Beschreibung des Match&Merge-Verfahrens

Im Match&Merge-Verfahren werden die gelieferten Datensätze an den vorhandenen GND-Datensätzen abgeglichen. Das Verfahren findet im OCLC/PICA-CBS statt, dem Katalogisierungssystem der GND. Hierzu werden die gelieferten Datensätze in das interne PICA-Datenformat konvertiert. Über Namens-Indizes werden dann für jeden Datensatz  passende GND-Datensätze vorselektiert und über den Abgleich verschiedener Elemente daraus derjenige mit dem höchsten Matchwert ermittelt.

Vergleichselemente sind:

* Name (MARC 100, 400)
* Geburts-, Sterbejahr (MARC 548 mit $4 “datl”)
* Beruf (MARC 550 mit $4 “berc” oder “beru”)
* Geburts-, Sterbeort (MARC 551 mit $4 “ortg” oder “orts”)

Voraussetzung für einen Match ist, dass der Vergleich des Namenselements einen hohen Matchwert erhält und zudem das Geburts- oder Sterbejahr identisch ist sowie zusätzlich der Beruf oder der Geburts-/Sterbeort oder die jeweils andere Jahresangabe. Dies bedeutet, dass sowohl im gelieferten Datensatz wie auch im GND-Datensatz zumindest ein Lebensdatum zwingend vorhanden sein müssen, damit es einen Match geben kann.

Analyse einzelner Bestände

Daten des Landesarchivs Baden-Württemberg

labw\_1\_7168.xml

Beschreibung

Bestände des LABW, die in den Titeln (unittitle) Personennamen und in der Zusammenfassung (abstract) Lebensdaten enthalten.

Beispiele

<unitid>J 191 Queisser, Hans-Joachim Prof.</unitid>

<abstract>Direktor am Max-Planck-Institut Stuttgart / Halbleiterphysiker<lb/>\* 06.07.1931 Berlin</abstract>

<unittitle>Xander, Karl Prof.</unittitle>

<abstract>Rektor FH.-Ulm<lb/>\* 18.09.1924 Ulm</abstract>

<unittitle>Zywietz, Otto</unittitle>

<abstract>Kommunalpolitiker<lb/>\* 14.01.1924<lb/>+ 22.03.1993 Kirchzarten</abstract>

Herausforderung

* Ansetzung der Titel/des Abstracts unklar
* Ansetzung des unittitle und abstract variiert
* Datenqualität generell unklar

Bewertung

* sehr gut für geeignet, insofern die Ansetzung homogen ist
* detaillierte Lebensdaten sind im Abstract vorhanden

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | LABW |
| Datei | labw\_1\_7168.xml |
| Anzahl der Entitäten | 17.235 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 4.029 (23,4%) |
| Davon individuelle Personen | 3.996 |

labw\_2\_2940185.xml

Beschreibung

Verzeichnis von Grabsteinen, in der Zusammenfassung sind teilweise die Inschriften festgehalten.

Beispiele

<abstract>Inschrift (1):<lb/>Landauer, Lina | --<lb/>Familienstand: ledig<lb/>Sprache: deutsch</abstract>

<abstract>Inschrift (1):<lb/>Netter, Albert | --<lb/>Beruf: Frontsoldat<lb/>Familienstand: unbekannt<lb/>Geb. 23.02.1891 | --<lb/>Gest. 26.11.1918 | -- (Dienstag)<lb/>Sprache: deutsch<lb/>Inschrift (2):<lb/>Netter, Therese | --, geb. Loeser<lb/>Familienstand: verheiratet<lb/>Geb. 20.07.1850 | --<lb/>Gest. 04.08.1904 | -- (Donnerstag)<lb/>Sprache: deutsch<lb/>Inschrift (3):<lb/>Netter, Nathan | --<lb/>Familienstand: unbekannt<lb/>Geb. 03.08.1846 | --<lb/>Gest. 13.12.1918 | -- (Freitag)<lb/>Sprache: deutsch</abstract>

