

Analysera textsamlingar i Mink

Webbinarium för SND, 2025-02-26

Anne Schumacher Olsson
Arild Matsson
Martin Hammarstedt

spraakbanken.gu.se

Innehåll

1. Om Språkbanken Text
2. Korpusarbete och varför vi byggde Mink
3. Om Mink
4. Överblick – Minks användargränssnitt
5. Demo – några specifika användningsfall

Språkbanken Text – vilka är vi?

- **forskningsinfrastruktur** för språkliga data och en **språkteknologisk** forskningsenhet
- del av **Språkbanken**, en **nationell infrastruktur** till stöd för forskning baserad på språkliga data



Från en idé till verklighet – Språkbanken föds

 **1965** – Sture Allén, en riktig språkdetectiv, programmerar i maskinkod för att analysera 1600-talsbrev.

 **1970** – Han slänger ut en vision i Dagens Nyheter: "Sverige behöver en språkbank!"

 **1975** – Logoteket (a.k.a. Språkbanken) ser dagens ljus, med nationell finansiering.

Från ordbok till språkteknologisk superkraft

 **2000-talet** – Språkbanken går från att bara jobba med ordböcker till att bli en high-tech-fabrik för språkteknologi.

 **Idag** – Språkbanken består av fyra avdelningar:

 **Språkbanken Text:** textbaserad språkteknologi (Göteborg)

 **Språkbanken Tal:** talteknologi (KTH)

 **Språkbanken Sam:** kulturarv och språkpolitik (Isof, Uppsala)

 **Språkbanken CLARIN:** koordinerar de svenska CLARIN-aktiviteterna (Uppsala)

 **Juni** – firar vi Språkbanken Texts 50-årsjubileum!

Korpusarbete och våra “husdjur”

Korp 
Språkbankens ordforskningsplattform

Sparv 
Språkbankens analysplattform

Mink 
Språkbankens dataplattform

Strix 
Språkbankens textforskningsplattform



Enkel | Utökad | Avancerad | Jämförelse

Sökhistorik

stadsteater (substantiv) **Sök**

☐ i fri följd och även som ☐ förlöd ☐ mellanled ☐ efterled och ☐ skiftlägesoberoende

Relaterade ord (SWE-FN)
begravningsplats, arbetsplats, avlastningsplats
avställningsyta, betesområde, betesplats ...

KWIC: träffar per sida: 25 | sortera inom korpus på: förekomst | Statistik: sammanställ på: ord ☐ Visa statistik ☐ Visa ordbild

KWIC | Statistik | Ordbild

Antal träffar: 440

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > » Gå till sida av 18 Visa kontext

GP 2013

Att de båda nu i två trevliga uppsättningar spelas på en liten scen i	Stadsteaterns	källare är alltså mycket passande.
Malin Stenbergs undersöker könsroller och drifter i nya föreställningen Djävlar på hjul, med premiär på Göteborgs	stadsteater	i morgon.
kulturråd emeritus, tidigare chef för	Stadsteatrarna	i Borås (1990-92) och Stockholm (1992-2002) samt svenskt kulturråd i USA (2002-2006).
Var: Göteborgs	stadsteater	
Göteborgs	Stadsteater	fortsätter sin imponerande satsning på ny svensk dramatik.
	Stadsteatern	, Källaren
Göteborgs	Stadsteater	19.00
Han jobbade på	Stadsteatern	, var etablerad, känd.
Anställd: På Göteborgs	stadsteater	.
Var: Göteborgs	stadsteater	
väll har han den stora äran att få besöka Göteborg, eller så är det vi som ska tacka för att denne poetiske möbelsnickare vill komma till	Stadsteaterns	Foajébar och dela med sig av sin pop på svenska.
Nya Studion,	Stadsteatern	, 19.00
Skönt då att du nyligen fick fast tjänst på Göteborgs	stadsteater	?
Malin Stenberg sitter i ett sjukhusvitt fikarum på	Stadsteatern	och gestikulerar med ena handen.
Göteborgs	Stadsteater	19.00
	Stadsteater	i dag och

