

The German Production Pipeline: Mannheim - OCR & Knowledge Graphs



Extracting research data and creating
knowledge graphs from historical documents

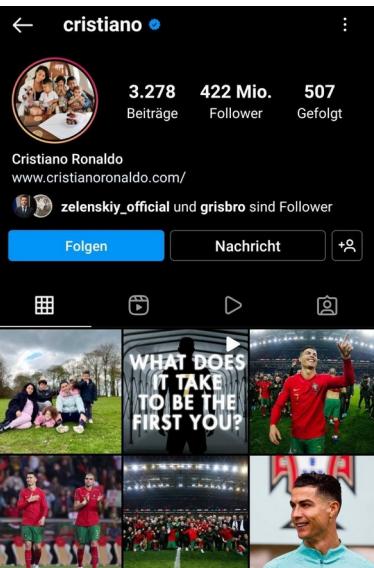
Jan Kamlah, Renat Shigapov
UB Mannheim

04.05.2023

Introduction

Goal of **BERD**@NFDI

BERD@NFDI is a consortium to build a platform for **collecting, processing, analyzing and preserving** Business, Economic and Related Data – all in one place.



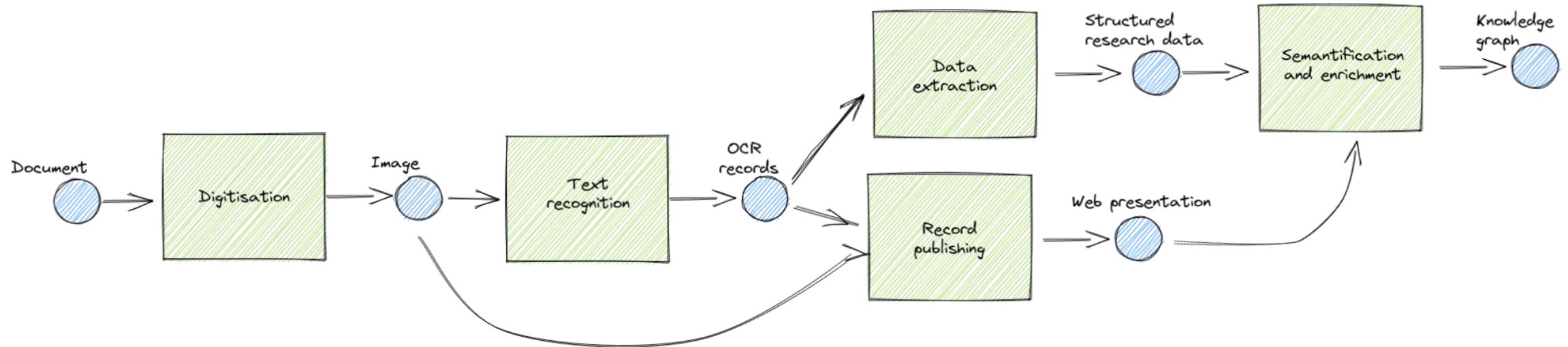
At the core of BERD@NFDI: Unstructured Data & Algorithms



BERD@NFDI

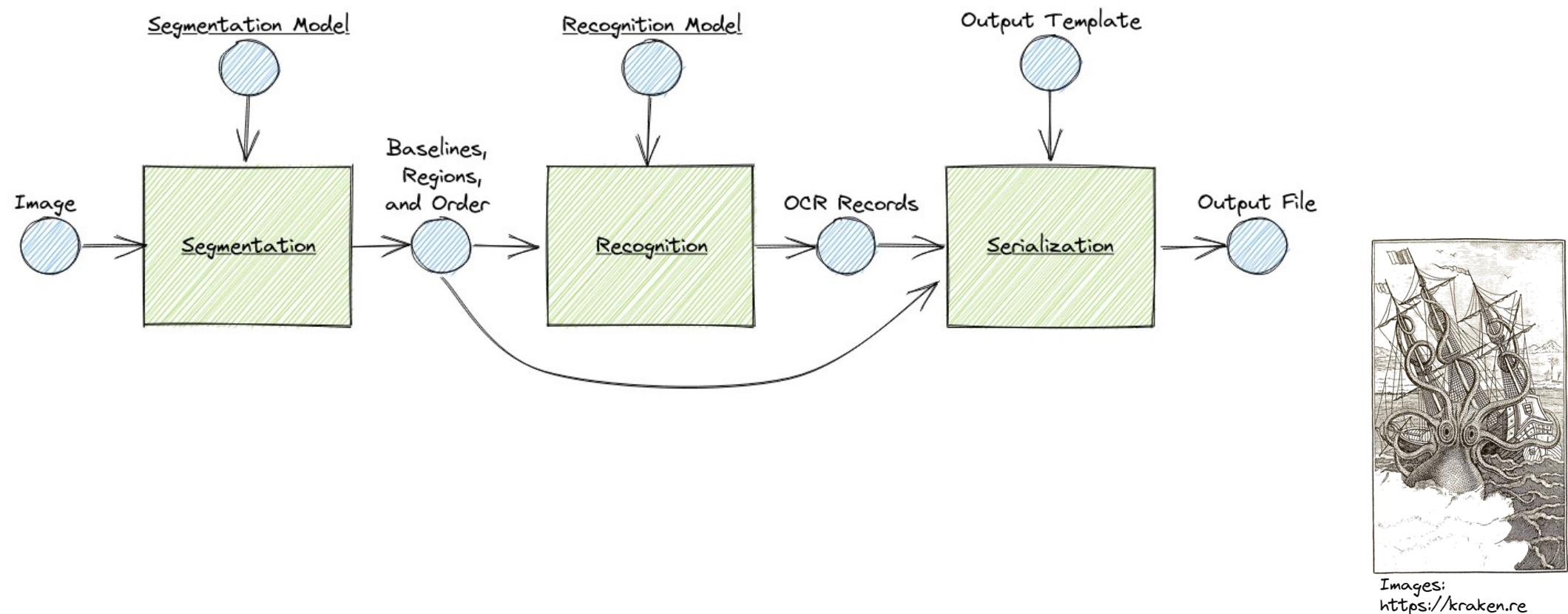


BERD-Workflow



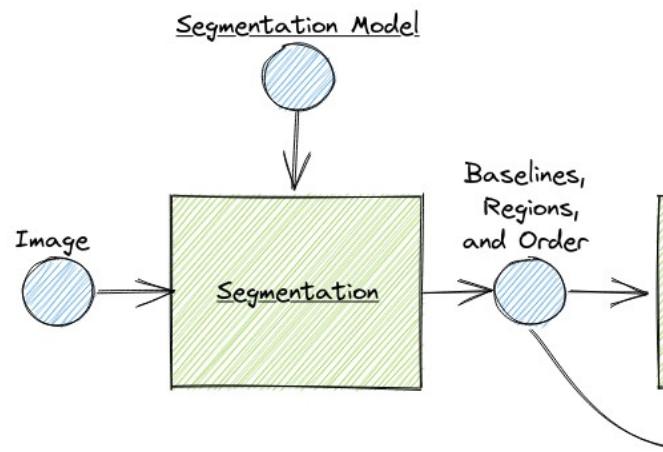
OCR/HTR

Text recognition Workflow



Segmentation Layout analysis

Layout segmentation is the process of recognising blocks of content from the background in a digitised document.

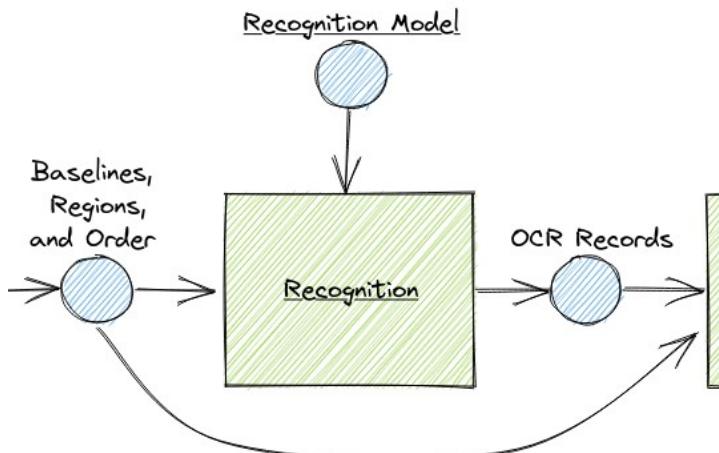


Region classification, on the other hand, is understood to be the classification of the blocks found during layout segmentation into different categories.

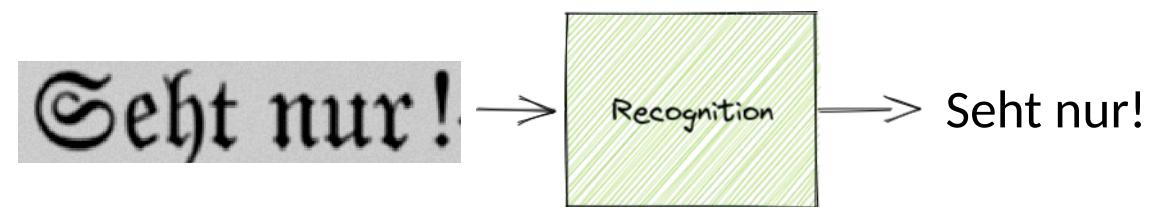
- Common approaches
- polygon detection
 - **pixel classification**
 - object detection

The ordering of the recognised elements are typically based on heuristic rules

Text Recognition



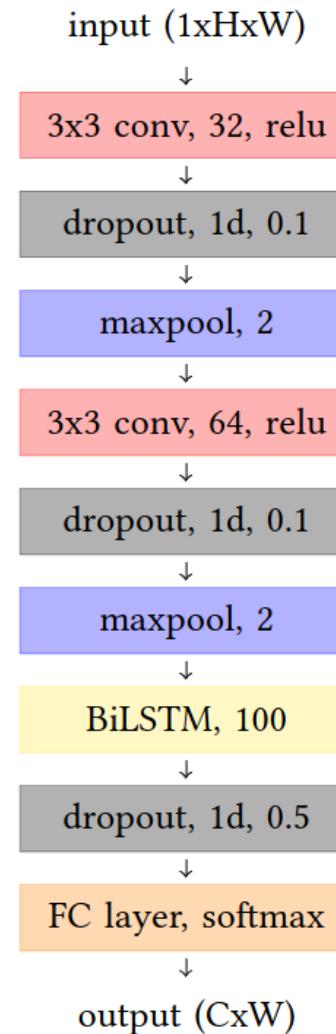
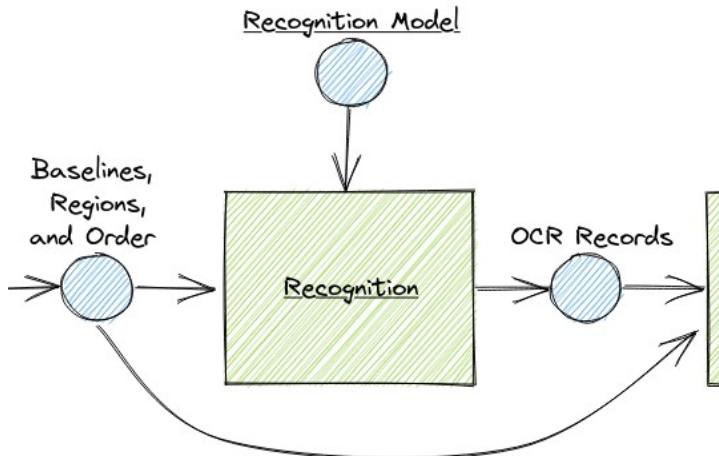
The view of modern textrecognition is to have an **input of variable size** (usually text line based) and to produce an **output of the same or different size** (group of glyphs).



Common approaches

- CTC (Connectionist Temporal Classification)
- Seq2Seq (Sequence-to-Sequence)

Text Recognition



The artificial neural network employed is a combination convolutional and recurrent neural network trained with the CTC loss function.

Network architecture
(H : sequence height,
 W : sequence length,
 C : alphabet size)

OCR/HTR Modern text recognition

Since the use of neural networks (models) the
the **possibilities** of text recognition software grow rapidly.

Neural networks must and **can be** trained.

To ensure the quality of research data
out-of-the-box methods and models are usually not sufficient.

Work-specific training on the basis of good basic models is an
efficient and effective method to achieve results with research data quality.

Work-specific Training

Optimisation of an already trained neural network (base model) according to **user-defined requirements** (task) and **the specific characteristics of a printed work** or a homogeneous series of printed works.

These optimisations are usually so profound, that the **generic nature of the base model is affected**.