<abstract>Inschrift (1):<lb/>Tänzer, Aron | --<lb/>Gatte: Rosa<lb/>Amt: Rabbiner (Feldrabbiner)<lb/>Familienstand: verwitwet<lb/>Geb. 30.01.1871 | --<lb/>Gest. 26.02.1937 | 15. Adar 5697 (Freitag)<lb/>Sprache: hebräisch und deutsch<lb/>Inschrift (2):<lb/>Tänzer, Berta | --<lb/>Familienstand: unbekannt<lb/>Geb. 1876 | --<lb/>Gest. 1943 | --, in: Lager Theresienstadt<lb/>Sprache: hebräisch und deutsch<lb/>Inschrift (3):<lb/>Tänzer, Rosa | --<lb/>Gatte: Aron / Arnold<lb/>Familienstand: verheiratet<lb/>Geb. 04.05.1875 | --<lb/>Gest. 12.09.1912 | 28. Elul 5672 (Donnerstag)<lb/>Sprache: hebräisch und deutsch</abstract>

Herausforderung

* die Namen von Personen sind als Freitext in der Zusammenfassung angegeben
* es gibt keine einheitliche Struktur und Ansetzung
* Text Mining-Verfahren sind notwendig, um zunächst den Aufbau zu erkennen, Personennamen und –daten zu extrahieren, erst dann können Match-Verfahren gegen die GND angewendet werden

Bewertung

* grundsätzlich gut für Normdatenanreicherung geeignet, allerding ist eine Aufbereitung der Daten (ggf. sogar Text Mining-Verfahren) notwendig

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | LABW |
| Datei | labw\_2\_2940185.xml |

Deskriptoren

Beschreibung

Tabellarische Auflistung von Personen, die in den Daten des Landesarchivs Baden-Württemberg verwendet werden. Dazu wird ein interne interne ID vergeben, die in den einzelnen Beständen wiederverwendet werden.

Beispiele

"5112";"6";"54248";"Ferdinand I.; König, Rumänien, 1865-1927 | 119346397"

"5121";"6";"54249";"Lutz, M."

"5138";"6";"54255";"Fürstenberg, Fürsten von"

"5139";"6";"54256";"Thurn und Taxis, Fürsten von"

"5143";"6";"54257";"Mayer, Samuel, Dr., Rabbiner, Hechingen"

"5152";"6";"54258";"Spiegel-Borlinghausen, von, Adolf Karl, Preußischer Kommissarius"

Herausforderung

* die personenbezogenen Daten sind in einem Feld gespeichert
* Ansetzung unklar, aber vermutlich einheitlich

Bewertung

* grundsätzlich gut für eine Normdatenanreicherung geeignet

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | LABW-Deskriptoren |
| Datei | deskriptoren\_hstas\_personen.csv |
| Anzahl der Entitäten | 46.994 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 558 (1,2%) |
| Davon individuelle Personen | 506 |

|  |  |
| --- | --- |
| Einrichtung | LABW-Deskriptoren |
| Datei | deskriptoren\_stas\_personen.csv |
| Anzahl der Entitäten | 35.156 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 49 (0,1%) |
| Davon individuelle Personen | 47 |

Hessisches Staatsarchiv

ead\_hstad\_r\_4\_ohne34239.xml

Beschreibung

Bestände des Hessischen Staatsarchivs, die in den Titeln (unittitle) Personennamen und einfache Lebensdaten enthalten.

Beispiele

<ead:unittitle>Albrecht, Elisabeth geb. Merck (1890-1970) / Porträt, mit Perlenkette und aufgeschlagenem Buch gegen Ölbaumhain / linksvorblickendes, sitzendes Kniestück</ead:unittitle>

<ead:unittitle>Alice Großherzogin v. Hessen und bei Rhein geb. Prinzessin v. Großbritannien (1843-1878) / Porträt, aufgebahrt auf dem Totenbett, Brustbild</ead:unittitle>

<ead:unittitle>Weber, Alex / Porträt, Brustbild</ead:unittitle>

Herausforderung

* Ansetzung der Titel unklar
* Datenqualität unklar
* große Dateigröße
* detaillierte Personendaten nicht vorhanden
  + ggf. sind weitere Informationen zu den Personen vorhanden? Wenn ja, wo?

