

EIN MODELL ZUR ANNOTATION HISTORISCHER WERBEANZEIGEN

UNTERSUCHUNGSgegenstand: DER STURM

zugehörige Website bei GitLab:



AUSGANGSLAGE

- Gesamte Werbeseite wird als Werbung annotiert
- Einzelne Werbungen werden nicht erkannt, sind nicht durchsuchbar
- Fehlerhaftes OCR durch Schriftmischung und graphische Elemente
- Keine konsistente Analyse der Werbungen möglich

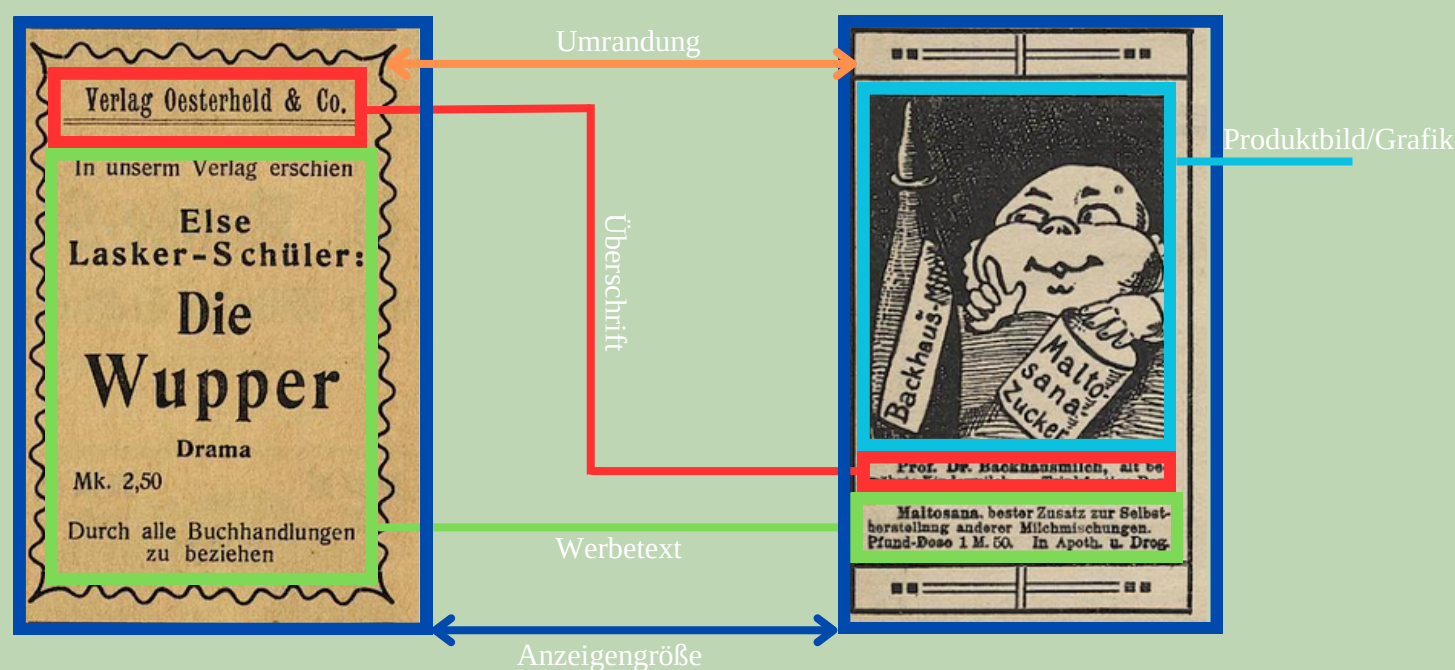
https://digi.ub.uni-heidelberg.de/diglit/sturm1910_1911/0011/image.info,textocr

VORGEHEN



METHODIK

Betrachtung grundlegender Aufbau einer Werbeanzeige



Übertragung in XML
am Beispiel der Anzeige für "Die Wupper"

```
<sturm:textcontent>
  <head>Verlag Oesterfeld & Co.</head>
  <sturm:price>2.50 Mark</sturm:price>
  <cb n="1">
    <lb>In unserm Verlag erschien</lb>
    <lb>Else</lb>
    <lb>Lasker-Schüler</lb>
    <lb>Die</lb>
    <lb>Wupper</lb>
    <lb>Drama</lb>
    <lb>Mk. 2,50</lb>
    <lb>Durch alle Buchhandlungen</lb>
    <lb>zu beziehen</lb>
  </cb>
</sturm:textcontent>
```

Manuelle Annotation von Werbeanzeigen

```
<sturm:layout ornamental="300009936" source="aat">
  <sturm:style type="bold">Else Lasker-Schüler</sturm:style>
  <sturm:font>Antiqua</sturm:font>
  <sturm:style type="bold">Die Wupper Drama</sturm:style>
  <sturm:font>Antiqua</sturm:font>
  <dimensions scope="columns">
    <dim>
      <width>1cb</width>
      <height>2cb</height>
    </dim>
  </dimensions>
</sturm:layout>
```

Die Wupper: <https://doi.org/10.11588/diglit.31770#0440>
Backhaus: <https://doi.org/10.11588/diglit.31771#0054>