

Svensk text

SND-ID: ext0278-1.

Skapare/primärforskare

[Peter Dahlgren](#) - Göteborgs universitet

Forskningshuvudman

[Göteborgs universitet](#) - Institutionen för journalistik, medier och kommunikation (JMG)

Beskrivning

Samling med språkresurser på svenska speciellt anpassat för att snabbt och enkelt kunna läsas in av programspråk som Python, R eller dylikt. Bland språkresurserna finns namn på kvinnor (förnamn), män (förnamn), städer, kommuner, län, huvudstäder, länder, nationaliteter, yrken, myndigheter, massmedier med mera.

Syftet är att samla svenska resurser som är fria att använda, finns i flera öppna standardiserade format (exempelvis CSV och JSON), har en enkel datastruktur som enkelt kan maskinläsas, består av små filer utan onödigt krångel och har exempelkod (R och Python) för att snabbt kunna användas.

Materialet utökas och uppdateras kontinuerligt under öppna licenser.

Dataformat / datastruktur

[Text](#)

Geografisk utbredning

Geografisk plats: [Sverige](#)

Ansvarig institution/enhet

Institutionen för journalistik, medier och kommunikation (JMG)

Forskningsområde

[Språkteknologi \(språkvetenskaplig databehandling\)](#) (Standard för svensk indelning av forskningsämnen 2011)

Nyckelord

[Politik](#), [Språk](#), [Nationalitet](#), [Yrken](#), [Städer](#), [Geografi](#), [Namn](#), [Datorhantering av naturligt språk](#), [Programmeringsspråk](#)

Tillgänglighetsnivå

Åtkomst till data via extern aktör
Data är fritt tillgängliga

Upphovsrätt

Särskild upphovsrätt (t.ex. Creative Commons) för respektive material beskrivs på <https://github.com/peterdalle/svensktext>

Hemsida

[Data via GitHub](#)

CLARIN Virtual Collection Registry

[Lägg till i samling](#)

En virtuell samling är kopplad till ett specifikt forskningsändamål och innehåller länkar till dataresurser i olika digitala arkiv. Samlingen är lätt att skapa, få åtkomst till och citera.

Read more about virtual collections on the [CLARIN website](#).

Ladda ner metadata

[DataCite](#)

[DDI 2.5](#)

[DDI 3.3](#)

[DCAT-AP-SE 2.0](#)

[JSON-LD](#)

[PDF](#)

[Citation \(CSL\)](#)