The **training material** (ground truth) is usually based on **manually created** or **post-corrected transcriptions** of the work and should cover all necessary features.

Mannheim - OCR - Challenges



Tables

- small font size
 - small line spacing
 - unclean print
 - no separators
 - vertical and horizontal texts
 - texts in rows and/or across columns cells

Berliner Börse vom 4. Januar 1895.

Amtlich festgestellte Kurse.

U mrechnung s G ä p e.

1 Franc = 0,90 fl. 1 Lira = 0,90 fl. 1 Rei. = 0,90 fl.
 1 Pfennig = 0,90 fl. 1 Pfennig Gold-Grosch. = 2,68 fl. 1 Guld.
 Dansk Kr. = 1,70 fl. 1 Groschen Pfennig. = 0,05 fl. 7 Gulds.
 100 fl. = 1,125 fl. 1 Gulden. = 1,70 fl. 1 Rubel Banco
 = 1,50 fl. 1 Krone = 1,125 fl. 1 Rubel = 3,20 fl. 1 Rei. =
 4,00 fl. 1 Dollar = 4,25 fl. 1 Reale Sterling = 20,00 fl.

Wechsel.

Bank-Doll.

Amsterdam, Rott.	100 fl.	8 £.	2½	—
do.	100 fl.	2 M.	—	—
Brüssel u. Antwerp.	100 Frts.	8 £.	3	—
do.	do.	100 Frts.	2 M.	—
Skandin. Pläze.	100 Kr.	10 £.	5 *	—
Kopenhagen	100 Kr.	10 £.	3½	—
London	1 £	8 £.	—	—
do.	1 £	3 M.	2	—
Lissab. u. Óporto	1 Milreis	14 £.	—	—
do.	do.	1 Milreis	3 M.	4½
Madrid u. Batt.	100 Pct.	14 £.	5	—
do.	do.	100 Pct.	2 M.	—
New-York	100 \$	vista	5	—
Paris	100 Frts.	8 £.	2½	—
do.	100 Frts.	2 M.	—	—
Budapest	100 fl.	8 £.	4	—
do.	100 fl.	2 M.	—	—
Wien, östl. Währ.	100 fl.	8 £.	164,25 bfl. G	—
do.	100 fl.	2 M.	4	163,35 bfl. G
Schweiz. Pläze	100 Frts.	10 £.	3	—
Italien. Pläze	100 lire	10 £.	—	—
do.	do.	100 lire	2 M.	5
St. Petersburg	100 R. S. £.	8 £.	219,65 bfl.	—
do.	100 R. S. £.	3 M.	4½	217,05 bfl.
Warschau	100 R. S. £.	8 £.	4½	219,60 bfl.

Geld-Sorten, Banknoten und Knippons.

Münz-Duk. fl.—	Engl. Bank. 1 £. 20,395 fl.
Rand-Duk. fl. 9,65 bfl.	Fri. Bln. 100 fl. 81,20 bfl.
Söderg. v. St. 20,35 bfl.	Döllnitz. Noten. 169,10 bfl.
20 Krts. - St. 16,22 bfl.	Italiens. Noten. 76,20 bfl.
8 Guld. —	Nordische Noten. 112,40 bfl.
Dollars v. St. —	Dest. Bln. v. 100 fl. 164,40 bfl.
Imper. v. St. —	do. 1000 fl. 164,40 bfl.
do. v. 500 fl. —	Russ. Bln. v. 100 R. 220,00 bfl.
do. neue	ult. Jan. —
do. 500 g fl. —	—
do. do. 500 g —	ult. Febr. —
Amerit. Noten	—
1000 u. 500 g —	—
do. kleine 4,155 G	Schweiz. Not. —
do. Gp. v. N. Y. 4,16 bfl. G	Russ. Zollkuponnot. 324,10 bfl. G
Belg. Noten 81,15 bfl.	do. kleine 324,00 bfl.
Zinsfuß der Reichsbank: Wechsel 3%, Lomb. 3½ u. 4%.	—

Fonds und Staats-Papiere.

Dtche. Näch. Anl. 4	St. 3.-Tm. Stilte zu 4%
do. do.	1.4.10 5000—200 106,00 G
do. do.	verfd. 5000—200 104,75 bfl.
do. do.	verfd. 5000—200 96,25 bfl.
do. do. ult. Jan.	96,30 a. 25 bfl.
Preuß. Rent. Anl. 4	verfd. 5000—150 105,80 G
do. do.	1.4.10 5000—150 104,90 bfl.
do. do.	1.4.10 5000—150 96,50 B
do. do. ult. Jan.	96,50 bfl.
do. St. Schuldt. 3½	1.1.7 3000—75 100,90 G
Kurmärk. Schuldt. 3½	1.5.11 3000—150 100,70 bfl.
Über-Deichs. Dbl. 3½	1.7 3000—300 —
Aachener St. II. 93	1.4.10 5000—500 101,20 B
Alten-St. 87,89	verfd. 5000—500 100,60 bfl. G
do. do. 1894	1.4.10 5000—500 100,60 bfl. G
Augsb. d. v. 1889	1.4.10 2000—1* 100,80 G
Varmer St.-Ant. 3½	verfd. 5000—500 100,70 G
do. do.	1.4.10 5000—500 104,00 B
Berl. Stadt-Dbl. 3½	verfd. 5000—100 101,80 bfl. G
do. do. 1892	1.1.7 5000—100 103,25 bfl. G
Breslau St.-Anl. 4	1.4.10 5000—200 104,90 G
Cassel Stadts-Anl. 3½	1.1.7 5000—200 101,50 G
Charlottb. do. 4	1.1.7 2000—100 101,00 bfl. G
do. do. 1889	1.4.10 5000—100 —
do. do.	1.4.10 2000—100 100,00 bfl. G
Krefelder do. 3½	verfd. 5000—500 —
Danziger do. 4	1.4.10 2000—200 —
Defaufer do. 91,4	1.1.7 5000—200 —
Düsseldorf 1876	1.5.11 10000—200 101,00 bfl. G
do. do. 1890	1.1.7 2000—500 101,00 bfl. G
do. do. 1894	1.3.9 2000—500 101,00 bfl. G
Ebersf. St.-Dbl. 3½	1.1.7 5000—500 —
Grüntz. do. 3½	1.4.10 1000—200 101,00 G
Essen. do. IV. V. 3½	1.1.7 3000—200 —
Halle (P.) do. 1886	1.4.10 1000—200 —
do. do. 1892	1.1.7 2000—200 —
Karlsr. St.-Anl. 863	1.5.11 2000—200 94,30 G
do. do. 893	1.5.11 2000—200 94,30 G
Kieler do. 89,3	1.1.7 2000—500 101,00 bfl. G
Kölner do. 94,3	1.1.7 10000 500 101,90 G
Königslz. 91 I. II. 1	1.4.10 2000—500 —
do. 91 III 3½	1.4.10 2000—500 100,60 G
do. 1893	1.4.10 2000—500 100,60 G
Regnitz. do. 1892	1.1.7 2000—200 —
do. do. 1892	1.1.7 2000—200 —
Magdeb. do. 91, IV 3½	verfd. 5000—100 —
do. do.	1.1.7 5000—100 104,30 B ll. f.
Mainzer St.-Anl. 4	1.6.12 2000—200 —
do. do. 88 3½	1.3.9 2000—200 —
Mannheim do. 88 3½	1.1.7 1000—200 —
do. do. 90 4	1.2.8 1000—200 —
Ostpreuß. Prin.-D. 3½	1.1.7 3000—100 101,25 bfl. G
Polen. Prod. Anl. 3½	1.1.7 5000—100 101,10 bfl. G
co. Statt.-Anl. 3½	1.1.7 5000—200 101,10 G
Potsdam St.-Anl. 92 4	1.4.10 2000—200 —

Es betrug der in Preußen dominizierten Aktiengesellschaften und Kommanditgesellschaften auf Aktien													
Bestand am 31. 12. 1908 Zugang im Jahre 1909 Abgang im Jahre 1909 Bestand am 31. 12. 1909													
Generbegruppe:	in der		davon		durch		Abhang		Mehr-		davon		
	über- haupt	son- der- heit	mithin	mit tätige	Reu- neu- grün- ge- gan- gen- dung	im Gesell- schafts- stand	der Firma	zg	zu- wachs (+)	ver- brau- ch	über- haupt	son- der- heit	
	haupt	oder	oder	oder	oder	oder	oder	oder	oder	oder	haupt	oder	
3. = Zahl R. = Grundkapital in Millionen Mark	befind- lich												
1) Land- und Forst- wirtschaft . . . R. 3 1 2 — — — 1 — 1 2 — 2 — 2	3	1	2	—	—	—	1	—	—	—	1	2	— 2
2) Tierzucht und Fischerei . . . R. 11 — 11 1 — 1 1 — 1 1 — 12 — 12	11	—	11	1	—	1	—	—	—	—	1	12	— 12
3) Bergbau, Hütten- und Salinen (je für sich getrennt) R. 968,21 1) 14,06 954,15 36,38 36,41 89,70 8,10 2,00 1,78 12,10 + 77,61 1 045,81 25,27 1020,54	968,21 1)	14,06	954,15	36,38	36,41	89,70	8,10	2,00	1,78	12,10 +	77,61	1 045,81	25,27 1020,54
3A. Bergbau und Hüttenbetrieb, Me- tall- oder Ma- schinen- mit einander verbun- den	R. 3 34 1) — 34 — (7) — — — (1) — — 34 — 34	34	—	(7)	—	—	—	—	—	—	—	34	— 34
4) Industrie d. Steine R. 281,33 15,08 266,24 5,68 8,31 14,06 1,92 — 2,53 4,78 + 9,28 290,61 17,31 273,30	281,33	15,08	266,24	5,68	8,31	14,06	1,92	—	2,53	4,78 +	9,28	290,61	17,31 273,30
5) Metallverarbeitung R. 99 4 95 5 (7) — 1 — (4) 1 — 6 105 4 101	99	4	95	5	(7)	—	1	—	(4)	1 —	6	105	4 101
6) Industrie der Ma- schinen u. v. d. ähnlich . . . R. 171,16 1,82 169,34 16,30 6,70 24,21 1,00 — 0,85 1,85 + 22,36 193,52 1,30 192,22	171,16	1,82	169,34	16,30	6,70	24,21	1,00	—	0,85	1,85 +	22,36	193,52	1,30 192,22
7) Chemische Industrie R. 323 25 298 12 (34) 12 10 2 (12) 16 4 319 21 298	323	25	298	12	(34)	12	10	2	(12)	16	4	319	21 298
8) Industrie d. Leucht- stoffe u. w. . . R. 285,04 5,23 279,81 5,75 5,05 10,80 3,00 — (2) 4,84 + 5,96 290,99 3,12 287,87	285,04	5,23	279,81	5,75	5,05	10,80	3,00	—	(2)	4,84 +	5,96	290,99	3,12 287,87
9) Textilindustrie R. 119 7 112 7 (8) 1 — 3 (1) 3 — 3 53 3 50	119	7	112	7	(8)	1	—	3	(1)	3 —	3	53	3 50
10) Papierindustrie R. 72,51 2,10 70,41 1,03 7,65 8,68 — — — — — 61 71,95 2,82 69,13	72,51	2,10	70,41	1,03	7,65	8,68	—	—	—	—	61	71,95	2,82 69,13
11) Lederindustrie R. 29 1 28 2 — 2 — — — — — 2 45 3 42	29	1	28	2	—	2	—	—	—	—	2	45	3 42
12) Industrie der Holz- und Schnüffstoffe R. 36 6 30 2 (5) 2 1 — (3) 1 — 1 8,68 81,19 2,47 78,73	36	6	30	2	(5)	2	1	—	(3)	1 —	1	8,68	81,19 2,47 78,73
13) Industrie der Nah- rungs- und Genuß- mittel R. 54,56 0,40 54,17 1,15 1,15 — — — (1) 0,07 0,07 + 1,08 55,65 3,80 51,85	54,56	0,40	54,17	1,15	1,15	—	—	—	(1)	0,07 +	1,08	55,65	3,80 51,85
14) Befleidungs- gewerbe R. 617,38 11,70 605,68 7,20 24,33 31,54 2,48 — 5,70 9,93 + 21,61 638,99 15,44 623,55	542	29	513	11	(26)	11	8	—	(10)	9 +	2	544	34 510
15) Baugewerbe R. 17,00 1,00 16,00 1,00 1,00 2,00 — — — — — 1 12 1 11	11	1	10	1	(1)	1	—	—	—	—	1	12	1 11
16) Polysgraphische Ge- werke R. 69,27 1,85 67,42 — 3,55 3,55 — 10,00 1,35 11,35 — 2,00 19,00 1,00 18,00	35	3	32	—	(7)	—	—	—	(1)	(1)	1	34	3 31
17) Handelsgewerbe R. 51,77 2,03 49,74 1,40 0,32 1,72 0,10 — 1,18 1,28 + 0,45 52,29 2,07 50,15	51,77	2,03	49,74	1,40	0,32	1,72	0,10	—	1,18	1,28 +	0,45	52,29	2,07 50,15
18) Verfertigungs- gewerbe R. 3224,15 96,34 3127,81 49,39 87,44 145,95 19,38 13,00 1,75 34,15 + 111,80 3335,95 127,60 3208,35	507	53	454	30	(36)	33	7	3	(2)	11	22	529	59 470
19) Verkehrsge- werbe R. 400,64 1,50 399,14 3,60 1,00 4,60 — — — — — 2 85 1 84	331	13	318	6	(11)	6	5	—	(7)	5 +	4,60	405,24	1,50 403,74
20) Verkehrsge- werbe R. 846,01 11,80 834,21 6,09 6,60 12,69 2,35 — 1,78 4,12 + 8,56 894,57 13,28 841,29	32	3	29	—	(2)	—	1	—	(2)	1 +	2	72	6 66
21) Kunst-, Theater- u. Schauspielungs- gewerbe R. 47,41 1,60 46,81 — 0,03 0,03 0,15 — 0,35 0,50 + 0,48 46,94 0,45 46,49	40	3	37	3	(1)	3	3	—	(1)	3	—	31	2 29
22) Sonstige Ge- sellschaften R. 18,54 1,09 17,45 7,00 0,50 7,50 1,09 — 0,33 1,42 + 6,08 24,62 0,60 24,62	206	19	187	4	(4)	5	3	—	(1)	3 +	2	208	22 186
23) Zusammen R. 296,94 16,06 280,88 5,50 3,83 10,65 2,04 — 1,20 3,24 + 7,40 304,35 14,48 289,87	3 092	212	2,880	108 (18)	119	49	10	(60)	69 + 50	3 142	922	9,915	