Bewertung

* gut für Normdatenanreicherung geeignet, insofern die Ansetzung homogen ist, es fehlen leider detaillierte Lebensdaten (exaktes Datum, Orte)

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | Hessische Staatsarchive |
| Datei | ead\_hstad\_r\_4\_ohne34239.xml |
| Anzahl der Entitäten | 36.195 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 3.619 (~10%) |
| Davon individuelle Personen | 1.691 |

Hessisches Landesarchiv

HHStAW\_469\_33\_2015\_11\_16\_10\_53\_58.xml u.a.

Beschreibung

Zivilprozessakten und Urteile, Strafprozesse und Urteile mit detaillierten Angaben zu den betreffenden Personen.

Beispiele

<verzeichnung bezeichnung="Blesinger, Therese geb. Travers" laufzeitbis="1920" laufzeitvon="1908" beschreibungsmodellid="12" sortiermerkmal="" isinverzeichnung="false" status="10" verborgenbis="0" signatur="7989" id="2449319">

<beschreibung beschreibungselementid="57[MB1] " beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Frankfurt am Main-Höchst</beschreibung>

<beschreibung beschreibungselementid="59[MB2] " beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Wiesbaden</beschreibung>

<beschreibung beschreibungselementid="63[MB3] " beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Wiesbaden</beschreibung>

<beschreibung beschreibungselementid="52[MB4] " beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Blesinger, Therese geb. Travers</beschreibung>

   <beschreibung beschreibungselementid="58[MB5] " beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">1916-04-14</beschreibung>

</verzeichnung>

<verzeichnung bezeichnung="Hoppe, Peter Daniel" laufzeitbis="1922" laufzeitvon="1919" beschreibungsmodellid="12" sortiermerkmal="" isinverzeichnung="false" status="10" verborgenbis="0" signatur="122" id="3049469">

<beschreibung beschreibungselementid="57" beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Solingen-Dorp</beschreibung>

<beschreibung beschreibungselementid="51" beschreibungsgruppeid="27" gruppenindex="0" elementindex="0">Namensänderung</beschreibung>

<beschreibung beschreibungselementid="56[MB6] " beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">1888-09-23</beschreibung>

<beschreibung beschreibungselementid="66" beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Wunsch zur Annahme des Familiennamens Hoppe-Boeken</beschreibung>

<beschreibung beschreibungselementid="52" beschreibungsgruppeid="28" gruppenindex="0" elementindex="0">Hoppe, Peter Daniel</beschreibung>

<beschreibung beschreibungselementid="2" beschreibungsgruppeid="27" gruppenindex="0" elementindex="0">1919-1922</beschreibung>

</verzeichnung>

Herausforderung

* Datenqualität unklar
* große Dateigröße

Bewertung

* sehr gut für Normdatenanreicherung geeignet
* Daten sind sehr gut strukturiert und vermutlich einheitlich angesetzt

Hinweis

* die übrigen Dateien des Datengebers sind ähnlich aufgebaut und ähnlich gut geeignet
  + HHStAW\_469\_33\_2015\_11\_16\_10\_53\_58.xml (diese)
  + HHStAW\_469\_6\_2015\_11\_16\_10\_56\_35.xml
  + HStAD\_G\_16\_A\_2015\_11\_16\_10\_59\_20.xml
  + HStAD\_G\_21\_B\_2015\_11\_16\_10\_24\_02.xml
  + HStAD\_G\_31\_P\_2015\_11\_16\_10\_30\_08.xml
  + HStAD\_G\_35\_E\_2015\_11\_16\_11\_05\_51.xml
  + HStAD\_R\_12\_P\_2015\_11\_16\_10\_27\_31.xml
  + HStAM\_251\_Wehlheiden\_2015\_11\_16\_10\_21\_39.xml