Antal träffar: 440

GP 2013

Att de båda nu i två trevliga uppsättningar spelas på en liten scen i	Stadsteaterns	källare är alltså mycket passande.
könsroller och drifter i nya föreställningen Djävlar på hjul, med premiär på Göteborgs	stadsteater	i morgon.
kulturråd emeritus, tidigare chef för	Stadsteatrarna	i Borås (1990-92) och Stockholm (1992-2002) samt s
Var: Göteborgs	stadsteater	
Göteborgs	Stadsteater	fortsätter sin imponerande satsning på ny svensk dra
	Stadsteatern	, Källaren
Göteborgs	Stadsteater	19.00
Han jobbade på	Stadsteatern	, var etablerad, känd.
Anställd: På Göteborgs	stadsteater	.
Var: Göteborgs	stadsteater	
g, eller så är det vi som ska tacka för att denne poetiske möbelsnickare vill komma till	Stadsteaterns	Foajébar och dela med sig av sin pop på svenska.
Nya Studion,	Stadsteatern	, 19.00
Skönt då att du nyligen fick fast tjänst på Göteborgs	stadsteater	?
Malin Stenberg sitter i ett sjukhusvitt fikarum på	Stadsteatern	och gestikulerar med ena handen.
Göteborgs	Stadsteater	19.00
	Stadsteater	i dag och
	Stadsteater	i dag och
n och Borås äntligen att få se Rusalka från Göteborgsoperan eller I sista minuten från	Stadsteatern	, Fragmente från Folkteatern, dans från Atalante eller c
Institutioner som man har betalat så många miljoner för: Operan 310 mkr,	Stadsteatern	81 mkr, Folkteatern 50 mkr, Atalante mer blygsamma
Var:	Stadsteaterns	källare 19.00
	STADSTEATERN	
	STADSTEATERNS	FOAJÉBAR
Foajébaren,	Stadsteatern	, 12.00
	Stadsteaterns	trånga källare kongenial spelplats för dubbla Beckett
Efter bärande roller på Göteborgs	stadsteater	och i flera långfilmer och tv-produktioner är han just

Ladda ner träffsida som... 

JSON

Korpus

GP 2013

Textattribut

datum: 2013-02-04
artikelförfattare: Tony Fischier
artikelavdelning: Kultur Nöje

Ordattribut

ordklass: substantiv
grundform: stadsteater
lemgram: stadsteater (substantiv)

msd: NN.UTR.SIN.DEF.GEN 

betydelse:
• stadsteater

förled: stadsteater (substantiv)
stad (substantiv)

efterled: teater (substantiv)
n (substantiv)

sammansatta ordformer:
• stadsteater+ns
Visa fler (3)

sammansatta lemgram:
• stadsteater (substantiv) + n (substantiv)
Visa fler (2)

dependensrelation: Huvud
uttryck: namnuttryck
namntyp: platsnamn
namnundertyp: facilitetsentitet
namn: Stadsteaterns

Visa dependensträd

institutioner som man har betalat så många miljoner för: Operan 310

0 mkr, 50 mkr, Atalante mer b

namnigenkänning

ordklass: substantiv

grundform:

- stadsteater

lemgram:

- stadsteater (substantiv)

msd: NN.UTR.SIN.DEF.GEN ⓘ

betydelse:

- stadsteater

förled:

- stadsteater (substantiv)
- stad (substantiv)

efterled:

- teater (substantiv)
- n (substantiv)

sammansatta ordformer:

- stadsteater+ns
Visa fler (3)

sammansatta lemmgram:

