



{ COD1NG }  
{ DA V1NC1 }

Der Kultur-Hackathon

# Digitalisierte historische Zeitungen der Staatsbibliothek zu Berlin

Coding da Vinci Berlin 2017

Clemens Neudecker

@cneudecker



# Wer?

- Staatsbibliothek zu Berlin – Preußischer Kulturbesitz
- Europeana Newspapers
- DAHLIE
- ZEFYS Zeitungsinformationssystem



**Staatsbibliothek  
zu Berlin**  
Preußischer Kulturbesitz

**Z**EFYS Zeitungsinformationssystem



**Europeana**  
newspapers



**EUROPÄISCHE UNION**  
Europäischer Fonds für  
regionale Entwicklung  
Investition in Ihre Zukunft



Der Regierende Bürgermeister  
von Berlin  
Senatskanzlei  
Kulturelle Angelegenheiten



# Was?

- 4 Historische Berliner Tageszeitungen



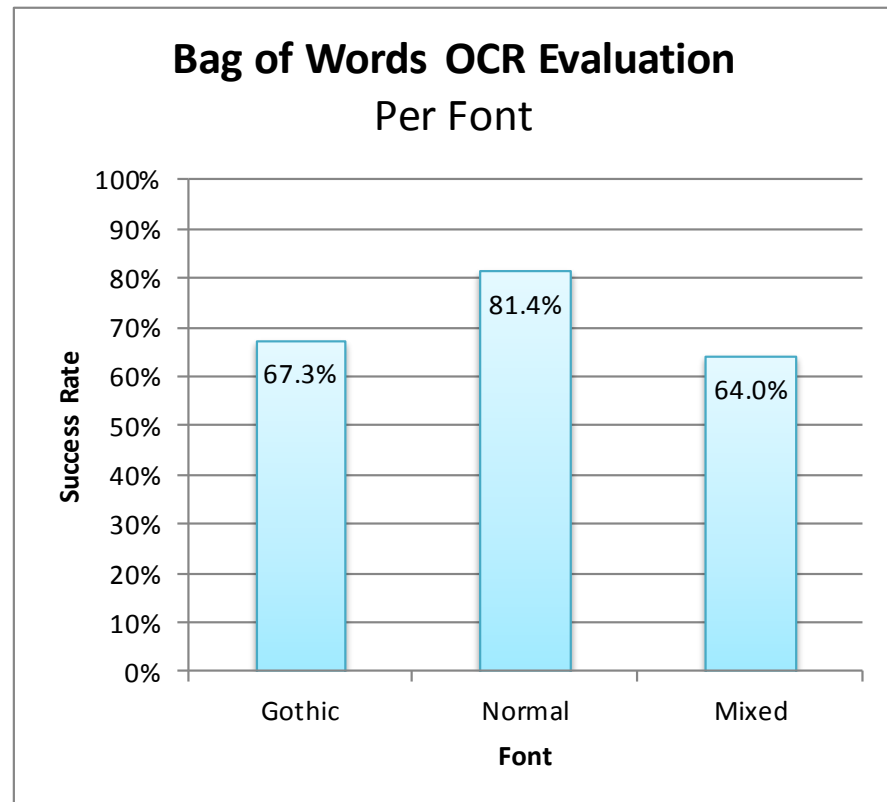
- Zeitraum 1872 – 1930
- Ca. 1,5 Mio. Seiten
- Images (Scans), Volltexte (OCR), Metadaten
- Lizenz: Public Domain, CC0

# Wie?

- OCR = Erkennung von Text in Bilddateien (Scans)



```
<TextLine ID="7268 110020" HPOS="90" VPOS="1254" WIDTH="715" HEIGHT="30">  
<String ID="7269 510010" RPOS="90" VPOS="1254" WIDTH="60" HEIGHT="27" CONTENT="A" WC="0.76" CC="007">  
<SP ID="7268 590150" HPOS="110" VPOS="1261" WIDTH="20">  
<String ID="7269 510010" RPOS="161" VPOS="1261" WIDTH="61" HEIGHT="18" CONTENT="fecundū" WC="0.70" CC="007">  
<SP ID="7268 590150" HPOS="210" VPOS="1261" WIDTH="15">  
<String ID="7269 510010" RPOS="241" VPOS="1261" WIDTH="70" HEIGHT="18" CONTENT="cubium" WC="0.63" CC="007">  
<SP ID="7268 590150" HPOS="310" VPOS="1261" WIDTH="15">  
<String ID="7269 510010" RPOS="340" VPOS="1261" WIDTH="38" HEIGHT="18" CONTENT="ē" WC="0.63" CC="007">  
<SP ID="7268 590150" HPOS="350" VPOS="1261" WIDTH="15">  
<String ID="7269 510010" RPOS="380" VPOS="1261" WIDTH="143" HEIGHT="30" CONTENT="in" WC="0.77" CC="0080007">  
<SP ID="7268 590150" HPOS="530" VPOS="1260" WIDTH="20">  
<String ID="7269 510010" RPOS="550" VPOS="1255" WIDTH="111" HEIGHT="31" CONTENT="in" WC="0.77" CC="0080007">  
<SP ID="7268 590150" HPOS="660" VPOS="1260" WIDTH="27">  
<String ID="7269 510010" RPOS="690" VPOS="1255" WIDTH="30" HEIGHT="30" CONTENT="in" WC="0.43" CC="009">  
<SP ID="7268 590150" HPOS="730" VPOS="1260" WIDTH="20">  
<String ID="7269 510010" RPOS="750" VPOS="1255" WIDTH="46" HEIGHT="17" CONTENT="in" WC="0.69" CC="170">  
</TextLine>
```