Isolated glyphs or glyph groups

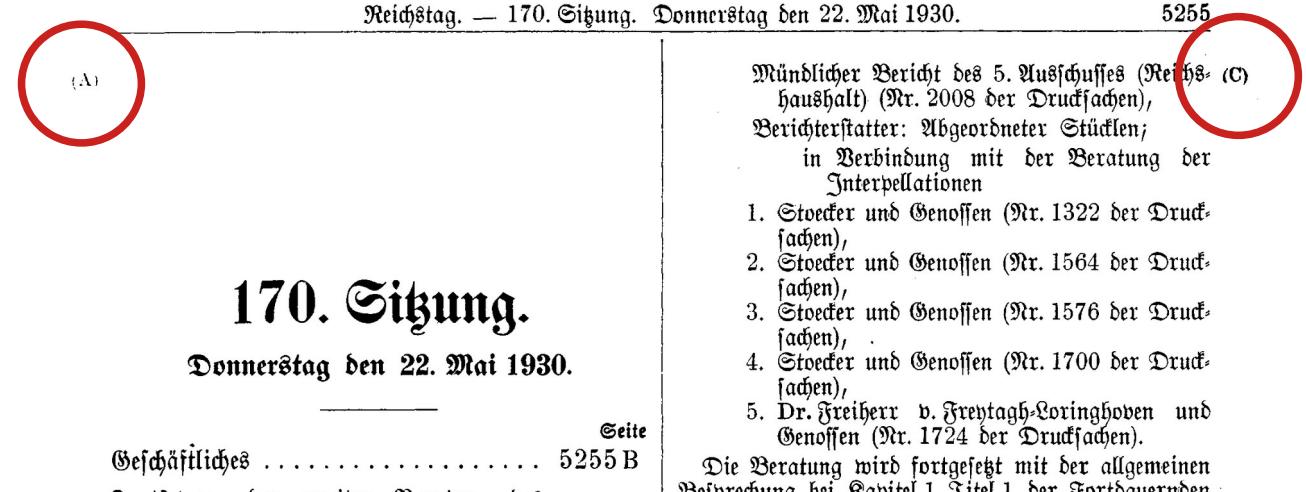
One problem that arises in layout recognition is the

- no classification
- false classification

The character accuracy of text recognition suffers because neuronal networks

- work context-dependently
- are trained with longer lines of text

Ort	Gebäude	Nr. im Meldebogen	Unterer ♂ in cm	Gewicht in kg	Gußjahr
Landkreis Feldbach: Hainfeld	Schloßkapelle	1	59	125	1700
Landkreis Graz: Reiteregg	Schloßkapelle	1	49	70	1670
Landkreis Judenburg: Rothenthurm Wasserberg	Kapelle Schloßturm	1 1	27 65	30 165	1676 —
Landkreis Murau: Murau Oberdorf	Schloßkirche Mariahof	2 1	82 25	335 25	1463 um 1300



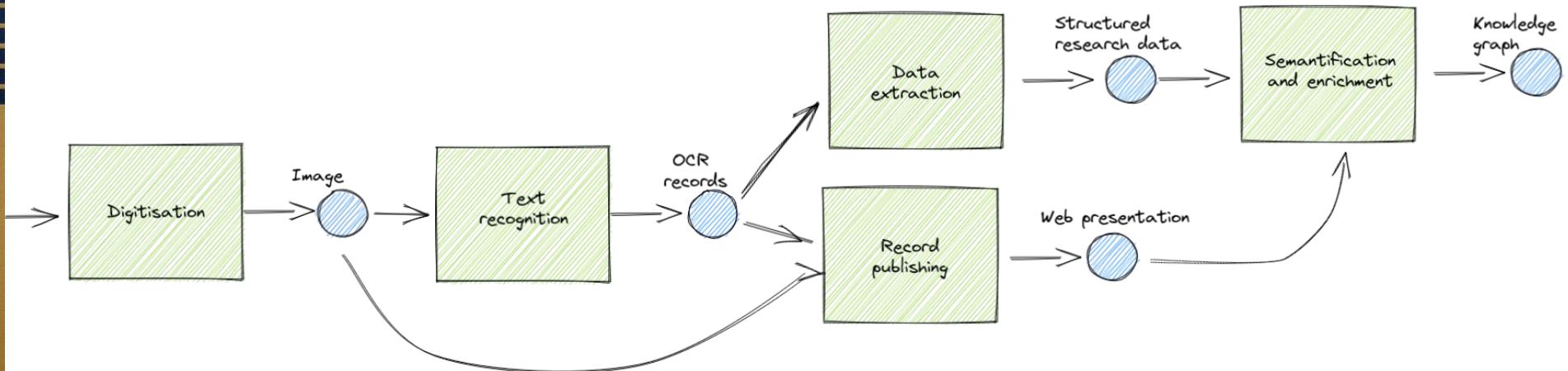
Services

- Open transcription instances
 - **eScriptorium**
<https://ocr-bw.bib.uni-mannheim.de/escriptorium>
 - **PERO-OCR**
<https://pero-ocr.bib.uni-mannheim.de/>
- **Open consultation hour** (every 2nd Thursday at 3.00–4.00 pm)

Providing models and information and training material

- **Kompetenzzentrum OCR-BW** (information and guides)
<https://ocr-bw.bib.uni-mannheim.de/anwendung/druckwerke/>
- **Generic text recognition models for historical fonts**
Models for Fraktur, Kurrent and Typerwriter: <https://ub-backup.bib.uni-mannheim.de/~stweil/tesstrain/kraken/>
- **Economic-historical layout models**
Model for historical reports (with two columns): https://digi.bib.uni-mannheim.de/~jkamlah/training/kraken/historical_reports/
- **Participate on the OCR-D GT Guidline development** (<https://ocr-d.de/en/gt-guidelines/transl/>)
- **GT-dataset for historical newspapers**
Reichsanzeiger (german newspaper): <https://github.com/UB-Mannheim/reichsanzeiger-gt>
ÖNB (austrian newspaper): <https://github.com/UB-Mannheim/AustrianNewspapers>

Workflow for Maschinen-Industrie 1937



Text recognition

eScriptorium

Transcriptionssoftware



Open source platform for **manual or automatic segmentation** and **text recognition** of historical manuscripts and prints.

Developer PSL (Paris)

Year of release 2018

Current version 0.13.5 (2023)

Operating system platform-independent

Programming language Python, JavaScript, Hypertext Markup Language

OCR engine kraken

Licence MIT licence

Models <https://ub-backup.bib.uni-mannheim.de/~stweil/eScriptorium/> (Fraktur)

Additional information <https://ocr-bw.bib.uni-mannheim.de/tag/escriptorium/>

kraken

OCR engine for layout segmentation and text recognition

kraken is an turn-key OpenSource OCR software solution optimised for historical documents and non-Latin writing systems, **with fully trainable layout analysis and text recognition.**

Developer Benjamin Kiessling

Year of publication 2015

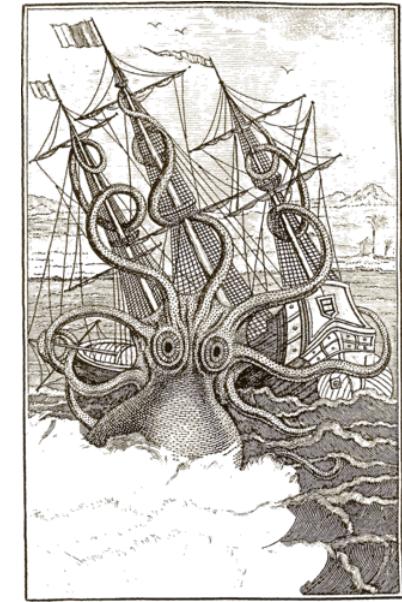
OCR technology BoundingBox and baseline-based (eScriptorium)

Output formats ALTO, PageXML, abbyyXML and hOCR

Programming language Python

Models <https://ub-backup.bib.uni-mannheim.de/~stweil/eScriptorium/> (Fraktur)

Licence MIT-Licence



Requirements for the text recognition

- High quality requirements for research data (>99.5 % character accuracy)
- Effort of factory-specific training difficult to estimate
- Profile separation
- Two columns