- stadsteater (substantiv) + n (substantiv)
Visa fler (2)

dependensrelation: Huvud

uttryck: namnuttryck

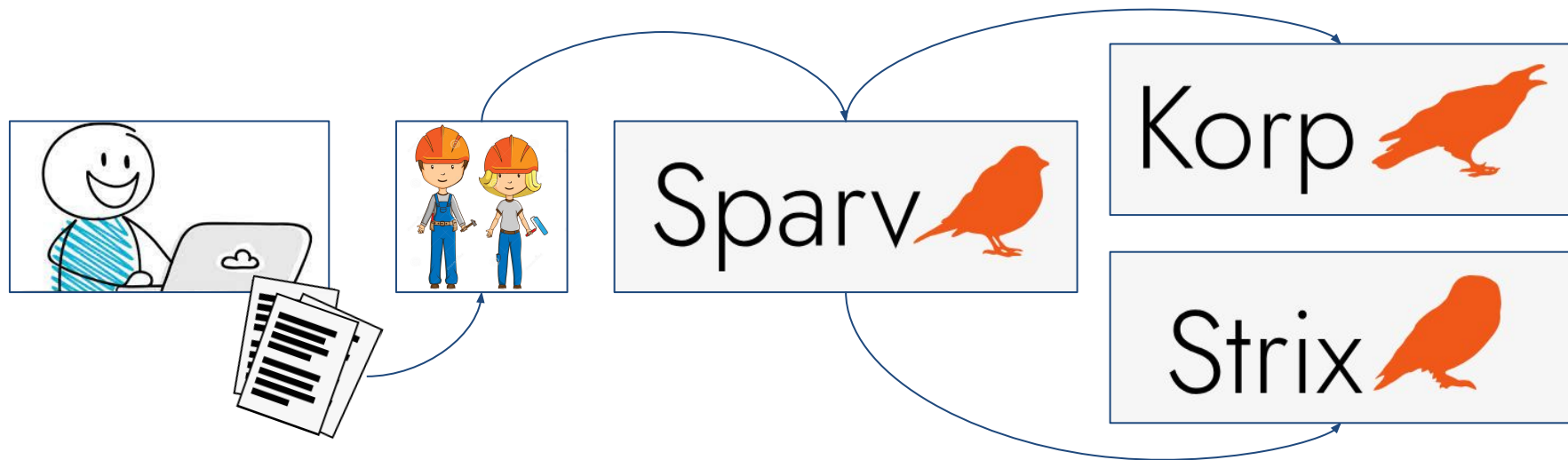
namntyp: platsnamn

namnundertyp: facilitetsentitet

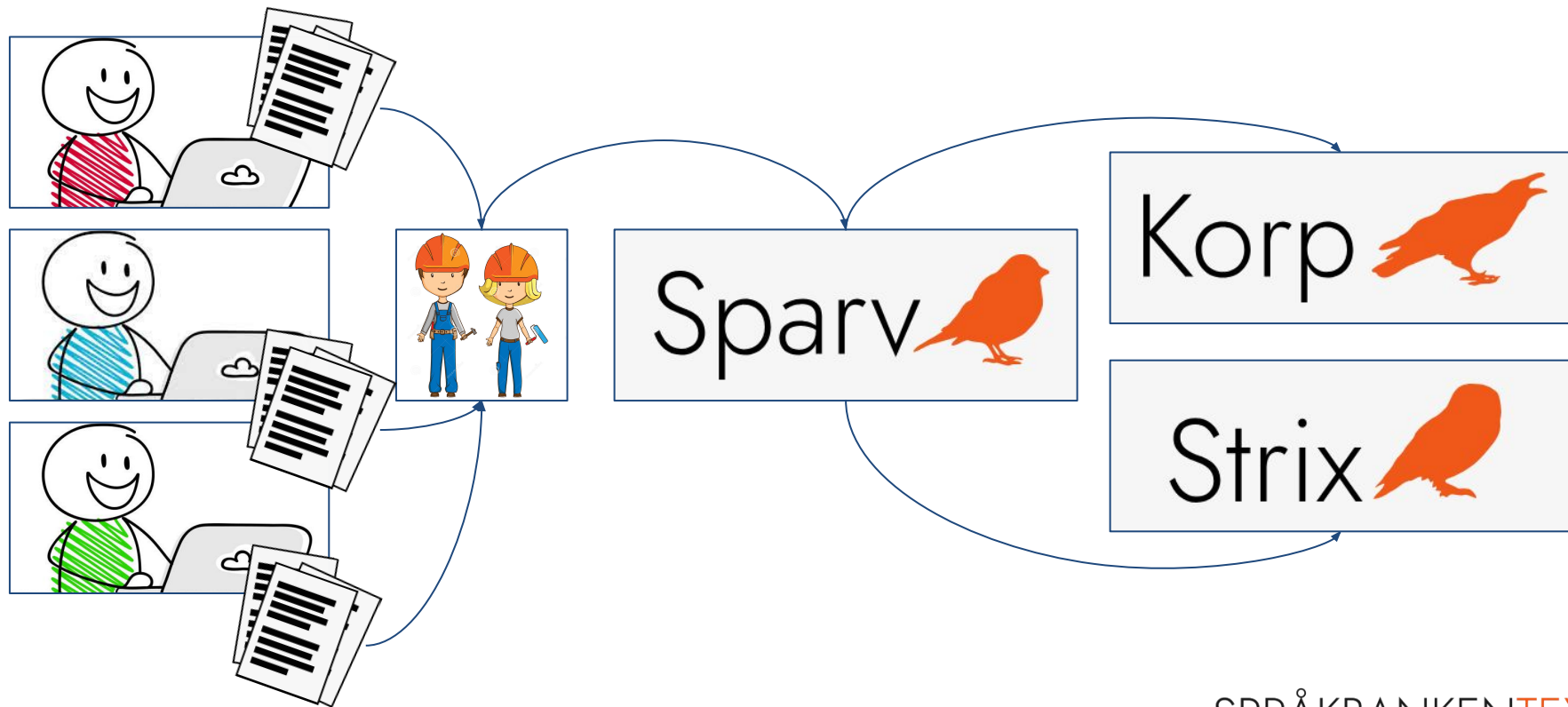
namn: Stadsteaterns

Visa dependensträd

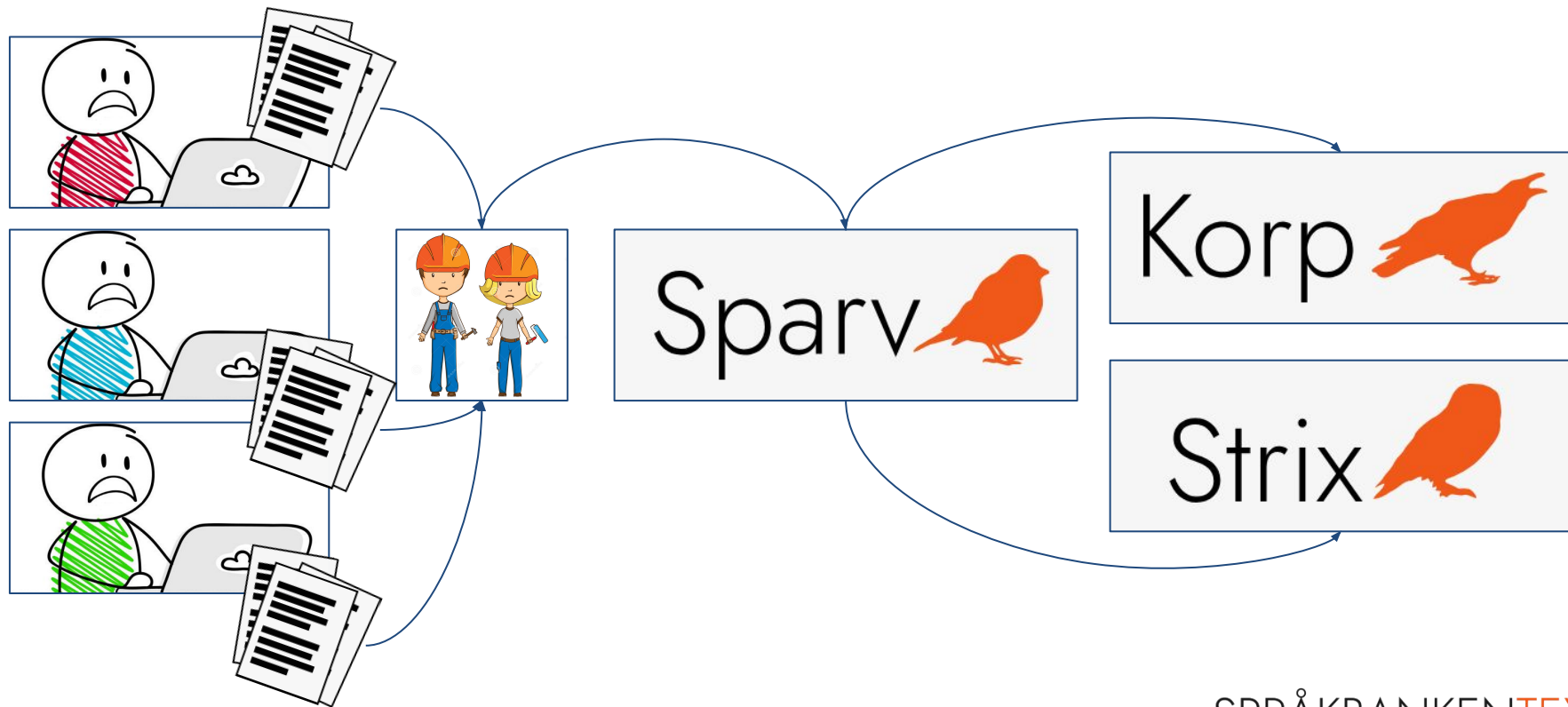
Korpusarbete på Språkbanken Text



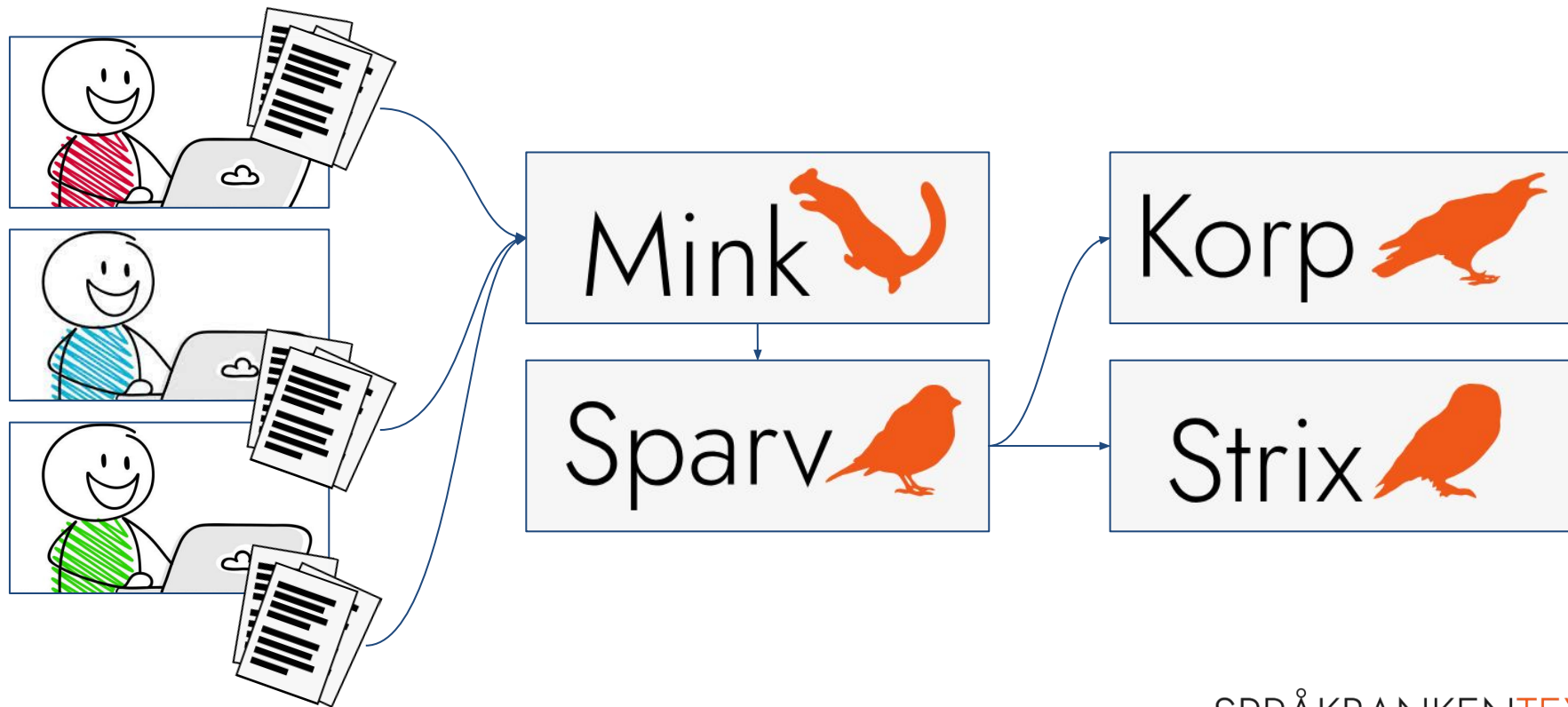
Korpusarbete på Språkbanken Text



Korpusarbete på Språkbanken Text



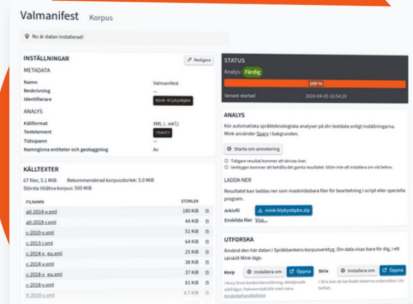
Vi behöver en Mink!



Överblick



