Forschungsdaten

aus Digitalisaten

Jan Kamlah, Stefan Weil
Universitätsbibliothek Mannheim

E-Science-Tage Heidelberg, 28.03.2019

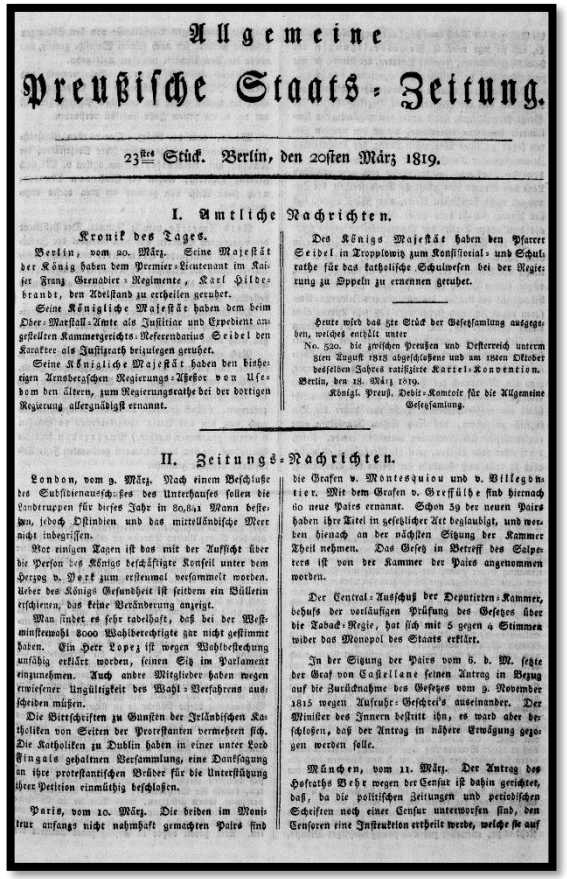
Motivation für Texterkennung und Strukturierung bei Digitalisierungsprojekten

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

- **Digitalisate**
  - Online Zugriff weltweit
  - Digitale Bestandserhaltung
- **Text durch Texterkennung (OCR)**
  - Recherchemöglichkeiten (Such-Funktion)
  - Kopiermöglichkeiten für den Volltext
- **Strukturierte Daten aus dem Volltext**
  - Strukturierte Suche
  - Auswertungen für Forschungsfragen

Übersicht

Digitalisate für Forschungsdaten an der UB Mannheim



Allgemeine
Preußische Staats-Zeitung.

23tes Stück. Berlin, den 20sten März 1819.

I. Amtliche Nachrichten.

Kronik des Tages.

Berlin, vom 20. März. Seine Majestät der König haben dem Premier-Lieutenant im Kaiser Franz Grenadier-Regimente, Karl Hildebrandt, den Adelstand zu ertheilen geruhet.

Seine Königliche Majestät haben dem beim Ober-Marstall-Amte als Justitiar und Expedient angestellten Kammergerichts-Referendarius Seidel den Karakter als Justizrath beizulegen geruhet.

Seine Königliche Majestät haben den bisherigen Aemsbergischen Regierungs-Assessor von Usedom den ältern, zum Regierungsrathe bei der dortigen Regierung allergnädigst ernannt.

Des Königs Majestät haben den Pfarrer Seidel in Tropplowitz zum Konsistorial- und Schulrathe für das katholische Schulwesen bei der Regierung zu Oppeln zu ernennen geruhet.

Heute wird das 5te Stück der Gesetzsammlung ausgegeben, welches enthält unter
No. 520. die zwischen Preußen und Oesterreich unterm 31sten August 1818 abgeschlossene und am 18ten Oktober desselben Jahres ratifizirte Kartel-Konvention.
Berlin, den 18. März 1819.
Königl. Preuß. Debit-Komtoir für die Allgemeine Gesetzsammlung.

II. Zeitungs-Nachrichten.

London, vom 9. März. Nach einem Beschlusse des Subsidienausschusses des Unterhauses sollen die Landtruppen für dieses Jahr in 80,821 Mann bestehen, jedoch Ostindien und das mittelländische Meer nicht inbegriffen.

Vor einigen Tagen ist das mit der Aufsicht über die Person des Königs beschäftigte Konseil unter dem Herzog v. York zum erstenmal versammelt worden. Ueber des Königs Gesundheit ist seitdem ein Bülletin erschienen, das keine Veränderung anzeigt.

Man findet es sehr tadelhaft, daß bei der Westminsterwahl 8000 Wahlberechtigte gar nicht gestimmt haben. Ein Herr Lopez ist wegen Wahlbestechung unfähig erklärt worden, seinen Sitz im Parlament einzunehmen. Auch andre Mitglieder haben wegen erwiesener Ungültigkeit des Wahl-Verfahrens ausscheiden müssen.

Die Bittschriften zu Gunsten der Irländischen Katholiken von Seiten der Protestanten vermehren sich. Die Katholiken zu Dublin haben in einer unter Lord Fingals gehaltnen Versammlung, eine Danksagung an ihre protestantischen Brüder für die Unterstützung ihrer Petition einmüthig beschlossen.