# Wo?

- <http://136.243.4.67/index.php/s/hp6TFyqvZ5ZuAlW>
- <http://www.theeuropeanlibrary.org/tel4/newspapers/search?provider-id=P01606>
- [https://www.europeana.eu/portal/de/search?q=europeana\\_collectionName%3A92\\*ewspapers\\*&f\[DATA\\_PROVIDER\]\[\]=Staatsbibliothek+zu+Berlin+-+Preu%C3%9Fischer+Kulturbesitz](https://www.europeana.eu/portal/de/search?q=europeana_collectionName%3A92*ewspapers*&f[DATA_PROVIDER][]=Staatsbibliothek+zu+Berlin+-+Preu%C3%9Fischer+Kulturbesitz)

# Wünsche

- OCR Verbesserung
- Anreicherung mit Metadaten
- Extraktion von Bildern
- Extraktion von Themen
- Extraktion von Strukturen
- Verknüpfung mit weiteren Daten
- ...und Du?

# Häh?

- Kontakt, Fragen
  - [clemens.neudecker@europeana-newspapers.eu](mailto:clemens.neudecker@europeana-newspapers.eu)
  - [@cneudecker](#)
- Dokumentation, Hintergründe
  - <http://www.europeana-newspapers.eu/public-materials/deliverables/>
  - <http://europeanane newspapers.github.io/>

# Danke! Fragen?

62. Jahrgang No. 361

# Berliner Volks-Zeitung

Morgen-Ausgabe

Sonntag, 2. August 1914

Erscheint tägl. zweimal, sonntags nur morgens. Abonnementspreis in Gross-Berlin u. vielen Orten der Provinzen Brandenburg, Pommern, Sachsen und Herzogtum Anhalt 20 Pf. wöchentlich, bzw. 85 Pf. monatlich, frei ins Haus, vierteljährl. M. 2.55. Abonnementspreis für auswärts bei Bezug durch die Post: monatlich 90 Pf. und vierteljährl. M. 2.70. — Anzeigenpreis für die Zeile 50 Pf. Stellenangebote u. Gesuche 40 Pf. Kl. Anzeigen: das Wort 8 Pf. Jedes festgedruckte Wort kostet 15 Pf. — Redaktion u. Haupt-Expd.: SW, Jerusalemstr. 46-49. Tel.: Amtszentrum, Nr. 10131-10142; für den Fernverkehr Nr. 14207-14208.

Chefredakteur: Karl Volkmann, Berlin SW.

mit täglichem Unterhaltungs-Blatt  
Illustrierter Familien-Zeitung und  
farbig illustriertem Witzblatt **ULK**

Filialen: Moritzplatz, Wiegnerstr. 1-6, Frankfurter Allee 61-62, Uferstr. 51, Prenzlauer Allee 34, Schönhauser Allee 144, Schillingstr. 4, Fehrbellin 1, Badstr. 50, Köpenickerstr. 56-57, Rosenthaler Str. 45, Hallesches Tor 3, Turnstr. 61, Potsdamer Str. 25, Leipzigerstr. 108, Zimmerstr. 60, Hohenstr. 20, Charlottenburg: Tautenburgstr. 2, Kunitzstr. 34, Schürstenerstr. 20, Nollnerstr. 20, 22, 24, 26, Friedenau: Rheinstr. 19, Hallesche-Grünwald: Heerstr. 10, Lichtenberg: Frankfurter Chaussee 130, Lichterfelde-W. Carst. 1-2, Neukölln: Berliner Str. 43, Frankow: Bornholmer Str. 1, Schöneberg: Hauptstr. 23-24, Martin-Luther-Str. 9, Stiglitz: Albrechtstr. 10, Tegel: Berliner Str. 24, Tempelhof: Berliner Str. 147, Wilmersdorf: Umlandstr. 24-26, Brandenburg: Stoliner 1, Frankfurt a. O.: Kaiserstr. 3, Potsdam: Brandenburger Str. 23, Spandau: Breite Str. 47, Stettin: Paradeplatz 5.

Druck und Verlag: Rudolf Mosse, Berlin SW.

## Deutschland macht mobil!

Der Kaiser hat gestern nachmittag um 5 Uhr 15 Minuten die Mobilmachung der gesamten deutschen Streitkräfte angeordnet. Der Chef des Großen Generalstabes begab sich ins Auswärtige Amt, wo er von diesem Beschluß Mitteilung machte. Kurz vorher war der Reichskanzler mit dem Staatssekretär v. Jagow zum Kaiser gefahren.

Als erster Mobilmachungstag gilt der heutige Sonntag.

Was seit Tagen unvermeidlich schien, das ist nunmehr geschehen: Das Deutsche Reich macht seine gesamte Kriegsmacht mobil. Nicht, weil es leichtsinnig und frivol den Krieg gewollt und gesucht hat, sondern weil diese Tat in diesem Augenblicke der natürlichste Akt der Notwehr ist. Es ist das eiserne Gebot der Selbsterhaltung, das uns zwingt, die Verteidigung unseres deutschen Vaterlandes gegen die Feinde in Ost und West mit allem Ernst und aller Energie aufzunehmen. Muß