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HHStAW\_469\_33\_2015\_11\_16\_10\_53\_58.xml |
| Anzahl der Entitäten | 5.895 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 16 (~0,3%) |
| Davon individuelle Personen | 16 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HHStAW\_469\_6\_2015\_11\_16\_10\_56\_35.xml |
| Anzahl der Entitäten | 9.224 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 67 (~0,7%) |
| Davon individuelle Personen | 66 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HStAD\_G\_16\_A\_2015\_11\_16\_10\_59\_20.xml |
| Anzahl der Entitäten | 4.266 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 3 (0,1%) |
| Davon individuelle Personen | 3 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HStAD\_G\_21\_B\_2015\_11\_16\_10\_24\_02.xml |
| Anzahl der Entitäten | 5.500 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 24 (0,4%) |
| Davon individuelle Personen | 23 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HStAD\_G\_31\_P\_2015\_11\_16\_10\_30\_08.xml |
| Anzahl der Entitäten | 12.180 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 31 (0,3%) |
| Davon individuelle Personen | 24 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HStAD\_G\_35\_E\_2015\_11\_16\_11\_05\_51.xml |
| Anzahl der Entitäten | 24.740 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 46 (0,2%) |
| Davon individuelle Personen | 46 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HStAD\_R\_12\_P\_2015\_11\_16\_10\_27\_31.xml |
| Anzahl der Entitäten | 8.022 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 762 (9,5%) |
| Davon individuelle Personen | 761 |

|  |  |
| --- | --- |
| Einrichtung | Hessisches Landesarchiv |
| Datei | HStAM\_251\_Wehlheiden\_2015\_11\_16\_10\_21\_39.xml |
| Anzahl der Entitäten | 59 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 0 |
| Davon individuelle Personen | 0 |

Max-Planck-Archiv

iii\_8\_10050\_10012\_ddbnormxml.xml

Beschreibung

Nachlass Max Rubner aus dem Archiv der Max-Planck-Gesellschaft. Bestände enthalten zum Teil ein Personenregister (insgesamt 713 Einträge), das oft nur aus Nachname und Vorname besteht. Die Titel der Bestände enthalten oft zusätzlich noch die Lebensdaten der Person.

Beispiele

<register art="Personen-Index">Thiele, A.</register>

<register art="Personen-Index">Thomas, Karl, Geb.: 28.09.1883</register>

  <item kind="file">

<col1>51</col1>

<title>Gruber, Max   (1882-1927)

Grützner, P.   (1909)

Gualdi, Tito (1903)

Günther, Carl (1900-1929)

Günther, D.   (1889-1879)

Gumm, F. (1927-1928)

Gurtner,     (1891)</title>

<col3>1882-1929</col3>

<register art="Personen-Index">Gruber, Max</register>

<register art="Personen-Index">Grützner, P.</register>

<register art="Personen-Index">Gualdi, Tito</register>

<register art="Personen-Index">Günther, Carl</register>

<register art="Personen-Index">Günther, D.</register>

<register art="Personen-Index">Gumm, F.</register>

<register art="Personen-Index">Gurtner</register>

</item>

Herausforderung

* spärliche Datenlage, oft nur Namen
* Lebensdaten müssen aus Titel der Akte extrahiert werden
* Ansetzung bei Registereintragen mit Lebensdaten ist unklar
* unklar ist, ob ggf. aus dem Kontext der Daten weitere Informationen gewonnen werden können; dies wäre zu klären

Bewertung

* schlecht für eine Anreicherung mit Normdaten geeignet

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | Max-Planck-Archiv |
| Datei | III\_8\_10050\_10012\_DDBNORMXML.xml |

III\_93\_10012\_DDBNORMXML.xml

Beschreibung

Nachlass Werner Heisenbergs des Archivs der Max-Planck-Gesellschaft. Die Daten enhalten viele Namen ohne weitere Daten als Freitext. Es gibt wieder ein Personenregister, allerdings nur mit 27 Einträgen.