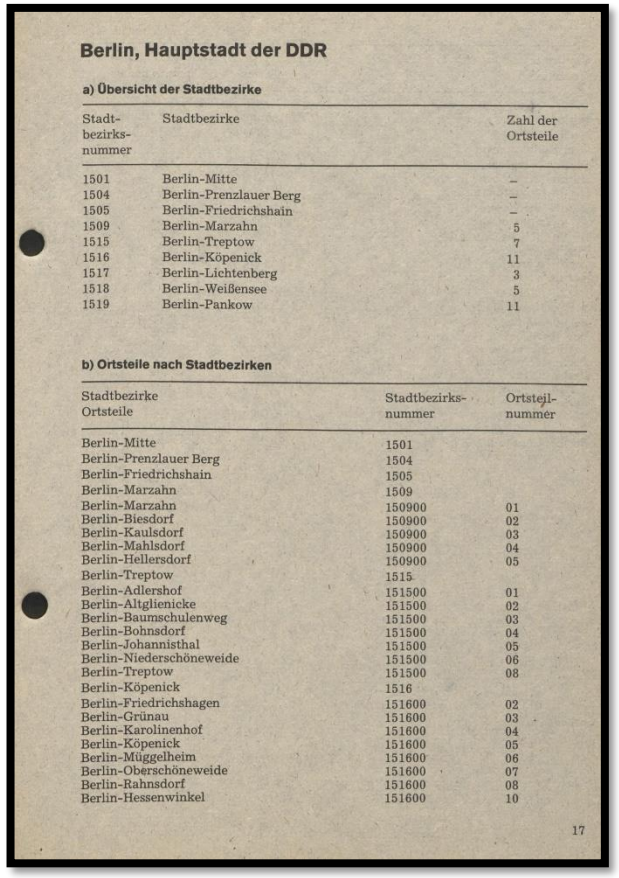
Paris, vom 10. März. Die beiden im Moniteur anfangs nicht nahmhaft gemachten Pairs sind die Grafen v. Montesquiou und v. Villegontier. Mit dem Grafen v. Greffülhe sind hiernach 60 neue Pairs ernannt. Schon 39 der neuen Pairs haben ihre Titel in gesetzlicher Art beglaubigt, und werden hienach an der nächsten Sitzung der Kammer Theil nehmen. Das Gesetz in Betreff des Salpeters ist von der Kammer der Pairs angenommen worden.

Der General-Ausschuß der Deputirten-Kammer, behufs der vorläufigen Prüfung des Gesetzes über die Tabaks-Regie, hat sich mit 5 gegen 4 Stimmen wider das Monopol des Staats erklärt.

In der Sitzung der Pairs vom 6. d. M. setzte der Graf von Castellane seinen Antrag in Bezug auf die Zurücknahme des Gesetzes vom 9. November 1815 wegen Aufruhr-Geschrei's auseinander. Der Minister des Innern bestritt ihn, es ward aber beschlossen, daß der Antrag in nähere Erwägung gezogen werden solle.

München, vom 11. März. Der Antrag des Hofraths Behr wegen der Censur ist dahin gerichtet, daß, da die politischen Zeitungen und periodischen Schriften noch einer Censur unterworfen sind, den Censoren eine Instruktion ertheilt werde, welche sie auf

Reichsanzeiger



Berlin, Hauptstadt der DDR

a) Übersicht der Stadtbezirke

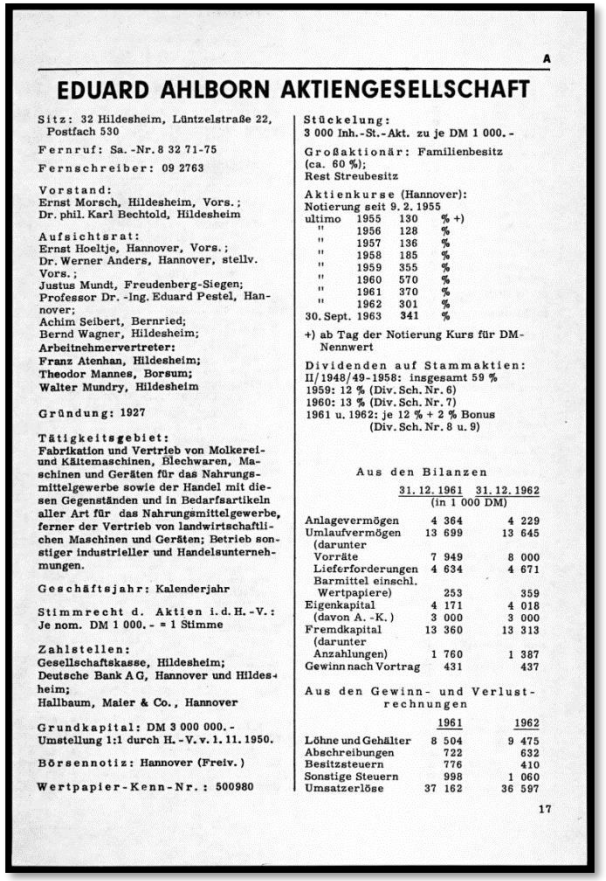
Stadtbezirks-nummer	Stadtbezirke	Zahl der Ortsteile
1501	Berlin-Mitte	–
1504	Berlin-Prenzlauer Berg	–
1505	Berlin-Friedrichshain	–
1509	Berlin-Marzahn	5
1515	Berlin-Treptow	7
1516	Berlin-Köpenick	11
1517	Berlin-Lichtenberg	3
1518	Berlin-Weißensee	5
1519	Berlin-Pankow	11

