

Dataset med annoteringar av tekniskt språk från fyra års tillståndsovervakning av pappersmaskinsindustri i norra Sverige

SND-ID: 2023-257. **Version:** 1. **DOI:** <https://doi.org/10.5878/hafd-ms27>

Tillhörande dokumentation

dataset_description.pdf (73.1 KB)

Citering

Löwenmark, K., Sandin, F., Liwicki, M., & Schnabel, S. (2023) Dataset med annoteringar av tekniskt språk från fyra års tillståndsovervakning av pappersmaskinsindustri i norra Sverige (Version 1) [Dataset]. Luleå tekniska universitet. Tillgänglig via: <https://doi.org/10.5878/hafd-ms27>

Skapare/primärforskare

[Karl Löwenmark](#) - Luleå tekniska universitet, Institutionen för system- och rymdteknik

[Fredrik Sandin](#) - Luleå tekniska universitet, Institutionen för system- och rymdteknik

[Marcus Liwicki](#) - Luleå tekniska universitet, Institutionen för system- och rymdteknik

[Stephan Schnabel](#) - Svenska Kullagerfabriken

Forskningshuvudman

[Luleå tekniska universitet](#) - Institutionen för system- och rymdteknik

Diarienummer hos huvudman

2019-02533

Beskrivning

Detta dataset består av tekniskt-språk-annoteringar från fyra års insamling från två pappersmaskiner i norra Sverige, strukturerat som en Pandas dataframe. Samma data finns också tillgänglig som en semikolonseparerad .csv-fil. Datan består av två kolumner, där den första kolumnen motsvarar annoteringens textinnehåll, och den andra titeln. Annoteringarna är skrivna på svenska, och processade så att alla egennamn ersatts av textsträngen 'egennamn'. Varje rad motsvarar en annotering med titel.

Data behandlas i Python med:

```
import pandas as pd
annotations_df = pd.read_pickle("Technical_Language_Annotations.pkl")
annotation_contents = annotations_df['noteComment']
annotation_titles = annotations_df['title']
```

Data innefattar personuppgifter

Ja

Typ av personuppgifter

Signerade annoteringar förekommer i bevarade rådata. Datasetet innehåller därmed pseudonymiserade personuppgifter.

Språk

[Svenska](#)

Tidsperiod(er) som undersökts

2018 - 2022

Dataformat / datastruktur

[Text](#)

[Övrigt](#)

Resurstyp

Korpus

Tänkt användning

Språkteknologiskt datorprogram, Mänsklig användning

Text corpus

- Antal språk
 - Enspråkig
- Språk
 - Svenska (swe)
 - Tekniskt språk (jargong)
- Modalitet
 - Skriftspråk
- Storlek
 - Poster: 2385
 - Uttryck: 1613
- Uppmärkning
 - Entitetsomnämnananden
 - Automatisk uppmärkning
 - Uppmärkta element: Övrigt
- Källa
 - <https://doi.org/10.5878/z34p-qj52>

Geografisk utbredning

Geografisk plats: [Sverige](#)

Geografisk beskrivning: Norrland

Ansvarig institution/enhet

Institutionen för system- och rymdteknik

Medverkande

Peter Wikström - SCA Munksund

Håkan Sirkka - Smurfit Kappa

Pär-Erik Martinsson - Luleå tekniska universitet, Institutionen för system- och rymdteknik

Kjell Lundberg - Smurfit Kappa

Per-Erik Larsson - Svenska Kullagerfabriken

Smurfit Kappa

RISE Research Institutes of Sweden

SCA Munksund

Svenska Kullagerfabriken

Finansiering

- Finansiär: Verket för innovationssystem
- Diarienummer hos finansiär: 2019-02533
- Projektnamn på ansökan: Kunskapsintegrering för klassificering av maskinsador
- Information om finansiering:
<https://www.vinnova.se/p/kunskapsintegrering-for-klassificering-av-maskinsador/>

Forskningsområde

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

Nyckelord

[Pappersmaskiner](#), [Övervakning av skick](#), [Språkteknologi](#), [Signalbehandling](#), [Pappersindustri](#), [Fault detection](#), [Natural language processing](#), [Teknisk språkteknologi](#), [Språkteknologi](#), [Prognostics and health management](#), [Intelligent fault diagnosis](#), [Technical language supervision](#)

Publikationer

Löwenmark, K., Taal, C., Nivre, J., Liwicki, M., & Sandin, F. (2022). Processing of Condition Monitoring Annotations with BERT and Technical Language Substitution: A Case Study. In Proceedings of the 7th European Conference of the Prognostics and Health Management Society 2022 (pp. 306–314).

DOI: <https://doi.org/10.36001/phme.2022.v7i1.3356>

URN: <urn:nbn:se:ltu:diva-95407>

SwePub: <oai:DiVA.org:ltu-95407>

Tillgänglighetsnivå

Åtkomst till data via SND

Tillgång till data är begränsad

Användning av data

[Att tänka på vid användning av data som delas via SND](#)

Licens

[CC BY 4.0](#)

Versioner

Version 1. 2023-12-21

Hemsida

[LTU:s hemsida för projektet KnowIT FAST](#)

[Vinnovas hemsida för projektet KnowIT FAST](#)

Relaterade forskningsdata i SND:s katalog

[Dataset med tillståndsovervakningsvibrationsdata annoterat med tekniskt språk, från pappersmaskinsindustri i norra Sverige](#)

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](#)

[MetaShare](#)

[MetaShare-CMDI](#)

[DDI 2.5](#)

[DDI 3.3](#)

[DCAT-AP-SE 2.0](#)

[JSON-LD](#)

[PDF](#)

[Citation \(CSL\)](#)

Publicerad: 2023-12-21

Senast uppdaterad: 2023-12-21