

imgflip.com





```
BIET,

Tidskrift för Finlands landthushållning. Utg. af Johannes Jernström och G. Stolpe.

H:fors: 8.c.

1880: 1 årg. (1),388s. och 1 tab.
1881: 2 " (5),384s. och 2 "
1882: 3 " (4),384s. och 2 "
1883: 4 " (),386s.
```

```
BIET,

Tidskrift för Finlands landthushållning. Utgaf Johannes Jernström och G. Stolpe.

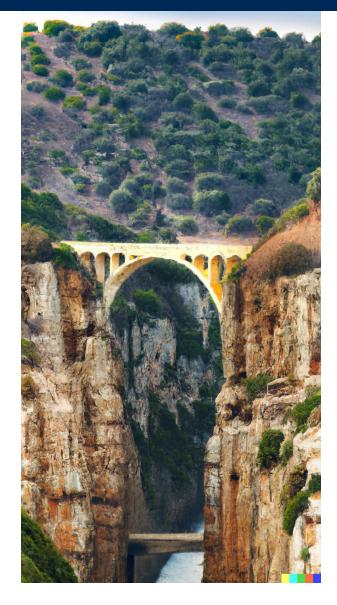
H.fors: 8.o.

1880: 1 årg. (1),388s. och 1 tab.
1881: 2 " (5),384s. och 2 "
1882: 3 " (4),384s. och 2 "
1883: 4 " (),386s.
```

the software is open sourced with a very permissive license. There's nothing stopping anyone from picking it up and doing pretty much whatever they'd like with it. **I don't have time to support** development of a GUI ...

https://github.com/mzucker/noteshrink





### Rakenteilla oleva silta

### Messydesk:

- Graafinen, selainpohjainen käyttöliittymä
- Olemassaolevien ratkaisujen paketointi
- Itsedokumentoiva rakenne
- Läpinäkyvyys (viitattavuus tärkeää)
- Tuki massa-ajoille!

työkalu tutkimukseen ja opetukseen



### Esimerkkejä

# PDF manipulointi

- kuvien ja tekstien ekstraktointi jatkokäyttöä varten
- aineistojen organisointi (esim. massa-asiasanoitus)



### Kuvat

- OCR (tekstintunnistus)
  - -- ja siihen liittyvä kuvien parantaminen
- Object detection
- kuvan löytäminen kuvasta
- tekstin löytäminen kuvasta
- auto cropping
- läpitulleen tekstin poistaminen
- aineiston luokittelu koneoppimisen avulla
- sisältö kuvailu (AI)



## **Teksti**

- kielentunnistus
- NER (luonnollisten entiteettien ekstraktointi)
- perusmuotoistus
- lauseenjäsennys
- topic modelling
- kielimallien hyödyntäminen

#### Avoimen Tiedon Keskus / JYU