b) Ortsteile nach Stadtbezirken

Stadtbezirke Ortsteile	Stadtbezirks-nummer	Ortsteil-nummer
Berlin-Mitte	1501	
Berlin-Prenzlauer Berg	1504	
Berlin-Friedrichshain	1505	
Berlin-Marzahn	1509	
Berlin-Marzahn	150900	01
Berlin-Biesdorf	150900	02
Berlin-Kaulsdorf	150900	03
Berlin-Mahlsdorf	150900	04
Berlin-Hellersdorf	150900	05
Berlin-Treptow	1515	
Berlin-Adlershof	151500	01
Berlin-Altglienicke	151500	02
Berlin-Baumschulenweg	151500	03
Berlin-Bohnsdorf	151500	04
Berlin-Johannisthal	151500	05
Berlin-Niederschöneweide	151500	06
Berlin-Treptow	151500	08
Berlin-Köpenick	1516	
Berlin-Friedrichshagen	151600	02
Berlin-Grünau	151600	03
Berlin-Karolinenhof	151600	04
Berlin-Köpenick	151600	05
Berlin-Müggelheim	151600	06
Berlin-Oberschöneweide	151600	07
Berlin-Rahnsdorf	151600	08
Berlin-Hessenwinkel	151600	10

17

Gemeindeverzeichnisse



A

EDUARD AHLBORN AKTIENGESELLSCHAFT

Sitz: 32 Hildesheim, Lüntzelstraße 22, Postfach 530

Fernruf: Sa.-Nr. 8 32 71-75

Fernschreiber: 09 2763

Vorstand:
Ernst Morsch, Hildesheim, Vors.;
Dr. phil. Karl Bechtold, Hildesheim

Aufsichtsrat:
Ernst Hoeltje, Hannover, Vors.;
Dr. Werner Anders, Hannover, stellv. Vors.;
Justus Mundt, Freudenberg-Siegen;
Professor Dr.-Ing. Eduard Pestel, Hannover;
Achim Seibert, Bernried;
Bernd Wagner, Hildesheim;
Arbeitnehmervertreter:
Franz Atenhan, Hildesheim;
Theodor Mannes, Borsum;
Walter Mundry, Hildesheim

Gründung: 1927

Tätigkeitsgebiet:
Fabrikation und Vertrieb von Molkerei- und Kältemaschinen, Blechwaren, Maschinen und Geräten für das Nahrungsmittelgewerbe sowie der Handel mit diesen Gegenständen und in Bedarfsartikeln aller Art für das Nahrungsmittelgewerbe, ferner der Vertrieb von landwirtschaftlichen Maschinen und Geräten; Betrieb sonstiger industrieller und Handelsunternehmungen.

Geschäftsjahr: Kalenderjahr

Stimmrecht d. Aktien i. d. H.-V.:
Je nom. DM 1 000.- = 1 Stimme

Zahlstellen:
Gesellschaftskasse, Hildesheim;
Deutsche Bank AG, Hannover und Hildesheim;
Hallbaum, Maier & Co., Hannover

Grundkapital: DM 3 000 000.-
Umstellung 1:1 durch H.-V. v. 1. 11. 1950.

Börsennotiz: Hannover (Freiv.)

Wertpapier-Kenn-Nr.: 500980

Stückelung:
3 000 Inh.-St.-Akt. zu je DM 1 000.-

Großaktionär: Familienbesitz (ca. 60 %);
Rest Streubesitz

Aktienkurse (Hannover):
Notierung seit 9. 2. 1955

ultimo	1955	130	% +)
"	1956	128	%
"	1957	136	%
"	1958	185	%
"	1959	355	%
"	1960	570	%
"	1961	370	%
"	1962	301	%
30. Sept.	1963	341	%

+) ab Tag der Notierung Kurs für DM-Nennwert

Dividenden auf Stammaktien:
II/1948/49-1958: insgesamt 59 %
1959: 12 % (Div. Sch. Nr. 6)
1960: 13 % (Div. Sch. Nr. 7)
1961 u. 1962: je 12 % + 2 % Bonus (Div. Sch. Nr. 8 u. 9)

Aus den Bilanzen

	31. 12. 1961 (in 1 000 DM)	31. 12. 1962 (in 1 000 DM)
Anlagevermögen	4 364	4 229
Umlaufvermögen	13 699	13 645
(darunter Vorräte	7 949	8 000
Lieferforderungen	4 634	4 671
Barmittel einschl. Wertpapiere)	253	359
Eigenkapital	4 171	4 018
(davon A.-K.)	3 000	3 000
Fremdkapital	13 360	13 313
(darunter Anzahlungen)	1 760	1 387
Gewinn nach Vortrag	431	437

Aus den Gewinn- und Verlustrechnungen

	1961	1962
Löhne und Gehälter	8 504	9 475
Abschreibungen	722	632
Besitzsteuern	776	410
Sonstige Steuern	998	1 060
Umsatzerlöse	37 162	36 597