<p>Phönix — Pincuss</p> <p>Phönix-Werk G. m. b. H., Spezialfabrik moderner Trocken- Apparate, Meerane (Sa.).</p> <p>Fernruf: 2491. Drahtanschrift: phönixwerk Gründung: 1907. Fabrikationsprogramm: Trockenapparate; Holzbearbeitungs-Maschinen. Kapital: RM 17.500. Anteilseigner: Nestmann, Meerane (100%). Geschäftsführer: Obering, A. Wackermann. Prokurist: J. Frenzel. Bankverbindungen: Meeraner Bank A.-G., Reichsbank, Meissner Bank. Postcheck-Konto: 115 485 Leipzig. Geschäftsjahr: 1/1.—31/12. Grundbesitz: 3700 qm, davon 1402 qm bebaut. Anlagen: Fabrikationshallen mit Montagehalle, elektr. Schleppbahnen, 100 m hoch, 100 m lang. Eigene Vertretungen in Berlin, Willy-Böckel, W 30, Martin-Luther-Str. 12. Besondere Angaben: Gegründet als Spezialfirma moderner Trocknungsanlagen. Hergestellt wurden Schlangenbündel-Trockenapparate System Otto und Getreidetrockner nach dem Zellensystem neben anderen allgemeinen Trocknungsanlagen. 1921 erfolgte neben dem Trocknerbau auch die Produktion von Papier- und Textilmaschinen. 1932 übernahm das Werk zusätzlich die Herstellung von Bewölk-Apparaturen für die Papierindustrie nach den Patenten Dr. Bruno Wieger, Berlin.</p> <hr/> <p>Piccolo-Automaten G. m. b. H., Berlin W 35, Kurfürstenstraße 16.</p> <p>Fernruf: 21 39 05. Drahtanschrift: piccoloautomat Gründung: 1922. Fabrikationsprogramm: Schokoladen-Verkaufsapparate (Tischautomaten, Kugelstechapparate). Kapital: RM 1000. Anteilseigner: Dr. Richard Schönthal, Wien. Geschäftsführer: Erwin Hantsch, Bla-Steglitz. Bankverbindung: Deutsche Bank u. Disconto - Ges., Berlin. Postcheck-Konto: 8012 Berlin. Geschäftsjahr: 1/1.—31/12. Anlagen: Mechanische Werkstätten. Gefolgschaft: 13 Arbeiter u. 6 Angestellte.</p> <hr/> <p>F. Piechatzek, Kran- u. Aufzug-Werke, Berlin N 65, Seestraße 51-56.</p> <p>Fernruf: 46 43 11. Drahtanschrift: läderszug Gründung: 1885. Fabrikationsprogramm: Krane u. Aufzüge, Hebezeuge u. Hubseileinheiten (Flaschenzüge, Laufkatzen, Winden, Elektro-Flaschenzüge). Geschäftsführer: Richard, Martin u. Paul Piechatzek, Prokurist: Paul Grünring, Alfred Knop, Otto Kunwald. Bankverbindung: Reichskredit-Gesellschaft A.-G., Berlin. Postcheck-Konto: 4847 Berlin. Geschäftsjahr: 1/1.—31/12. Grundbesitz: 3600 qm, davon 2000 qm bebaut; gepachtet sind 1250 qm mit 700 qm bebauter Fläche. Anlagen: Verarbeitungs- u. Fabrikationsräume in Heidenheim, München. Besondere Angaben: In München befindet sich ein Zweigwerk der Gesellschaft. Gefolgschaft: 350 Mitglieder.</p> <hr/> <p>Eduard Pincuss Armaturenfabrik, Sanitäre Einrichtungen, Berlin O 17, Gr. Frankfurter Str. 13.</p> <p>Fernruf: 50 13 18. Drahtanschrift: epal Gründung: 1859. Fabrikationsprogramm: Wasserleitungs-Armaturen. Inhaber: Arthur Landsberger, Bla-Charlottenburg; Ernst Reichenbach, Bla-Grunewald. Prokurist: Paul Diersch, Frieda Thierschmann.</p>	<p>Otto Pieron siehe Maschinenfabrik</p> <hr/> <p>Paul Pietzschmann Wasserwerkshau, Berlin-Spandau, Schönwalder Str. 34.</p> <p>Fernruf: 37 68 71. Gründung: 1902. Fabrikationsprogramm: Entwurf und Bau von Trink- und Brauchwasserwerken jeder Größe, Entsorgungsanlagen einschl. der erforderlichen Antriebsanlagen (Dampf — Diesel — Wasserkraft, Hoch- und Niederspannungsnetze). Inhaber: Ing. Paul Pietzschmann. Bankverbindungen: Dresdner Bank, Spandauer Bank, Spandau.</p> <p>Besondere Angaben: Der Firmeninhaber beschäftigt sich hauptsächlich mit der Errichtung halb- u. vollautomatisch arbeitender Wasserwerke u. besitzt hierüber große u. langjährige Erfahrungen.</p> <hr/> <p>Anton Piller Maschinenfabrik, Osterode (Harz), Abgust 24.</p> <p>Fernruf: 211. Drahtanschrift: apo Gründung: 1909. Fabrikationsprogramm: Ventilatoren für Heizungs-, Lüftungs-, Absauge u. sonstige Zwecke. Bankverbindung: Reichsbank, Städte Sparkasse, Osterode a. H. Postcheck-Konto: 40 278 Hannover.</p> <hr/> <p>Pilot. G. m. b. H. siehe Maschinenfabrik</p> <hr/> <p>Friedrich Piltz & Sohn K.-G., Heidenheim a. d. Brenz, Friedrichstr. 9.</p> <p>Fernruf: 697. Drahtanschrift: piltz Gründung: 1863. Fabrikationsprogramm: Genauigkeitswerkzeuge für Herstellung u. Kontrolle von Gewinde, Gewindeschleif- und drehenden Draht- u. Schraubenfertigkeiten. Geschäftsführer: Otto u. Walther Piltz. Gesellschafter: Die Gesellschafter. Bankverbindungen: Reichsbank, Deutsche Bank u. Disconto-Haus, Stuttgart. Postcheck-Konto: 2178 Stuttgart. Geschäftsjahr: Kalenderjahr.</p>
--	---

Image

**Piccolo-Automaten G. m. b. H.,
Berlin W 35, Kurfürstenstraße 146.**
Fernruf: 21 20 95. Drahtanschrift: piccoloautomat
Gründung: 1932.
Fabrikationsprogramm: Schokoladen-Verkaufsapparate
(Tischautomaten, Kugelstechapparate).
Kapital: RM 20 000.—.
Anteileigner: Dr. Richard Schönthal, Wien.
Geschäftsführer: Erwin Hantsch, Bln.-Steglitz.
Bankverbindung: Deutsche Bank u. Disconto - Ges.,
Berlin.
Postcheck-Konto: 8012 Berlin.
Geschäftsjahr: 1./1.—31./12.
Anlagen: Mechanische Werkstätten.
Gefolgschaft: 13 Arbeiter u. 6 Angestellte.



Structured research data

Firmenname	Piccolo-Automaten
Rechtsform	G.m.b.H.
Sitz	Berlin W 35 Kurfürstenstraße 146
Fernruf	212095
Drahtanschrift	piccoloautomat
Gründung	1932
Produkt	Schokoladen- Verkaufsapparate (Tischautomaten, Kugelstechapparate)
•	
•	
•	

Workspecific layout training

Abwärme — Ackermann	
Fabrikationsprogramm: Beton- u. Mörtelmaschiner; Bauwände, Bauputzgerüste, Holzputzgerüste.	
Anteilseigner: Arthur Schumann, Rieh., Findleisen; Ernst Schumann sen.	Prokurstent: Karl Roth, Eduard Reinschmidt, Heinrich Böckeler, Arthur Schumann, Rieh., Findleisen; Ernst Schumann sen.
Gesellschafter: Arthur Schumann, Rieh., Findleisen; Ing.-O. Seltz.	Bankverbindungen: Reichsbank, Deutsche Bank und Disconto-Ges., Siegen.
Postcheck-Konto: Stadt- u. Girobank, Leipzig;	Postcheck-Konto: 878-Dortmund.
Geschäftsjahr: 1/7.—30/6.	Geschäftsjahr: 1/7.—30/6.
Grundbemess: 46.000 qm, davon 3500 qm bebaut.	Grundbemess: 45.000 qm, davon 10.000 qm bebaut.
Anlagen: Maschinenbauwerkstätten.	Anlagen: Gießerei u. Modellbaufabrik; Maschinenbauwerkstätten.
Besondere Angaben: Die Firma ist Fabrikationsnachfolgerin der in Konkurs geratenen „Allgemeinen Bau- und Eisenwaren-Ges., m. b. H.“, Leipzig o. L.	Besondere Angaben: Die Firma hat ihren Ursprung in dem im 15. Jahrhundert errichteten Hütten- und Hau- und Eisenwerk „Schumann, Findleisen & Co.“. Diese Firma lautete 1846 die jetzige Firma gegründet (Eisengießerei zur Herstellung von gußeisernen Ofen, Stufen, Treppen, Geländern, Eisenbeschlägen). Der Walzwerkbau den größten Teil der Produktion ein. Die Werkshallen sind mit mechanischen Bearbeitungs-Maschinen und Elektricitätsanlagen ausgestattet.
Gefolgschaft: 260 Mitglieder.	Gefolgschaft: 350 Mitglieder.
M. Achgelis Söhne A.-G.	
Maschinenfabrik u. Eisengießerei, Wesermünde-G, An der Zweighahn 1.	
Fernruf: 101 u. 140. Drahtanschrift: achgeliswerke	Fernruf: 101 u. 140. Drahtanschrift: achgeliswerke
Gründung: 1883; seit 1918 A.-G.	Gründung: 1883; seit 1918 A.-G.
Fabrikationsprogramm: Schiffshefmaschinen in jeder Art.	Fabrikationsprogramm: Schiffshefmaschinen in jeder Art.
Kapital: RM 300.000.—	Kapital: RM 225.000.—
Vorstand: Ing. Karl Boos, Georg Brinkmann, Weser-	Vorstand: Ing. Karl Boos, Georg Brinkmann, Weser-
Prokurstent: Obing, Wils., Barth.	Prokurstent: Obing, Wils., Barth.
Aufsichtsrat: Vors.; Arthur Friedrichs, Bremerhaven,	Aufsichtsrat: Vors.; Arthur Friedrichs, Bremerhaven,
Inhaber: Ing. Karl Boos, Georg Brinkmann, Weser-	Inhaber: Ing. Karl Boos, Georg Brinkmann, Weser-
Bankverbindungen: Reichsbank, Commerz- u. Privat-	Bankverbindungen: Reichsbank, Commerz- u. Privat-
Bank-A.G., Berlin;	Bank-A.G., Berlin;
Postcheck-Konto: 118-120; Berlin.	Postcheck-Konto: 118-120; Berlin.
Acetylenwerk Ebersbach	Acetylenwerk Ebersbach
Inh. Eug. Zinsler,	Inh. Eug. Zinsler,
Ebersbach-F. (Wtthg.)	Ebersbach-F. (Wtthg.)
Fernruf: 210. Drahtanschrift: acetylwerk	Fernruf: 210. Drahtanschrift: acetylwerk
Gründung: 1886; seit 1918 A.-G.	Gründung: 1886; seit 1918 A.-G.
Fabrikationsprogramm: Autopen - Schwilpapparate; Autogen-Werkzeuge (Schweiß-, Schneide- u. Lötherben-), Wasserstoff-, Wasser-Reduzivverstille;	Fabrikationsprogramm: Autogen-Schwilpapparate; Autogen-Werkzeuge (Schweiß-, Schneide- u. Lötherben-), Wasserstoff-, Wasser-Reduzivverstille;
Inhaber: Th. Friedrich.	Inhaber: Th. Friedrich.
Bankverbindungen: Gewerbebank, Ebersbach/Fz; Reichsbank, Geislingen-Wilhelmsburg; Ebersbach-Fz; Reichsbank.	Bankverbindungen: Gewerbebank, Ebersbach/Fz; Reichsbank, Geislingen-Wilhelmsburg; Ebersbach-Fz; Reichsbank.
Postcheck-Konto: 3258; Stuttgart.	Postcheck-Konto: 25.577; Frankfurt a/M; 8617; Ludwigsburg.
Geschäftsjahr: Kalenderjahr.	Geschäftsjahr: Kalenderjahr.
Grundbemess: 1500 qm, davon 1000 qm bebaut.	Grundbemess: 1500 qm, davon 1500 qm bebaut.
Anlagen: Apparations- u. Schweißerei u. Klempner-; Schmiede-; Schlosserei-; Montage-, Eisen- u. Metallfertigerei; Eigene Vertretung: in Berlin u. im Ausland.	Anlagen: Apparations- u. Schweißerei u. Klempner-; Schmiede-; Schlosserei-; Montage-, Eisen- u. Metallfertigerei; Eigene Vertretung: in Berlin u. im Ausland.
Tochtergesellschaft: Südd. Spindelwerk, Ebersbach/Fz.	Tochtergesellschaft: Südd. Spindelwerk, Ebersbach/Fz.
Gefolgschaft: 70 Arbeiter (1 Lehrling) und 7 Angestellte (1 Lehrling).	Gefolgschaft: 70 Arbeiter (1 Lehrling) und 7 Angestellte (1 Lehrling).
Achenbach Söhne G.m.b.H., Buschhütten (Kr. Siegen).	Achenbach Söhne G.m.b.H., Buschhütten (Kr. Siegen).
Fernruf: Siegen 5011. Drahtanschrift: achenbachsöhne	Fernruf: Siegen 5011. Drahtanschrift: achenbachsöhne
Gründung: 1846.	Gründung: 1846.
Fabrikationsprogramm: Walzwerksan- und Walzen-	Fabrikationsprogramm: Walzwerksan- und Walzen-
Kapital: RM 1.554.000.—	Kapital: RM 1.554.000.—

