

OCR Ground Truth Package for Swedish Fraktur

Package contains page images and ALTO XML files for each page, with the proofreading done in Swedish by the Finnish native speakers.

Description

The pages of fraktur range from the year 1771 until 1915. The package can help in creating own post correction algorithms for OCR text recognition.

Note1. the tiff files exif metadata lacks resolution information, so if the coordinates of ALTO do not match, be aware that images has been done either 200 or 300 dpi.

Note2. The ground truth package does not contain the data of the 1918 or later due to copyright reasons.

User interface

Data downloads

<http://digi.kansalliskirjasto.fi/opendata>

API

-

License

[Terms of use](#)

Content type

Pageimages, ALTO XML, metadata

Language

Swedish

Data status

Primary source

Size

xx pages

Update frequency

-

Relationships

-

External information

Article: Creating and using ground truth OCR sample data for Finnish historical newspapers and journals ([PDF file](#))

Acknowledgements

Contact information

kk-tutkijapalvelut@helsinki.fi

Star rating

-