17

Aktienführer

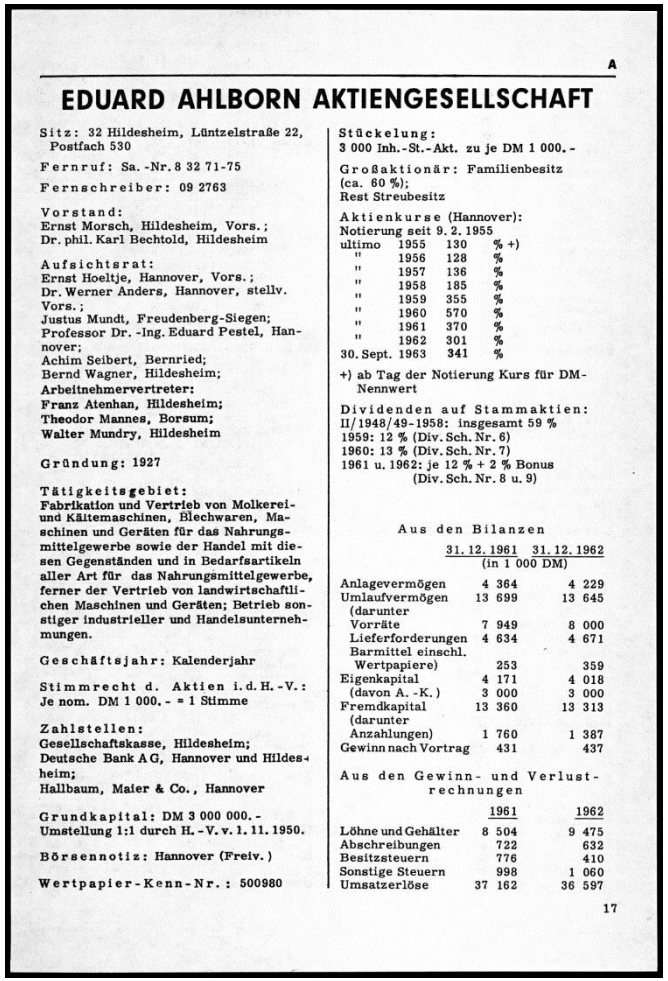
Aktienführer-Datenarchiv-Projekt

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

- Was ist der **Aktienführer**?
  - Jährlich erscheinende Publikation (Bücher bzw. CDs)
  - Firmenprofile aller an deutschen Börsen notierten Aktiengesellschaften
    - Wer war im Vorstand, Aufsichtsrat?
    - Aktienkurse, Bilanzdaten
- Was ist das **Ergebnis des Projekts**?
  - Zugang zu den **Unternehmensprofilen** der **letzten 140 Jahre** (1870–2018)
  - Bereitstellung einer **feinstrukturierten Datenbank** der **letzten 60 Jahre** (1956–2018)
- Was ist der **Rahmen des Projekts**?

DFG Deutsche Forschungsgemeinschaft

  - 1. Projektphase: 2013–2016 (24 M)
  - 2. Projektphase: 2017–2019 (24 M)



A

EDUARD AHLBORN AKTIENGESELLSCHAFT

Sitz: 32 Hildesheim, Lüntzelstraße 22, Postfach 530

Fernruf: Sa.-Nr. 8 32 71-75

Fernschreiber: 09 2763

Vorstand:
Ernst Morsch, Hildesheim, Vors.;
Dr. phil. Karl Bechtold, Hildesheim

Aufsichtsrat:
Ernst Hoeltje, Hannover, Vors.;
Dr. Werner Anders, Hannover, stellv. Vors.;
Justus Mundt, Freudenberg-Siegen;
Professor Dr.-Ing. Eduard Pestel, Hannover;
Achim Seibert, Bernried;
Bernd Wagner, Hildesheim;
Arbeitnehmervertreter:
Franz Atenhan, Hildesheim;
Theodor Mannes, Borsum;
Walter Mundry, Hildesheim

Gründung: 1927

Tätigkeitsgebiet:
Fabrikation und Vertrieb von Molkerei- und Kältemaschinen, Blechwaren, Maschinen und Geräten für das Nahrungsmittelgewerbe sowie der Handel mit diesen Gegenständen und in Bedarfsartikeln aller Art für das Nahrungsmittelgewerbe, ferner der Vertrieb von landwirtschaftlichen Maschinen und Geräten; Betrieb sonstiger industrieller und Handelsunternehmungen.

Geschäftsjahr: Kalenderjahr

Stimmrecht d. Aktien i.d.H.-V.:
Je nom. DM 1 000.- = 1 Stimme

Zahlstellen:
Gesellschaftskasse, Hildesheim;
Deutsche Bank AG, Hannover und Hildesheim;
Hallbaum, Maier & Co., Hannover

Grundkapital: DM 3 000 000.-
Umstellung 1:1 durch H.-V. v. 1. 11. 1950.

Börsennotiz: Hannover (Freiv.)