SPRÅKBANKENTEXT



Språkbanken Text är en forskningsinfrastruktur för språkdata. Med språkteknologiska metoder i framkant skapar vi och underhåller språkresurser för forskare världen över, och vi använder fri och öppen access såvitt det går.

Mink är vår insats för att lägga infrastrukturen i forskarnas händer. Här kan du applicera våra språkteknologiska metoder på texter som du själv har samlat in. Resultaten kan du ladda ner eller föra in i våra forskningsverktyg såsom Korp och Strix bakom inloggning.

Dina källdata och resultaten är privata (se [integritets- och datapolicy](#)). I kommande versioner kommer du att kunna dela datamängder med kollegor och publicera dem för allmän åtkomst.



De flesta med akademisk anknytning kan logga in direkt.



Andra användare behöver i regel registrera ett konto.

SPRÅKBANKENTEXT

Överblick: logga in

You are connecting to...

Språkbanken Text

SPRÅKBANKENTEXT

Språkbanken Text is a research infrastructure for language data and a language technology research unit at the Department of Swedish, Multilinguality, Language Technology, University of Gothenburg.

[Privacy Statement](#)

This service requests that you **identify yourself**. Select your preferred identity service below...

universitet

University of Gothenburg - Göteborgs Universitet

The University of Gothenburg Identity Provider is used by employees and students at the university.



GÖTEBORGS UNIVERSITET

Karlstad University - Karlstads universitet

Identity Provider for Karlstad University



Linköping University - Linköpings universitet

Identity Provider for employees and students at Linköping University.



