonar. Les meir om spaCy [her](#).

Data

NorNE

[NorNe](#) (Norwegian Named Entities) vart laga som eit samarbeid mellom Schibsted Media Group, Språkteknologigruppa ved Institutt for informatikk på Universitetet i Oslo og Språkbanken på Nasjonalbiblioteket. NorNE er eit tekstkorpus basert på dei same tekstene som [Norsk Dependenstrebank](#), men er i tillegg tagga med namneentitetar.

NorNE-korpuset er publisert under den same [lisensen](#) som Norsk Dependenstrebank.

NB-BERT

[NB-BERT-base](#) er ein generell BERT-base-modell som er bygd med utgangspunkt i den store digitale samlinga ved Nasjonalbiblioteket.

Modellen er basert på same struktur som [BERT Cased multilingual model](#), og er trent på eit breitt utval norske tekster – både bokmål og nynorsk – frå dei siste 200 åra.

Lisens

[The MIT License \(MIT\)](#)

Spørsmål og attendemeldingar

Vi set pris på spørsmål, kommentarar og alle slags attendemeldingar. Vi tek òg gjerne imot forslag til forbetringar og endringar. Informer oss gjerne om ressursar som vert utvikla med denne ressursen som utgangspunkt og som kan vere av interesse for andre.

Kontakt oss på e-post, sprakbanken@nb.no.

[ENGELSK TEKST PÅ ANDRE SIDA]