Wertpapier-Kenn-Nr.: 500980

Stückelung:
3 000 Inh.-St.-Akt. zu je DM 1 000.-

Großaktionär: Familienbesitz (ca. 60 %);
Rest Streubesitz

Aktienkurse (Hannover):
Notierung seit 9. 2. 1955

ultimo	1955	130	% +)
"	1956	128	%
"	1957	136	%
"	1958	185	%
"	1959	355	%
"	1960	570	%
"	1961	370	%
"	1962	301	%
30. Sept.	1963	341	%

+) ab Tag der Notierung Kurs für DM-Nennwert

Dividenden auf Stammaktien:
II/1948/49-1958: insgesamt 59 %
1959: 12 % (Div. Sch. Nr. 6)
1960: 13 % (Div. Sch. Nr. 7)
1961 u. 1962: je 12 % + 2 % Bonus (Div. Sch. Nr. 8 u. 9)

Aus den Bilanzen

	31. 12. 1961	31. 12. 1962
	(in 1 000 DM)	
Anlagevermögen	4 364	4 229
Umlaufvermögen	13 699	13 645
(darunter Vorräte	7 949	8 000
Lieferforderungen	4 634	4 671
Barmittel einschl. Wertpapiere)	253	359
Eigenkapital	4 171	4 018
(davon A.-K.)	3 000	3 000
Fremdkapital	13 360	13 313
(darunter Anzahlungen)	1 760	1 387
Gewinn nach Vortrag	431	437

Aus den Gewinn- und Verlustrechnungen

	1961	1962
Löhne und Gehälter	8 504	9 475
Abschreibungen	722	632
Besitzsteuern	776	410
Sonstige Steuern	998	1 060
Umsatzerlöse	37 162	36 597

17

Firmenprofil aus dem Aktienführer 1964 (**Buch**)

Prozesskette
Forschungdaten aus Digitalisaten

Bücher

Digitalisierung/ Vorverarbeitung

Generierung digitaler Ausgangsformate

Text- erkennung (OCR)

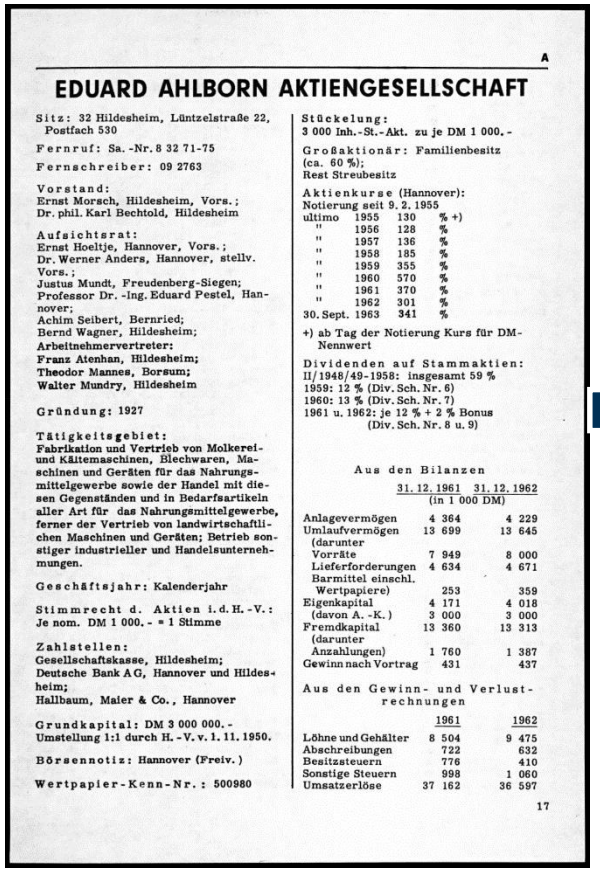
Generierung der digitalen Inhalte

Struktur- parsing

Strukturierung der digitalen Inhalte (Datenextraktion)

Befüllen der Datenbank

crass – Vorverarbeitung des Digitalisats



A

EDUARD AHLBORN AKTIENGESELLSCHAFT

Sitz: 32 Hildesheim, Lüntzelstraße 22, Postfach 530

Fernruf: Sa.-Nr. 8 32 71-75

Fernschreiber: 09 2763

Vorstand:
Ernst Morsch, Hildesheim, Vors.;
Dr. phil. Karl Bechtold, Hildesheim

Aufsichtsrat:
Ernst Hoeltje, Hannover, Vors.;
Dr. Werner Anders, Hannover, stellv. Vors.;
Justus Mundt, Freudenberg-Siegen;
Professor Dr.-Ing. Eduard Pestel, Hannover;
Achim Seibert, Bernried;
Bernd Wagner, Hildesheim;
Arbeitnehmervertreter:
Franz Atenhan, Hildesheim;
Theodor Mannes, Borsum;
Walter Mundry, Hildesheim

Gründung: 1927

Tätigkeitsgebiet:
Fabrikation und Vertrieb von Molkerei- und Kältemaschinen, Blechwaren, Maschinen und Geräten für das Nahrungsmittelgewerbe sowie der Handel mit diesen Gegenständen und in Bedarfsartikeln aller Art für das Nahrungsmittelgewerbe, ferner der Vertrieb von landwirtschaftlichen Maschinen und Geräten; Betrieb sonstiger industrieller und Handelsunternehmungen.

Geschäftsjahr: Kalenderjahr

Stimmrecht d. Aktien i. d. H.-V.:
Je nom. DM 1 000.- = 1 Stimme

Zahlstellen:
Gesellschaftskasse, Hildesheim;
Deutsche Bank AG, Hannover und Hildesheim;
Hallbaum, Maier & Co., Hannover

Grundkapital: DM 3 000 000.-
Umstellung 1:1 durch H.-V. v. 1. 11. 1950.

Börsennotiz: Hannover (Freiv.)