Lulea University of Technology - Luleå tekniska universitet

Identity Provider for Lulea University of Technology



Lund University - Lunds universitet

Identity provider for employees and students at Lund University)



Malmö University - Malmö universitet

Identity Services for Malmö University



Mälardalen University - Mälardalens universitet

The Mälardalen University Identity Provider is used by employees and students at the university.



SPRÅKBANKENTEXT

Överblick: ladda upp text



SPRÅKBANKENTEXT

Arild Matsson  Svenska 



Hem » Bibliotek

Bibliotek

KORPUSAR

Här listas de korpusar du har skapat. De är bara synliga för dig. Klicka på dem för att visa mer.

Valmanifest

67 filer

Installation färdig

Geo

1 fil

Annotering färdig

mink-g7yaowftnn

1 fil

Annotering färdig



Ny korpus

NY KORPUS

LÄGG TILL FILER

Dra och släpp filer hit, eller klicka här för att välja

Rekommenderad filstorlek1 MiB

Högsta tillåtna filstorlek10 MiB

Största tillåtna uppladdning100 MiB

Överblick: köra analys



SPRÅKBANKENTEXT
Arild Matsson · Svenska



Hem » Bibliotek » mink-crdncfncsy

mink-crdncfncsy Korpus

🔔 Annoteringen körs i bakgrunden. Det går bra att stänga webbsidan, ta en kaffe och kolla läget igen senare!

INSTÄLLNINGAR

Redigera

METADATA

Namn

mink-crdncfncsy

Beskrivning

—

Identifierare

mink-crdncfncsy

ANALYS

Källformat

Ren text (.txt)

Existerande

Inget

meningssegmentering

—

Tidsspann

—

Annotationer

Lexikala klasser, Morfologi, Läsbarhet, Lemmatisering, Sentiment, Dependenser, Betydelser

STATUS

Analys: **Arbetar** **Avbryt**

70 %

Senast startad 2025-02-20 14:33:19

ANALYS

Kör automatiska språkteknologiska analyser på din textdata enligt inställningarna. Mink använder Sparv i bakgrunden.

Starta annotering

LADDA NER

Resultatet kan laddas ner som maskinläsbara filer för bearbetning i script eller speciella program.

Nedladdningar finns här efter slutförd annotering.

KÄLLTEXTER

1 fil, 98 B Rekommenderad korpusstorlek: 5.0 MiB

Största tillåtna korpus: 500 MiB

SPRÅKBANKENTEXT

Överblick: använda resultat

Mink

SPRÅKBANKENTEXT

Arild Mattson Svenska

Hem » Bibliotek » Valmanifest 2014 » Resultat

Valmanifest 2014

Korpus

RESULTAT

Annoteringsprocessen skapar exportfiler för varje källfil, och de kan laddas ner enskilt eller som arkiv. Filernas struktur beror både på källfilerna och konfigurationen. De lagras på servern tills vidare, men ersätts med nya om du kör om annoteringen på nytt.
[Läs mer om exportfiler i användarhandboken](#)

Ladda ner all utdata [mink-gaudpeagi.zip](#)

FILNAMN

storlek

stats_export.frequency_list_csv/

970 KiB

c-2014-y.csv

730 KiB

csv_export/

all-2014-y.csv

3.3 MiB

f-2014-y.csv

710 KiB

fp-2014-y.csv

810 KiB

ld-2014-y.csv

1.9 MiB

m-2014-h.csv

870 KiB

mp-2014-y.csv

Mink

Logga ut Arild Mattson English Svenska Referera till Korp MENY

Korp

SPRÅKBANKENS ordforskningsplattform

Valmanifest 2014 vald – 110,71K av 1,00M token

Enkel Inläst Avancerad Jämför

Sökhistorik

ordklass

är

Lägg till token

eller

Lägg till grupp

och

Alternativ

Sök

☐ fr följd och inom mening

KWIC träffar per sida: 25

sortera inom korpus på: förekomst

Statistik

samtalstall på: ord

Visa statistik

Visa ordbild

KWIC Statistik Ordbild

KORPUS

Valmanifest 2014

TEXTATTRIBUT

text source: all-2014-y

text party: Alliansen

text källa: Hämtat från partiets officiella hemsida.

text id: all-2014-y

text type id: v

text party id: all

text type: Valmanifest - Riksdagsval

text year: 2014

Antal träffar: 8 917

Gå till sida: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ...