Base model segmentation

Abwärme — Ackermann	
Fabrikationsprogramm: Beton- u. Mörtelmaschiner; Bauwände, Bauputzgerüste, Holzputzgerüste.	
Anteilseigner: Frau Dr. Burton, Dr. Ing. Ernst Burton, Ernst Goebel (Geschäftsführer).	Anteilseigner: Frau Dr. Burton, Dr. Ing. Ernst Burton, Ernst Goebel (Geschäftsführer).
Kapital: RM 20.000.—	Kapital: RM 20.000.—
Antikurstent: Arthur Schumann, Rieh., Findleisen; Ernst Schumann sen.	Antikurstent: Arthur Schumann, Rieh., Findleisen; Ernst Schumann sen.
Gesellschafter: Arthur Schumann, Rieh., Findleisen; Ing.-O. Seltz.	Gesellschafter: Arthur Schumann, Rieh., Findleisen; Ing.-O. Seltz.
Bankverbindungen: Reichsbank, Deutsche Bank und Disconto-Ges., Siegen.	Bankverbindungen: Reichsbank, Deutsche Bank und Disconto-Ges., Siegen.
Postcheck-Konto: 878-Dortmund.	Postcheck-Konto: 79.000; Leipzig.
Geschäftsjahr: 1/7.—30/6.	Geschäftsjahr: 1/7.—30/6.
Grundbemess: 46.000 qm, davon 3500 qm bebaut.	Grundbemess: 45.000 qm, davon 10.000 qm bebaut.
Anlagen: Maschinenbauwerkstätten.	Anlagen: Gießerei u. Modellbaufabrik; Maschinenbauwerkstätten.
Besondere Angaben: Die Firma hat ihren Ursprung in dem im 15. Jahrhundert errichteten Hütten- und Hau- und Eisenwerk „Schumann, Findleisen & Co.“. Diese Firma lautete 1846 die jetzige Firma gegründet (Eisengießerei zur Herstellung von gußeisernen Ofen, Stufen, Treppen, Geländern, Eisenbeschlägen). Der Walzwerkbau den größten Teil der Produktion ein. Die Werkshallen sind mit mechanischen Bearbeitungs-Maschinen und Elektricitätsanlagen ausgestattet.	Besondere Angaben: Die Firma hat ihren Ursprung in dem im 15. Jahrhundert errichteten Hütten- und Hau- und Eisenwerk „Schumann, Findleisen & Co.“. Später nahmen die Herstellung von Walzen und der Walzwerkbau eine wichtige Rolle in der Produktion ein. Die Werkshallen sind mit mechanischen Bearbeitungs-Maschinen und Elektricitätsanlagen ausgestattet.
Gefolgschaft: 260 Mitglieder.	Gefolgschaft: 350 Mitglieder.
M. Achgelis Söhne A.-G.,	
Maschinenfabrik u. Eisengießerei, Wesermünde-G, An der Zweighahn 1.	
Fernruf: 101 u. 140. Drahtanschrift: achgeliswerke	Fernruf: 101 u. 140. Drahtanschrift: achgeliswerke
Gründung: 1883; seit 1918 A.-G.	Gründung: 1883; seit 1918 A.-G.
Fabrikationsprogramm: Schiffshefmaschinen in jeder Art.	Fabrikationsprogramm: Schiffshefmaschinen in jeder Art.
Kapital: RM 300.000.—	Kapital: RM 225.000.—
Vorstand: Ing. Karl Boos, Georg Brinkmann, Weser-	Vorstand: Ing. Karl Boos, Georg Brinkmann, Weser-
Prokurstent: Obing, Wils., Barth.	Prokurstent: Obing, Wils., Barth.
Aufsichtsrat: Vors.; Arthur Friedrichs, Bremerhaven,	Aufsichtsrat: Vors.; Arthur Friedrichs, Bremerhaven,
Inhaber: Ing. Karl Boos, Georg Brinkmann, Weser-	Inhaber: Ing. Karl Boos, Georg Brinkmann, Weser-
Bankverbindungen: Reichsbank, Commerz- u. Privat-	Bankverbindungen: Reichsbank, Commerz- u. Privat-
Bank-A.G., Berlin;	Bank-A.G., Berlin;
Postcheck-Konto: 118-120; Berlin.	Postcheck-Konto: 118-120; Berlin.
Abwärme Ausnutzung und Saugzug Ges. m. b. H. Abas,	
Berlin W. 35, Potsdamer Str. 28.	
Fernruf: 22.65.17. Drahtanschrift: abas	Fernruf: 22.65.17. Drahtanschrift: abas
Gründung: 1883; seit 1918 A.-G.	Gründung: 1883; seit 1918 A.-G.
Fabrikationsprogramm: Ventilatoren, Staubabscheider, Taschenfilterheizer.	Fabrikationsprogramm: Ventilatoren, Staubabscheider, Taschenfilterheizer.
Kapital: RM 200.000.—	Kapital: RM 200.000.—
Vorstand: Ing. Karl Boos, Georg Brinkmann, Weser-	Vorstand: Ing. Karl Boos, Georg Brinkmann, Weser-
Prokurstent: Obing, Wils., Barth.	Prokurstent: Obing, Wils., Barth.
Aufsichtsrat: Vors.; Arthur Friedrichs, Bremerhaven,	Aufsichtsrat: Vors.; Arthur Friedrichs, Bremerhaven,
Inhaber: Ing. Karl Boos, Georg Brinkmann, Weser-	Inhaber: Ing. Karl Boos, Georg Brinkmann, Weser-
Bankverbindungen: Reichsbank, Commerz- u. Privat-	Bankverbindungen: Reichsbank, Commerz- u. Privat-
Bank-A.G., Berlin;	Bank-A.G., Berlin;
Postcheck-Konto: 118-120; Berlin.	Postcheck-Konto: 118-120; Berlin.
Acetylenwerk Ebersbach	
Inh. Eug. Zinsler,	
Ebersbach-F. (Wtthg.)	
Fernruf: 210. Drahtanschrift: acetylwerk	Fernruf: 210. Drahtanschrift: acetylwerk
Gründung: 1886; seit 1918 A.-G.	Gründung: 1886; seit 1918 A.-G.
Fabrikationsprogramm: Autogen-Schwilpapparate; Autogen-Werkzeuge (Schweiß-, Schneide- u. Lötherben-), Wasserstoff-, Wasser-Reduzivverstille;	Fabrikationsprogramm: Autogen-Schwilpapparate; Autogen-Werkzeuge (Schweiß-, Schneide- u. Lötherben-), Wasserstoff-, Wasser-Reduzivverstille;
Inhaber: Th. Friedrich.	Inhaber: Th. Friedrich.
Bankverbindungen: Gewerbebank, Ebersbach/Fz; Reichsbank, Geislingen-Wilhelmsburg; Ebersbach-Fz; Reichsbank.	Bankverbindungen: Gewerbebank, Ebersbach/Fz; Reichsbank, Geislingen-Wilhelmsburg; Ebersbach-Fz; Reichsbank.
Postcheck-Konto: 3258; Stuttgart.	Postcheck-Konto: 25.577; Frankfurt a/M; 8617; Ludwigsburg.
Geschäftsjahr: Kalenderjahr.	Geschäftsjahr: Kalenderjahr.
Grundbemess: 1500 qm, davon 1000 qm bebaut.	Grundbemess: 1500 qm, davon 1500 qm bebaut.
Anlagen: Apparations- u. Schweißerei u. Klempner-; Schmiede-; Schlosserei-; Montage-, Eisen- u. Metallfertigerei; Eigene Vertretung: in Berlin u. im Ausland.	Anlagen: Apparations- u. Schweißerei u. Klempner-; Schmiede-; Schlosserei-; Montage-, Eisen- u. Metallfertigerei; Eigene Vertretung: in Berlin u. im Ausland.
Tochtergesellschaft: Südd. Spindelwerk, Ebersbach/Fz.	Tochtergesellschaft: Südd. Spindelwerk, Ebersbach/Fz.
Gefolgschaft: 70 Arbeiter (1 Lehrling) und 7 Angestellte (1 Lehrling).	Gefolgschaft: 70 Arbeiter (1 Lehrling) und 7 Angestellte (1 Lehrling).
Peter Acker,	
Gau-Odernheim (Rh. H.)	
Fernruf: 220. Drahtanschrift: achenbachsöhne	Fernruf: 220. Drahtanschrift: achenbachsöhne
Gründung: 1878	Gründung: 1878
Fabrikationsprogramm: Hohenförderer u. Pflege,	Fabrikationsprogramm: Hohenförderer u. Pflege,
Kapital: RM 70.000.—	Kapital: RM 70.000.—
Inhaber: Johann Acker, Jakob Acker	Inhaber: Johann Acker, Jakob Acker
Bankverbindungen: Volksbank, Alzey; Spars- u. Disconto-Ges., Neuwied.	Bankverbindungen: Volksbank, Alzey; Spars- u. Disconto-Ges., Neuwied.
Postcheck-Konto: 25.577; Frankfurt a/M; 8617; Ludwigs-	Postcheck-Konto: 25.577; Frankfurt a/M; 8617; Ludwigs-
Geschäftsjahr: Kalenderjahr.	Geschäftsjahr: Kalenderjahr.
Grundbemess: 1500 qm, davon 1500 qm bebaut.	Grundbemess: 1500 qm, davon 1500 qm bebaut.
Anlagen: Apparations- u. Schweißerei u. Klempner-; Schmiede-; Schlosserei-; Montage-, Eisen- u. Metallfertigerei; Eigene Vertretung: in Berlin u. im Ausland.	Anlagen: Apparations- u. Schweißerei u. Klempner-; Schmiede-; Schlosserei-; Montage-, Eisen- u. Metallfertigerei; Eigene Vertretung: in Berlin u. im Ausland.
Tochtergesellschaft: Südd. Spindelwerk, Ebersbach/Fz.	Tochtergesellschaft: Südd. Spindelwerk, Ebersbach/Fz.
Gefolgschaft: 70 Arbeiter (1 Lehrling) und 7 Angestellte (1 Lehrling).	Gefolgschaft: 70 Arbeiter (1 Lehrling) und 7 Angestellte (1 Lehrling).
Achenbach Söhne G.m.b.H.,	
Buschhütten (Kr. Siegen).	
Fernruf: Siegen 5011. Drahtanschrift: achenbachsöhne	Fernruf: Siegen 5011. Drahtanschrift: achenbachsöhne
Gründung: 1846.	Gründung: 1846.
Fabrikationsprogramm: Walzwerksan- und Walzen-	Fabrikationsprogramm: Walzwerksan- und Walzen-
Kapital: RM 1.554.000.—	Kapital: RM 1.554.000.—

Finetuned model segmentation

Weak point: Textlines

In text line recognition, the baseline is not always correct and sometimes lines are divided into small areas.

Anlagen:	Maschinen-, Werkzeugfabrik u. Kaltwalzwerk.
Gefolgschaft:	1930/31 31/32 32/33 33/34 34/35 35/36
Arbeiter	137 87 96 102 108 109
Angestellte	16 15 15 16 17 17
Umsatz (Mill RM)	1,1 0,9 0,7 0,85 0,85 0,9

Beispiel: Textzeilensegmentierung einer Tabelle

Work-specific training

Training material 26 pages

Results text regions Significant improvement

Results text lines Significant improvement for further processing

cbad_1800

Anlagen: Maschinen-, Werkzeugfabrik u. Kaltwalzwerk.						
Gefolgschaft:	1930/31	31/32	32/33	33/34	34/35	35/36
Arbeiter	137	87	96	102	108	109
Angestellte	16	15	15	16	17	17
Umsatz (Mill RM)	1,1	0,9	0,7	0,85	0,85	0,9

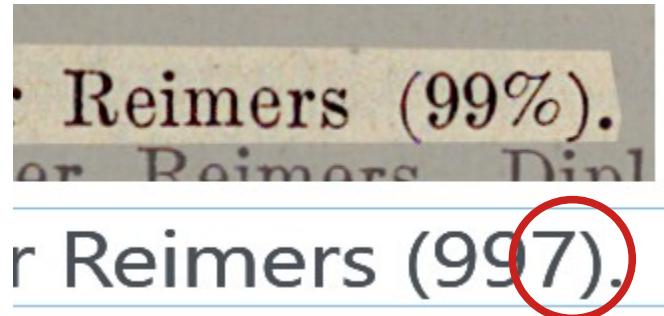
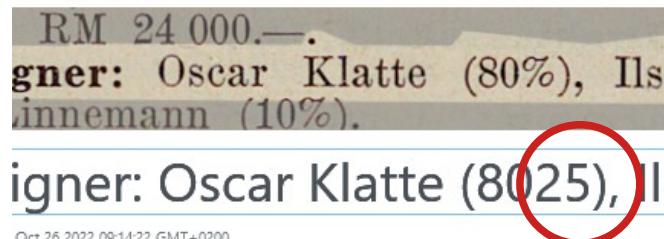
cbad_1800_trained

Anlagen: Maschinen-, Werkzeugfabrik u. Kaltwalzwerk.						
Gefolgschaft:	1930/31	31/32	32/33	33/34	34/35	35/36
Arbeiter	137	87	96	102	108	109
Angestellte	16	15	15	16	17	17
Umsatz (Mill RM)	1,1	0,9	0,7	0,85	0,85	0,9