Wertpapier-Kenn-Nr.: 500980

Stückelung:
3 000 Inh.-St.-Akt. zu je DM 1 000.-

Großaktionär: Familienbesitz (ca. 60 %);
Rest Streubesitz

Aktienkurse (Hannover):
Notierung seit 9. 2. 1955

ultimo	1955	130	% +)
"	1956	128	%
"	1957	136	%
"	1958	185	%
"	1959	355	%
"	1960	570	%
"	1961	370	%
"	1962	301	%
30. Sept.	1963	341	%

+) ab Tag der Notierung Kurs für DM-Nennwert

Dividenden auf Stammaktien:
II/1948/49-1958: insgesamt 59 %
1959: 12 % (Div. Sch. Nr. 6)
1960: 13 % (Div. Sch. Nr. 7)
1961 u. 1962: je 12 % + 2 % Bonus (Div. Sch. Nr. 8 u. 9)

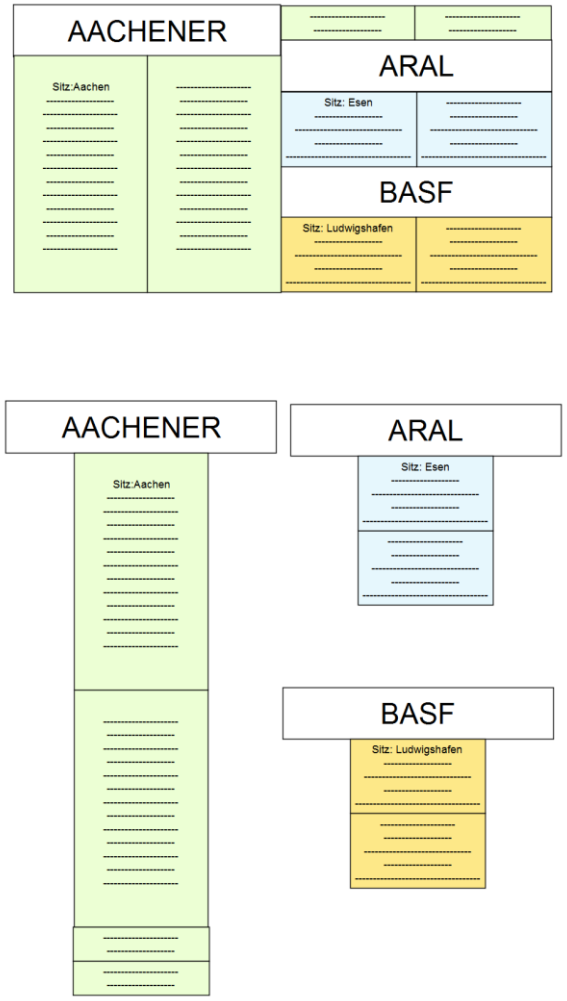
Aus den Bilanzen

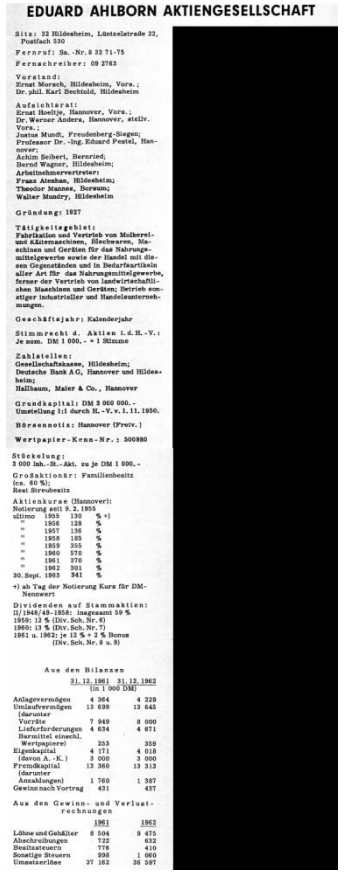
	31. 12. 1961	31. 12. 1962
	(in 1 000 DM)	
Anlagevermögen	4 364	4 229
Umlaufvermögen	13 699	13 645
(darunter Vorräte	7 949	8 000
Lieferforderungen	4 634	4 671
Barmittel einschl. Wertpapiere)	253	359
Eigenkapital	4 171	4 018
(davon A.-K.)	3 000	3 000
Fremdkapital	13 360	13 313
(darunter Anzahlungen)	1 760	1 387
Gewinn nach Vortrag	431	437

Aus den Gewinn- und Verlustrechnungen

	1961	1962
Löhne und Gehälter	8 504	9 475
Abschreibungen	722	632
Besitzsteuern	776	410
Sonstige Steuern	998	1 060
Umsatzerlöse	37 162	36 597

17





OCR Software (Übersicht)

kommerzielle Software

fett = eingesetzt in Bibliotheken

ABBYY Finereader
BIT-Alpha
Readiris
OmniPage

Adobe Acrobat
CorelDraw
Microsoft OneNote
…

Tesseract
OCRopus / Kraken / Calamari
CuneiForm
…

freie Software

ABBYY Cloud OCR
Google Cloud Vision
Microsoft Azure Computer Vision
OCR.space Online OCR …

Cloud OCR

Tesseract OCR

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

- Open Source
- Komplettlösung „All-in-1“
- Mehr als 100 Sprachen / mehr als 30 Schriften
- Liest Bilder in allen gängigen Formaten (nicht PDF!)
- Erzeugt Text, PDF, hOCR, ALTO, TSV
- Große, weltweite Anwender-Community
- Technologisch aktuell (Texterkennung mit neuronalem Netz)
- Aktive Weiterentwicklung u. a. im DFG-Projekt OCR-D