Visa kontext

VALMANIFEST 2014

nas förord: Alliansens löfte till väljarna Alliansen har en gemensam vision för Sverige.

skyddelst som ger stöd om vi hamnar snett, om vi blir sjuka eller om vi för stunden eller inte längre orkar av egen kraft.

igår att utveckla sin förmåga och där alla kan känna sig trygga stort.

stärkt

Moderna Mini Flar

Logga ut Svenska English Meny

Sirix

SPRÅKBANKENS ordforskningsplattform

1 av 2 korpusar valda (11 av 78 dokument, 110,71K av 918,71K token)

Titel:

Feminism för alla, Feministiskt initiativ - Valplattform riksdag 2014

Korpusnamn: ID:

VALMANIFEST 2014

Dokumentstreck:

2255 token

Ar:

2014

Vänstaste subattribut:

politik (15) rättsighet (14) kvinna (13) mänskliga (12) diskriminering (11) rätt (11) samhälle (11) arbete (10) skola (10) ansvar (8)

Stäng dokument

Dokumenttry

Statistik

Utförsk

ordattribut

Dependensrelation

CJ

Sök i dokumentet

1 Feministiskt initiativ - Valplattform riksdag 2014

2

3 F1

4

5 Feminism för alla

6

7 Det är nu det händer - var med och skriv historik! Feministiskt initiativ väljer demokrati med en genomgripande politik m den öppna handen som är redo att samla de växande rörelser som vill ha jämlikhet, jämställdhet, utjäpnighet och social politik.

8

9 Feministiskt initiativs utgångspunkt är att alla människor ska fördes vill genom livet. Frihet från våld och diskriminering, all för ett demokratiskt samhälle. Denna syn på världet ställer vi mot den orättvisa som ryftar till att de flesta människor i världen, mestadels kvinnor, lever i fattigdom, utsatthet och våld. Världens kvinnor står i centrum för utvecklingen, och vi är oss utrymme att genom våra

10

Textattribut

Strukturella attribut

Ordattribut

LENGD

initiativ (subattribut)

Demo

1. En enkel körning
2. Metadata i källfiler
3. Resultat i tabellformat
4. Jämföra korpusar

Demo: enkel körning

- Källdata: enkel textfil
- Visa inställningar
- Visa körprocessen
- Visa export-XML
- Visa installera-knapparna

Demo: metadata i källfiler

- Förberedd korpus “Valmanifest 2014” med:
 - egna attribut på textnivå: parti m m
 - analys
- Visa export-XML
 - egna attribut + analys
- Visa sökning i Korp:
 - [pos = "JJ" & _.text_party = "Feministiskt initiativ"]
- Spara och visa jämförelse med [... party = "Vänsterpartiet"]

Demo: resultat i tabellformat

- Förberedd korpus “Valmanifest 2014” med:
 - analysannotationer
- Visa CSV-fil
 - Öppna med LibreOffice
 - Kolumner separeras med tabbstopp
 - Rader som börjar med “#” innehåller strukturella attribut
 - Visa token, pos, lemma, sentiment_label (göm resten)

Demo: jämföra korpusar

- Två förberedda korpusar “Valmanifest 2014” och “Valmanifest 2018”
- Visa statistik i Korp
 - Sök på ord = vad som helst
 - Sammanställ på attityd
 - Visa trend i positiva/negativa

Tack!

spraakbanken.gu.se/mink