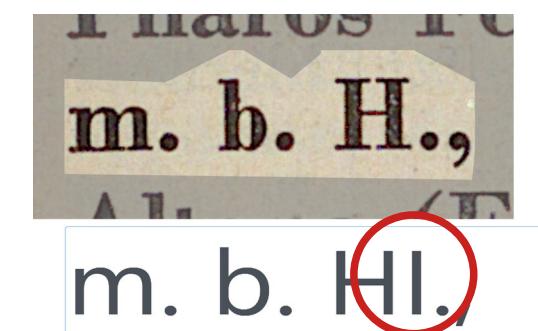
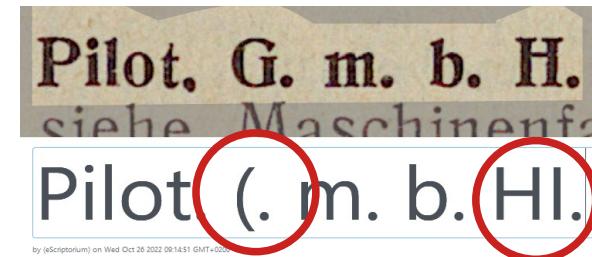
Weaknesses in the text recognition

The model already delivers very good results with text recognition rates > 98 %.
Some specific errors can be identified:

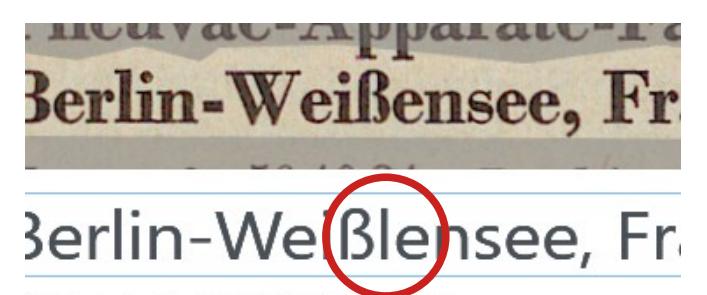
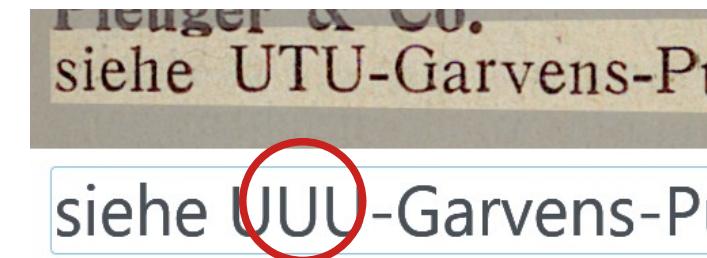
Percentage sign



String G.m.b.H.



Rare characters /
character combinations



Work-specific training

Training material 26 pages

Results text recognition The recognition rate could be increased to > 99.85 %.

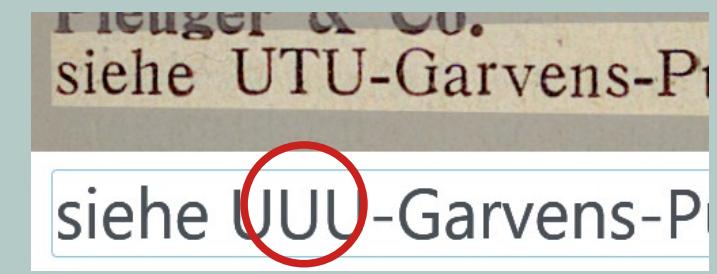
Percentage sign



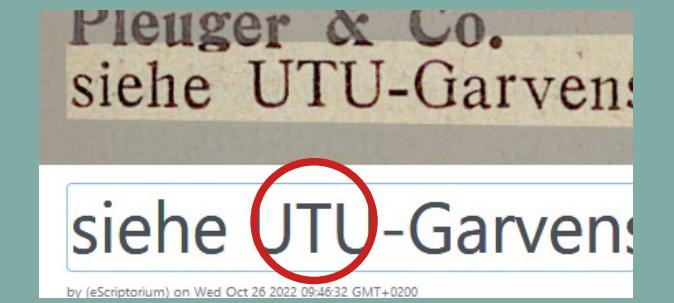
String G.m.b.H.



Rare characters / character combinations



digitue_trained



Structuring unstructured text

Blatt

pypi package 0.1.6

NLP-helper for OCR-ed pages in [PAGE XML](#) format.

Installation

pip install `blatt`



- <https://pypi.org/project/blatt>
- <https://github.com/UB-Mannheim/blatt>

Functionality

- Reads PAGE-XML of OCR-ed text
- Does geometric segmentation based on coordinates of segments
- Joins segments splitted into two or more pages
- Removes hyphens and new lines
- Uses a colon „:“ as a separator for structuring
- Saves structured data as a CSV-file

Knowledge Graph creation

Digital image

**Piccolo-Automaten G. m. b. H.,
Berlin W 35, Kurfürstenstraße 146.**

Fernruf: 21 20 95. Drahtanschrift: piccoloautomat
Gründung: 1932.

Fabrikationsprogramm: Schokoladen-Verkaufsapparate
(Tischautomaten, Kugelstechapparate).

Kapital: RM 20 000.—.

Anteileigner: Dr. Richard Schönthal, Wien.

Geschäftsführer: Erwin Hantsch, Bln.-Steglitz.

Bankverbindung: Deutsche Bank u. Disconto - Ges.,
Berlin.

Postscheck-Konto: 8012 Berlin.

Geschäftsjahr: 1./1.—31./12.

Anlagen: Mechanische Werkstätten.

Gefolgschaft: 13 Arbeiter u. 6 Angestellte.

Structured data (e.g., CSV)

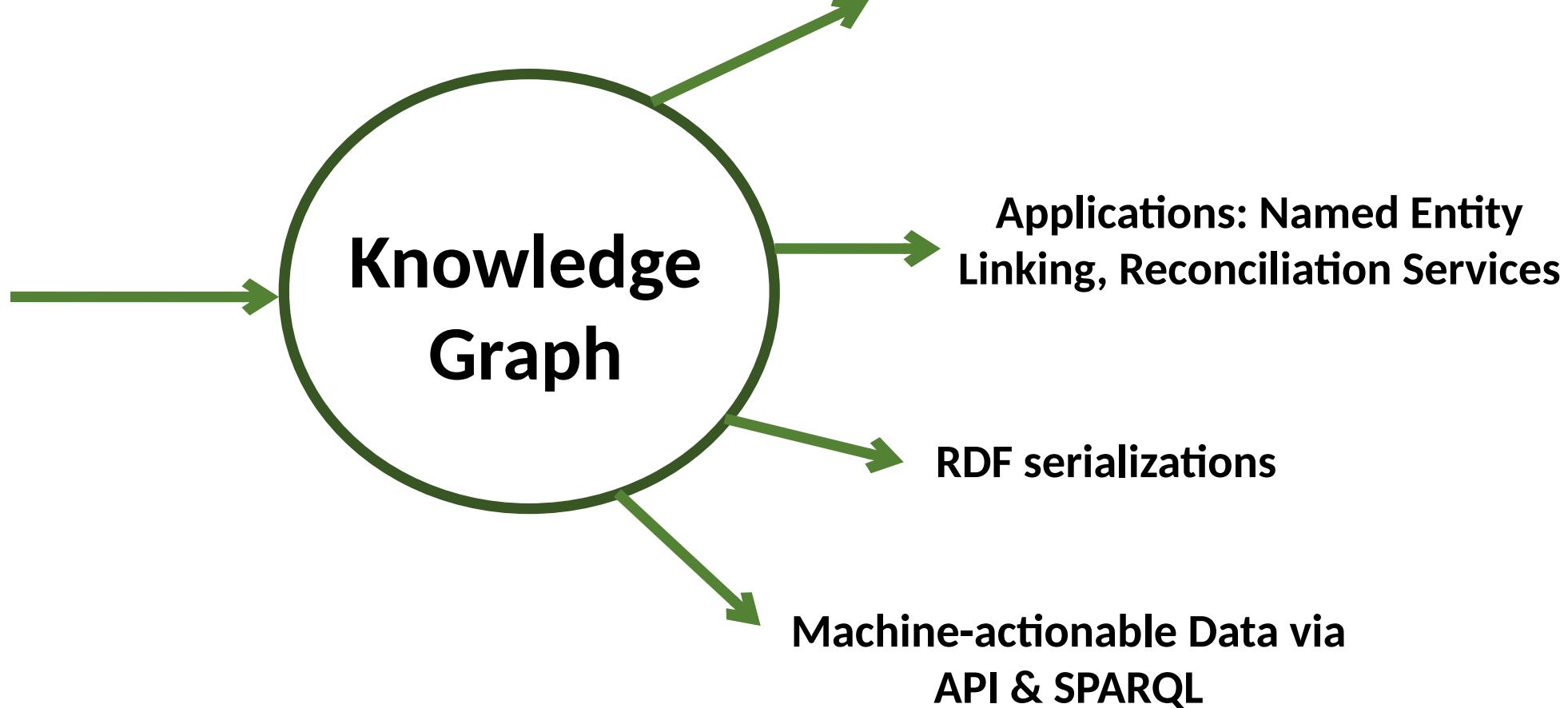
Firmenname	Piccolo-Automaten
Rechtsform	G.m.b.H.
Sitz	Berlin W 35 Kurfürstenstraße 146
Fernruf	212095
Drahtanschrift	piccoloautomat
Gründung	1932
Postscheck-Konto	8012 Berlin

Actionable FAIR data

- Findable
- Accessible
- Interoperable
- Reusable

Scanning, OCR, structuring, semantifying: towards actionable FAIR data

Firmenname	Piccolo-Automaten
Rechtsform	G.m.b.H.
Sitz	Berlin W 35 Kurfürstenstraße 146
Fernruf	212095
Drahtanschrift	piccoloautomat
Gründung	1932
Postscheck-Konto	8012 Berlin



<https://wikiba.se>



- Open-source production-ready
- Many supporting tools and extensions
- Growing community of users & developers
- Export in CSV, JSON and RDF-serialisations
- User-friendly frontend for manual edits
- History of changes is saved
- Every data value is citable

<https://www.wikidata.org>



- 102+ million entities
- Collaborative editing by both humans and machines
- Strong impact on science and data science

Knowledge Graph (KG) creation workflow

Firmenname	Piccolo-Automaten
Rechtsform	G.m.b.H.
Sitz	Berlin W 35 Kurfürstenstraße 146
Fernruf	212095
Drahtanschrift	piccoloautomat
Gründung	1932
Postscheck-Konto	8012 Berlin



- Create the initial properties with datatypes „strings“ corresponding to the properties in the book
- Upload the data via API using the Python scripts and WikidataIntegrator-library
- Semantify:
 - Match cities to geographic entities at Wikidata using the OpenRefine reconciliation API. Create geo-entities, enrich them with data from Wikidata.
 - Create entities for legal forms, classes and people.



**Knowledge
Graph**

← → ⌂ https://mbi.kgi.uni-mannheim.de 80%

≡ MBI

Inhaltsverzeichnis

Overview

Access to data

Data model

Use Cases

Get involved

MaschinenBauIndustrie

Aus MBI

Quelltext anzeigen

Versionsgeschichte

Diskussion

0 A

⋮

Overview

The MaschinenBauIndustrie (MBI) Knowledge Graph contains structured data from the book "[Die Maschinen-Industrie im Deutschen Reich](#)" written by Herbert Patschan in 1937. The book was scanned, OCR-ed, structured and semantified at [UB Mannheim](#). The data includes basic data for the German companies from machine industry.

Access to data

- The MBI data is accessible by both humans and machines
- The MBI data is currently not available openly
- [List of properties](#)
- [List of items](#)
- [Advanced search for items](#)
- [SPARQL endpoint](#)
- [API](#)
- [Export to CSV](#)
- Every entity can be downloaded in .json, .rdf, .ttl, .nt and .jsonld formats. An example for Daimler-Benz A.-G.: <https://mbi.kgi.uni-mannheim.de/entity/Q707.json>, <https://mbi.kgi.uni-mannheim.de/entity/Q707.rdf>, <https://mbi.kgi.uni-mannheim.de/entity/Q707.ttl>, <https://mbi.kgi.uni-mannheim.de/entity/Q707.nt> and <https://mbi.kgi.uni-mannheim.de/entity/Q707.jsonld>.



MBI-KG: Page for “Daimler-Benz A.-G.”