Tesseract an der UB Mannheim

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

- Verwendung im DFG-Projekt „Aktienführer“ https://digi.bib.uni-mannheim.de/aktienfuehrer/
- Volltexte für Deutscher Reichsanzeiger und Vorgänger https://digi.bib.uni-mannheim.de/periodika/reichsanzeiger

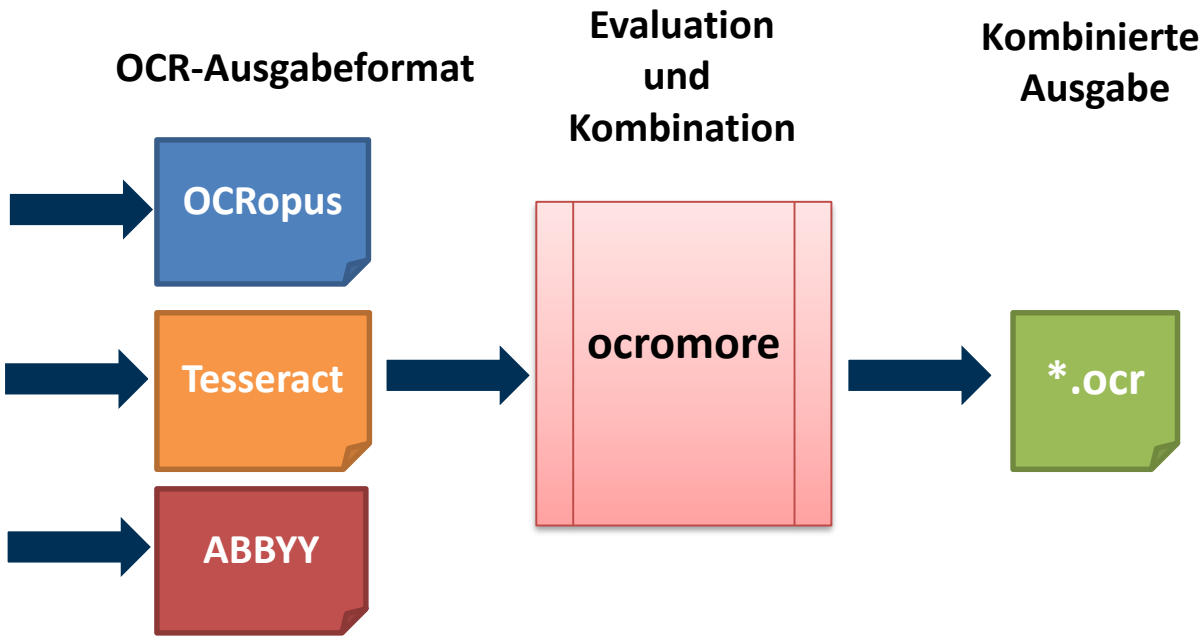
- DFG-Projekt „OCR-D“ http://www.ocr-d.de/, Modulprojekt „Optimierter Einsatz von OCR-Verfahren – Tesseract als Komponente im OCR-D-Workflow“: Schnittstellen, Stabilität, Performance und praktische Einsetzbarkeit, Erweiterungen wie z. B. Konfidenzen

OCR-D
Koordinierte Förderinitiative zur Weiterentwicklung von Verfahren der Optical Character Recognition (OCR)

OCROMORE

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

ocromore – Kombination und Optimierung von OCR-Resultaten




```
R1: Eduard Ahlborn Axtiengesellschaft
R2: Edoard Anlborn Aktiengesellschaft
R3: Eduard Ahlbrn  Aktiengesellschaft
```

Input: Mehrere OCR-Resultate

Edu¦ard	A¦hlborn	A¦xtiengesellschaft
Ed¦oard	An¦lborn	Ak¦tiengesellschaft
Edu¦ard	A¦hlb¦rn	Ak¦tiengesellschaft

Aufteilung und Vergleich auf Wortebene mit Textausrichtung (Alignment)

u¦	99; 00;
¦o	00; 60;
u¦	90; 00;

Wahl der Zeichen nach Konfidenz der Verfahren

Eduard	Ahlborn	Aktiengesellschaft

OCROMORE

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

ocromore – Verbesserung der Zeichengenauigkeit

OCR-Engine	Aktienführer (AKF)	UNLV
ABBYY	99,35 %	98,46 %
OCRopus (default en-model)	-	92,49 %
OCRopus (trained)	98,76 %	-
Tesseract	99,00 %	98,23 %
ocromore (MSA)	**99,60 %**	**98,65 %**

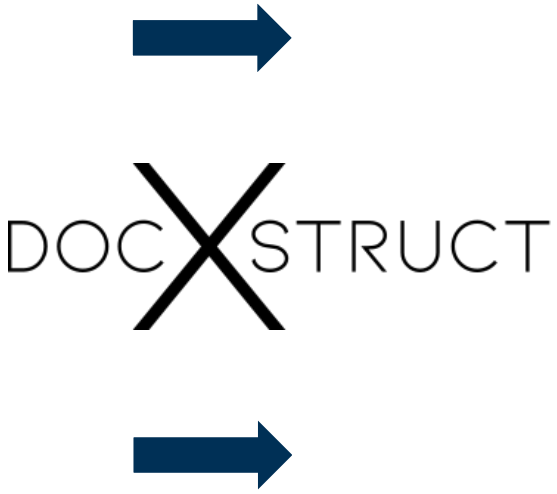
Erhöhung der Zeichengenauigkeit (AKF) : 0,25 %
Erhöhung der Wortgenauigkeit (AKF) : 2,03 %
Erhöhung der Zeichengenauigkeit (UNLV) : 0,19 %
Fehlerreduktion der Zeichengenauigkeit (AKF) : 38,5 %

