

Detta dataset består av tekniskt-språk-annoteringar från fyra års insamling från två pappersmaskiner i norra Sverige, strukturerat som en Pandas dataframe. 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']
```

This dataset consists of four years of technical language annotations from two paper machines in northern Sweden, structured as a Pandas dataframe. The data consists of two columns, where the first column corresponds to annotation note contents, and the second column corresponds to annotation titles. The annotations are in Swedish, and processed so that all mentions of personal information are replaced with the string 'egennamn', meaning "personal name" in Swedish. Each row corresponds to one annotation with the corresponding title.

Data can be accessed in Python with:

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