<https://mbi.kgi.uni-mannheim.de/wiki/Item:Q707>

Daimler-Benz A.-G. (Q707)

Aus MBI

Keine Beschreibung vorhanden

► In weiteren Sprachen

Basisdaten

GRÜNDUNG <small>Englisch</small>	1890
	▼ 0 Fundstellen

instance of <small>Englisch</small>	company <small>Englisch</small>
	▼ 0 Fundstellen

CITY <small>Englisch</small>	Stuttgart-Untertürkheim
	▼ 0 Fundstellen

STREET <small>Englisch</small>	Mercedes-Straße 53
	▼ 0 Fundstellen

POSTSCHECK-KONTO <small>Englisch</small>	470 Stuttgart
	▼ 0 Fundstellen

string

<https://mbi.kgi.uni-mannheim.de/wiki/Item:Q707>

POSTSCHECK-KONTO Englisch ▼ 0 Fundstellen

FERNRUF Englisch □ 30 141, 30 241, 31 141, 31 151

▼ 0 Fundstellen

DRAHTANSCHRIFT Englisch □ daimlerbenz untertürkheim

▼ 0 Fundstellen

legal form Englisch □ Aktiengesellschaft Englisch

▼ 0 Fundstellen

inception Englisch □ 1890

▼ 0 Fundstellen

country Englisch □ German Reich Englisch

▼ 0 Fundstellen

headquarters location Englisch □ Stuttgart-Untertürkheim Englisch

▼ 0 Fundstellen

thing

<https://mbi.kgi.uni-mannheim.de/wiki/Item:Q707>

Rohdaten

RAW_TEXT Englisch

□ Daimler-Benz A.-G.,\nStuttgart-Untertürkheim,\n\nMercedes-Straße 53.\nFernruf: 30 141, 30 241, 31
141, 31 151.\nDrahtanschrift: daimlerbenz
untertürkheim\nGründung:
1890.\nFabrikationsprogramm: Personenkraftwagen,
Lastkraft-\nwagen (Benzin- u. Dieselantrieb),
Lieferungswagen, Spe-\nzialfahrzeuge aller Art wie
Omnibusse, Sattelschlepper,\nFeuerwehrfahrzeuge,
Kommunalfahrzeuge; Luftfahrt-\nSchiffs- und
Bootsmotore, Triebwagenmotore, schnell und\nlangsam
laufende Dieselmotore für ortsfeste
Verwendung.\nSpezial- und Großmotore,
Rohölmotore.\nKapital: 26 172 000.–.\nVorstand: Dir.
Wolfgang von Hentig, Berlin; Dir. Dr.\nning. Otto Hoppe,
Dir. Dr. ing. e. c. Wilhelm Kissel, Dir.\nCarl Schippert,
Stuttgart-Untertürkheim; Dir. Jakob Wer-\nlin, München;
Stellv.: Dir. Dr. ing. Wilhelm Haspel, Dir.\nMax Sailer,
Stuttgart-Untertürkheim; Dir. Arnold Frei-\nherr Gedult
von Jungenfeld, Gaggenau; Dir. Carl C.,\nMüller, Berlin-\nMarienfelde.\nAufsichtsrat: Vors.: Dr. E. G. von Stauss,
Staatsrat,\nBerlin; Stellv.: Hans Rummel,
Vorstandsmitglied der\nDeutschen Bank u. Disconto-\nGes., Berlin; Dr. Carl Jahr,\nKommerzienrat,
Mannheim.\nBankverbindungen: Deutsche Bank u.
Disconto-Ges.,\nCommerz- u. Privat-Bank A.-G.,
Dresdner Bank, Bank\nnder Deutschen Arbeit A.-G.,
Reichsbank-Giro-Konto,\nStädtische Girokasse,
Stuttgart.\nPostscheck-Konto: 470
Stuttgart.\nGeschäftsjahr: Kalenderjahr.\nGrundbesitz: 2
073 904 qm, davon 569 160 qm bebaut.\nAnlagen:
Werke in: Stuttgart-Untertürkheim; Mann-\nheim;

MBI-KG: Page for “Stuttgart-Untertürkheim”

<https://mbi.kgi.uni-mannheim.de/wiki/Item:Q5452>

  <https://www.wikidata.org/wiki/Q897283>

Stuttgart-Untertürkheim Englisch (Q5452)

Aus MBI

Keine Beschreibung vorhanden

► In weiteren Sprachen

Basisdaten

instance of Englisch

administrative territorial entity of Germany 

▼ 0 Fundstellen

Aussagen

coordinate location Englisch

48°46'48.7"N, 9°15'3.2"E 

▼ 0 Fundstellen

Identifikatoren

Wikidata QID Englisch

Q897283 

▼ 0 Fundstellen

(Q5452)

tem Discussion

Untertürkheim (Q897283)

district of Stuttgart, Germany

► In more languages

Statements

instance of

Stadtbezirk

▼ 0 references

image



UT-Panorama.jpg
716 × 156; 33 KB

▼ 0 references

country

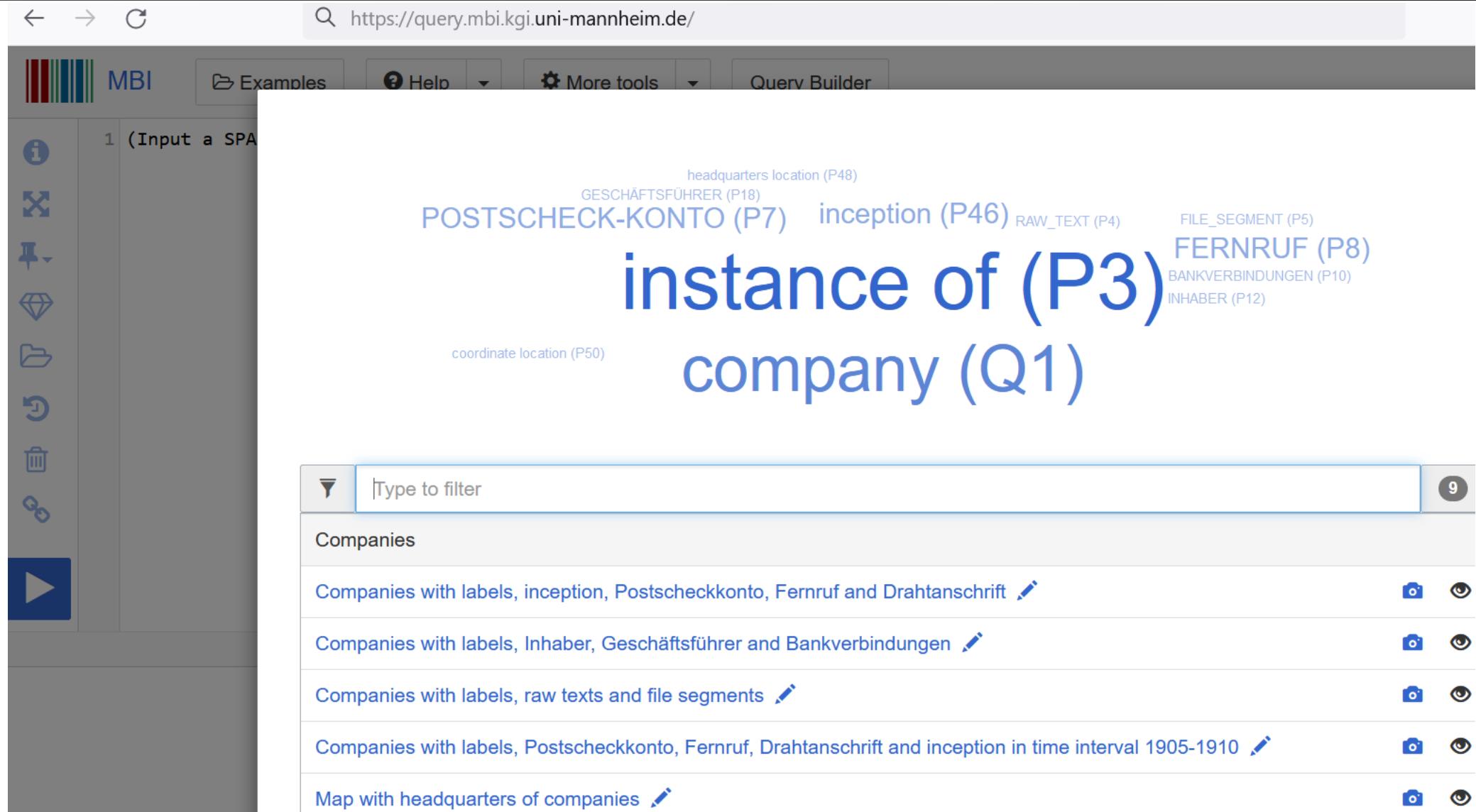
Germany

► 1 reference

located in the administrative area of Stuttgart

Knowledge Graph applications

MBI-KG: SPARQL endpoint



The screenshot shows a web-based SPARQL endpoint interface for the MBI-KG dataset. The URL in the address bar is <https://query.mbi.kgi.uni-mannheim.de/>. The interface includes a navigation bar with back, forward, and refresh buttons, and links for Examples, Help, More tools, and Query Builder. On the left, there is a sidebar with various icons for navigation and search. The main content area displays a search result for the query "instance of (P3) company (Q1)". The result is presented in a large, bold blue font. Surrounding the main term are several blue labels representing properties and their types: "headquarters location (P48)", "GESCHÄFTSFÜHRER (P18)", "POSTSCHECK-KONTO (P7)", "inception (P46)", "RAW_TEXT (P4)", "coordinate location (P50)", "FILE_SEGMENT (P5)", "FERNRUF (P8)", "BANKVERBINDUNGEN (P10)", "INHABER (P12)". Below the main term, a filter bar with a magnifying glass icon and a "Type to filter" input field is shown. A list of filters is provided, each with a pencil icon for editing and eye icons for viewing: "Companies", "Companies with labels, inception, Postscheckkonto, Fernruf and Drahtanschrift", "Companies with labels, Inhaber, Geschäftsführer and Bankverbindungen", "Companies with labels, raw texts and file segments", "Companies with labels, Postscheckkonto, Fernruf, Drahtanschrift and inception in time interval 1905-1910", and "Map with headquarters of companies". A notification badge with the number "9" is visible in the top right corner of the filter bar.

MBI-KG: SPARQL endpoint

MBI-KG: SPARQL endpoint



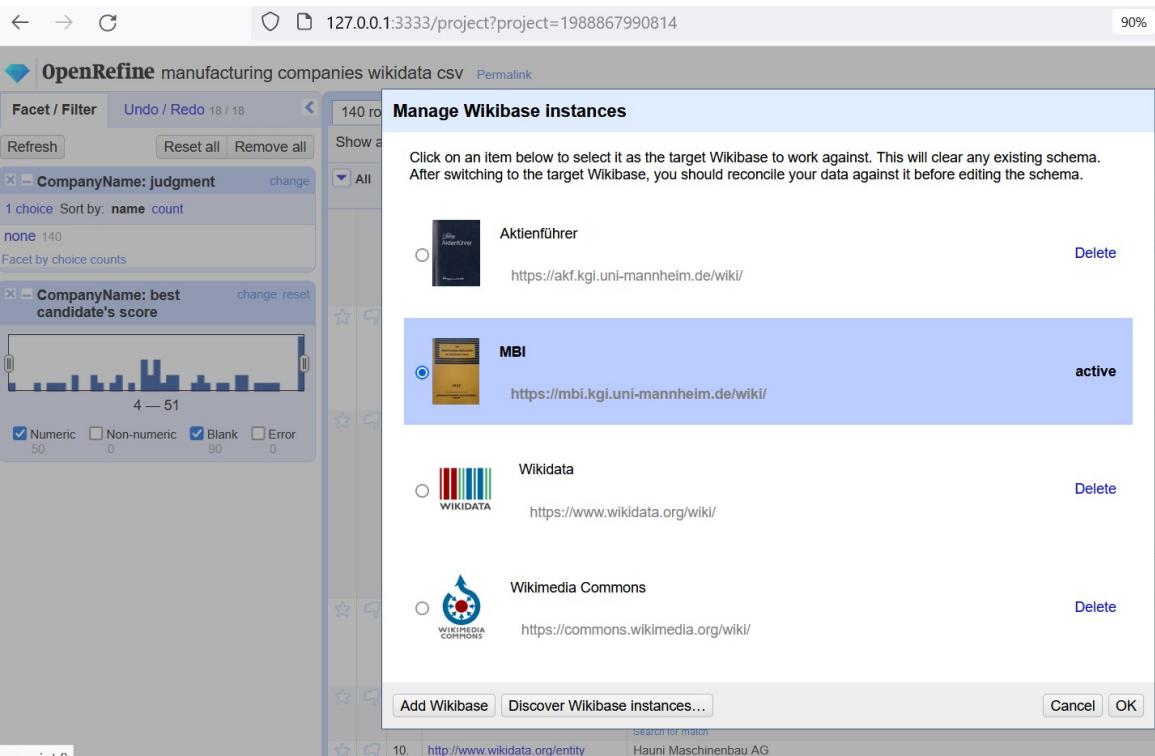
BERD@NFDI





OpenRefine

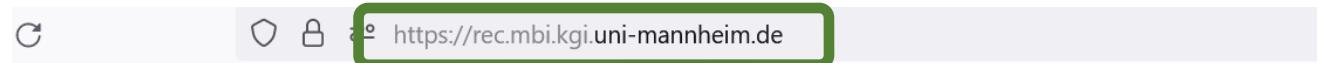
OpenRefine is connected to MBI KG



The screenshot shows the OpenRefine interface with a facet analysis of company names. The 'Manage Wikibase instances' dialog is open, listing four Wikibase instances:

- Aktienführer
- MBI (selected and marked as active)
- Wikidata
- Wikimedia Commons

The 'MBI' entry shows the URL <https://mbi.kgi.uni-mannheim.de/wiki/>.