AKF: 18 große Dateien (mit insgesamt ca. 100.000 Zeichen) 1957-1976 alle 3-4 Jahre
UNLV: University of Nevada Las Vegas standardized test set

docxstruct – Segmentklassifizierung und Datenextraktion

```
EDUARD AHLBORN AKTIENGESELLSCHAFT

Sitz: 32 Hildesheiim, Lüntzelstraße 22,
Postfach 530
Fernruf: Sa.-Nr.8 32 71-75
Fernschreiber: 09 2763
Vorstand:
Ernst Morsch, Hildesheim, Vors.;
Dr. phil. Karl Bechtold, Hildesheim
Aufsichtsrat:
Ernst Hoeltje, Hannover, Vors.;
Dr. Werner Anders, Hannover, stellv.
Vors.;
Justus Mundt, Freudenberg-Siegen;
Professor Dr. -Ing. Eduard Pestel, Han-
nover;
Achim Seibert, Bernried;
Bernd Wagner, Hildesheim;
Arbeitnehmervertreter:
Franz Atenhan, Hildesheim;
Theodor Mannes, Borsum;
Walter Mundry, Hildesheim
Gründung: 1927
```



```
"Sitz": [
     {
          "numID": "32",
          "city": "Hildesheiim",
          "street": "Lüntzelstraße 22",
          "additional_info": [
               "Postfach 530"
          ]
     }
],
"Vorstand": [
     {
          "name": "Ernst Morsch",
          "first_name": "Ernst",
          "last_name": "Morsch",
          "city": "Hildesheim",
          "funct": "Vors."
     },
     {
          "name": "Karl Bechtold",
          "first_name": "Karl",
          "last_name": "Bechtold",
          "city": "Hildesheim",
          "title": "Dr.phil."
     }
],
```

Segmentierung des Aktienführers

Strukturerkennung von Sitz und Vorstand

Ein Blick **zurück** und nach **vorne**

Derzeitiger Stand:

- Alle Tools als **Open Source** öffentlich auf GitHub:
  https://github.com/UB-Mannheim/Aktienfuehrer-Datenarchiv-Tools
  https://github.com/tesseract-ocr/tesseract

Ausblick:

- Ausbau des **Forschungsdatenzentrums** (FDZ) der UB Mannheim; Aktienführer als Kernstück
- **BERD-Center** (Business and Economic Research Data Center)
  - Kooperationsprojekt Universität Mannheim und ZEW, Landesförderung BW
  - Aufbau eines Kompetenzzentrums für Forschungsdaten der Wirtschaftswissenschaften
  - Möglichkeit zur Nachnutzung der entwickelten Softwaretools aus Aktienführer-Datenarchiv-Projekt
- Kooperationsprojekt **OCR-BW** (gemeinsam mit Tübingen, Landesprojekt BW, bewilligt)
  - Aufbau eines Kompetenzzentrums Volltexterkennung für Bibliotheken und Archive
- DFG-Projekt **Reichsanzeiger** (bewilligt)

Literatur

- Weil, S., & Zumstein, P. (2016). Mit freier Software Text in Digitalisaten erkennen. https://speakerdeck.com/zuphilip/mit-freier-software-text-in-digitalisaten-erkennen-ocr-praxis-an-der-ub-mannheim
- Baierer, K., & Zumstein, P. (2016). Verbesserung der OCR in digitalen Sammlungen von Bibliotheken. *027.7 Zeitschrift für Bibliothekskultur / Journal for Library Culture, 4*(2), 72-83. https://doi.org/10.12685/027.7-4-2-155
- Kamlah, J., Stegmüller, J. (2018). Ocromore – Combining multiple OCR-engine results to improve character recognition accuracy. https://zenodo.org/record/1493860
- Kamlah, J., Stegmüller, J., Schumm, I., Zumstein, P. (2019). Automatisierte Optimierung und Strukturierung von OCR-Ergebnissen mit nachnutzbaren Werkzeugen. https://ub-madoc.bib.uni-mannheim.de/48940
- Weil, S. (2019). Vom Bild zum Text. Automatisierte Texterkennung in historischen Drucken mit der freien Software Tesseract. https://nbn-resolving.org/urn:nbn:de:0290-opus4-163511

Bildquellen

- Titelseite:
  https://pixabay.com/de/vectors/flach-design-symbol-icon-www-2126884/
  https://pixabay.com/de/vectors/flach-design-symbol-icon-www-2126880/
  https://pixabay.com/de/vectors/werkzeug-schraubenschl%C3%BCssel-3456474/
  https://commons.wikimedia.org/wiki/File:Opensource.svg
- DFG-Logo: https://www.dfg.de/
- GitHub Logos: https://github.com/logos
- OCR-D Logo: http://www.ocr-d.de/