MBI reconciliation for OpenRefine

This web service can be used to align datasets to MBI items in OpenRefine.

Use the following URL in OpenRefine: <https://rec.mbi.kgi.uni-mannheim.de/en/api>.

Use the following URL to access the manifest.json: <https://rec.mbi.kgi.uni-mannheim.de/static/manifest.json>.

Replacing "en" by another language code will display items and properties in your language, when they are available.

This interface works with OpenRefine from 2.6 rc2 onwards. It is not compatible with Google Refine.

- Documentation of the protocol
- Documentation of this application for developers and Wikibase admins
- Source
- Bug tracker
- Service status

Tutorials for OpenRefine users

- Official documentation in English
- Official documentation in Japanese by 山本泰智
- Tutorial by Antoine Courtin (in French)
- Comparison of reconciliation services for OpenRefine by Aaron Tay (in English)
- Using OpenRefine to Reconcile Name Entities by Karen H. (in English)
- Reconciling in OpenRefine - eine Einführung von histHub (in German)

Videos

MBI-KG: Data matching and enrichment via OpenRefine



BERD@NFDI

OpenRefine manufacturing companies wikidata csv [Permalink](#)

Facet / Filter Undo / Redo 19 / 19 <

Refresh Reset all Remove all

CompanyName: judgment change invert reset
2 choices Sort by: name count
matched 15 exclude
none 125
Facet by choice counts

CompanyName: best candidate's score change reset

4 — 51
 Numeric 15 Non-numeric 0 Blank 0 Error 0

15 matching rows (140 total) Schema Issues 1 Preview Extensions Wikibase ▾

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first < previous 1 ▾ next » last »

All	WikidataQID	CompanyName	Qid	FERNRUF	DRAHTANSCHRIFT	headquarters location	legal form
6.	http://www.wikidata.org/entity/Q163580	Henschel & Sohn A.-G. Choose new match	Q1744	5521	lokomotive	Kassel Choose new match	Aktiengesellschaft Choose new match
12.	http://www.wikidata.org/entity/Q314543	Maschinenfabrik Eßlingen Choose new match	Q2800	Stuttgart 30 656	maschinenfabrik		
18.	http://www.wikidata.org/entity/Q465370	DEMAG Aktiengesellschaft Choose new match	Q739	26 471	demag	Duisburg Choose new match	Aktiengesellschaft Choose new match
30.	http://www.wikidata.org/entity/Q831095	Beth Choose new match	Q404				Aktiengesellschaft Choose new match
61.	http://www.wikidata.org/entity/Q1616970	Heyligenstaedt & Comp. G. m. b. H. Werkzeugmaschinenfabrik Choose new match	Q1809	2941/42	heylingenstaedt		GmbH Choose new match
78.	http://www.wikidata.org/entity/Q1914433	Mayer & Cie Choose new match	Q2987				
79.	http://www.wikidata.org/entity/Q1920279	Menck & Hambrock G. m. b. H. Maschinenfabrik Choose new match	Q3024	42 91 19	menckbrock	Altona Choose new match	GmbH Choose new match
81.	http://www.wikidata.org/entity/Q1949805	Motorenfabrik Hatz G. m. b. H. Choose new match	Q3150	Pocking 50	hatzmotor	Ruhstorf (Rott) Choose new match	GmbH Choose new match
84.	http://www.wikidata.org/entity/Q1998329	Nordischer Maschinenbau Rud. Baader Choose new match	Q3346	20 863	nordmaschine		Kommanditgesellschaft Choose new match
103.	http://www.wikidata.org/entity/Q2583690	Windmöller & Hölscher G. m. b. H. Choose new match	Q4996	291	maschinenfabrik lengerichwestfl		GmbH Choose new match
114.	http://www.wikidata.org/entity/Q15811438	Gebr. Becker Choose new match	Q304	Hamburg 36 35 86 u. Wedel 145		Hamburg Choose new match	
131.	http://www.wikidata.org/entity/Q55262762	Maschinenfabrik Buckau R. Wolf Aktiengesellschaft Choose new match	Q2786	41 942 u. 42 551	maschinenbau (u.) lokomobile		Aktiengesellschaft Choose new match
135.	http://www.wikidata.org/entity/Q65966715	Maschinenfabrik Moenus A.-G. Choose new match	Q2876	70 411	moenus	Frankfurt (M.)-West Choose new match	Aktiengesellschaft Choose new match
136.	http://www.wikidata.org/entity/Q66133341	Kölsch-Fölzer-Werke A. G. Choose new match	Q2287	5301	kölschfölzer		Aktiengesellschaft Choose new match
137.	http://www.wikidata.org/entity	Eisenwerk Wülfel	Q988	68 61 61	eisenwerk	Hannover-Wülfel	

MBI-KG: Federated query over MBI-KG and AKF-KG

https://query.mbi.kgi.uni-mannheim.de/#%23Companies in MBI-KG with the labels identical to aliases of companies at Wikidata

```
1 #Companies in MBI-KG with the labels identical to aliases of companies at Wikidata
2 PREFIX wd-wd: <http://www.wikidata.org/entity/>
3 PREFIX wd-wdt: <http://www.wikidata.org/prop/direct/>
4 SELECT DISTINCT ?MBI_company ?MBI_companyLabel ?WD_company ?WD_companyLabel WHERE {
5   ?MBI_company rdfs:label ?MBI_companyLabel;
6   wdt:P3 wd:Q1.
7   FILTER((LANG(?MBI_companyLabel)) = "de")
8   SERVICE <https://query.wikidata.org/sparql> {
9     ?WD_company rdfs:label ?WD_companyLabel,
10    skos:altLabel ?MBI_companyLabel;
11    wd-wdt:P452 ?industry;
12    wd-wdt:P17 wd-wd:Q183.
13    FILTER((LANG(?WD_companyLabel)) = "de")
14  }
15 }
```

Table ▾ ? 13 results in 21416 ms </> Code

MBI_company	MBI_companyLabel	WD_company	WD_companyLabel
Q <http://mbi.kgi.uni-mannheim.de/entity/Q289>	Bayerische Motoren Werke Aktiengesellschaft	wd:Q26678	BMW
Q <http://mbi.kgi.uni-mannheim.de/entity/Q452>	Blohm & Voß	wd:Q218715	Blohm + Voss
Q <http://mbi.kgi.uni-mannheim.de/entity/Q687>	Continental	wd:Q163241	Continental AG

MBI-KG: Federated query over MBI-KG and AKF-KG

Screenshot of a web browser showing a federated SPARQL query results page.

The URL in the address bar is: <https://query.mbi.kgi.uni-mannheim.de/> #%23MBI-KG-companies and AKF-KG-companies with identi

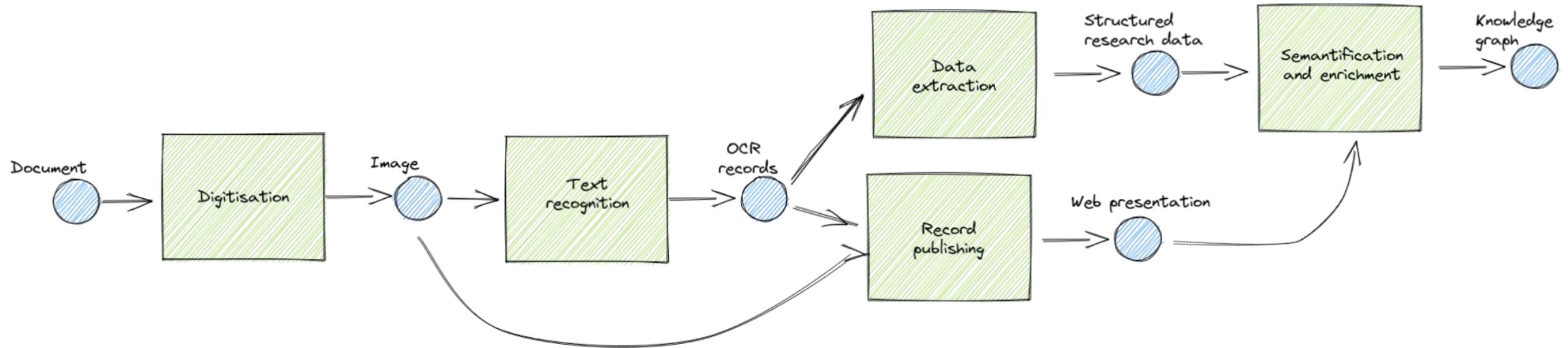
```

2 PREFIX akfd: <http://akf.kgi.uni-mannheim.de/entity/>
3 PREFIX akfdt: <http://akf.kgi.uni-mannheim.de/prop/direct/>
4 SELECT DISTINCT ?MBI_company ?MBI_companyLabel ?AKF_company ?AKF_companyLabel WHERE {
5   ?MBI_company rdfs:label ?MBI_companyLabel;
6   wdt:P3 wd:Q1.
7   FILTER((LANG(?MBI_companyLabel)) = "de")
8   SERVICE <https://query.akf.kgi.uni-mannheim.de/proxy/wdq/bigdata/namespace/wdq/sparql> {
9     ?AKF_company rdfs:label ?AKF_companyLabel;
10    akfdt:P346 ?MBI_companyLabel;
11    akfdt:P325 akfd:Q1.
12    FILTER((LANG(?AKF_companyLabel)) = "de")
13  }
14 }
```

The results table shows 8 results in 8695 ms:

MBI_company	MBI_companyLabel	AKF_company	AKF_companyLabel
<http://mbi.kgi.uni-mannheim.de/entity/Q289>	Bayerische Motoren Werke Aktiengesellschaft	Q <http://akf.kgi.uni-mannheim.de/entity/Q434>	Bayerische Motoren Werke AG
<http://mbi.kgi.uni-mannheim.de/entity/Q536>	Braunschweigische Maschinenbauanstalt	Q <http://akf.kgi.uni-mannheim.de/entity/Q466>	BMA Braunschweigische Maschinenbauanstalt
<http://mbi.kgi.uni-mannheim.de/entity/Q569>	Brown Boveri & Cie Aktiengesellschaft	Q <http://akf.kgi.uni-mannheim.de/entity/Q473>	ABB AG
<http://mbi.kgi.uni-mannheim.de/entity/Q739>	DEMAG Aktiengesellschaft	Q <http://akf.kgi.uni-mannheim.de/entity/Q718>	DEMAG AG
<http://mbi.kgi.uni-mannheim.de/entity/Q988>	Eisenwerk Wülfel	Q <http://akf.kgi.uni-mannheim.de/entity/Q548>	Eisenwerk Wülfel

Thank you for your attention



- 1) https://gitlab.com/scripta/escriptorium/-/blob/develop/app/escriptorium/static/images/escriptorium_hd.png
- 2) <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjamo3bi97-AhWuhf0HZdcDcMQFnoECAkQAQ&url=https%3A%2F%2Fwww.theses.fr%2F2021UPSLP023.pdf&usg=AOvVaw05nUc2Tj-5cUDB5n1koxOj>
- 3) <https://kraken.re/main/index.html>
- 4) https://commons.wikimedia.org/wiki/File:Python_logo_and_wordmark.svg
- 5) https://commons.wikimedia.org/wiki/File:Wikibase_logo.svg#/media/Datei:Wikibase_logo.svg
- 6) <https://commons.wikimedia.org/wiki/File:Wikidata-logo-en.svg#/media/Datei:Wikidata-logo-en.svg>
- 7) <https://openrefine.org/>