Beispiele

<item kind="file">

<col1>1687</col1>

<title>1946-1947, A-G</title>

<text>Enthält: Abel, Kurt.- Ablakos, Stefan.- Achelis, [Johann Daniel].- Arfsten, Arvest Newton.- Artmann, Kurt.- Auersbach, K. - Auvers.-

Babinger, Franz.- Back, E[rnst Emil Alexander].- Volkshochschule Bad Pyrmont e.V.- Bäuerle, Hans.- Bahro, Hans.- Bank, Hans.- Barthel, Ernst.- Bartram, Theodor.- Barz, Walter Heinz.- Bauer, Georg F.- Bauer, Karl.- Bauer, Arnold.- Becker, [Richard].- Beauclair, Gotthard de.-  Beauclair, Wilfried de.- Bechert, Karl.- Benjamin Verlag GmbH.- Bein, Willy.- Bertuch&apos;s Original Unterrichtsbriefe.- Berve, Helmut.- Beulig, Gerhard.- Arbeitskreis für christliche Kultur und Lebensgestaltung Bielefeld.- Biermann, Ludwig.- Biophysikalische Konferenz (Göttingen1947).- Birkmeyer, Karl M.- Bliesener, Robert.- Blücher.- Bökemann, Karl.- Bonhoeffer, Karl Friedrich.- Allgemeiner Studentenausschuss der Universität Bonn.- Borchardt, Werner.- Borger, G.-  Borrmann, G.- Brandes, Werner.- Braun, Hanns.- Braunschweig, A.- Naturwissenschaftlicher Verein Bremen.- Brieger, Gotthard.- Brill, Hermann L.- Brix, Hermann.- Brockhaus, Hans.- Broog, Tim.- Brosi, Elisabeth.- Brüche, E.- Bücking, Helmuth.- Bund Naturschutz in Bayern e.V.- Burger, H.- Burghauser, Franz.- Bussche, Axel v. d.- Buttersack.- Butenandt, Adolf.-

Akademische Buchhandlung Calvör.- Castrop-Rauxel.- Clar, Anneliese.- Clara, Henny.- Claussen, Joachim.-

Dähling, Heinz.- Dahnke, Heinrich.- Dänzer, Hermann.- DANA-Bilderdienst.- Deutsche Beiträge (Nymphenburger Verlagshandlung).- Deutsches Museum.- Deutschmann, Martin.- Dieckmann, Johann Seeden.- Dieminger, W.- Doehring, Lic.- Dölger, Franz.- Dogigli, Giovannii.- Dolch, Heiner.- Domagk, Gerhard.- Doerfler, Fritz.- Bildungswerk Dortmund.- Droste, G. v.- Druckzersetzer GmbH.- Duhn, A. v.- Duhn, Johann Hermann. v.</text>

<text>Altsignatur: (1)</text>

<text> Ebbinghaus.- Eckert, Hans-Ulrich.- Elfremov, N. v.- Einfalt, H.- Elster, Hans Johann.- Engelmann, Werner.- Erler, Werner.- Ernst, Josef.- Escales, Erich.- Eskau, Robert F.- Euler, Marita.- Europäische Rundschau.- Studiengemeinschaft der Evangelischen Akademie.- Evangelische Michaelsbruderschaft.- Express-Verlag GmbH.-

Falk, Gottfried.- FIAT-Review.- FIAT Agency.- Filipczyk, Ernst.- Fischer, Heinrich.- Flachsbart.- Fleidl, Anny.- Fleischer, Carl Friedrich.- Fleischmann, R.- Förster, Ludwig.- Förtner, W. Siegfried.- Forschungen und Fortschritte (Verlag).- Forstwissenschaftliche Fakultät der Universität Göttingen.- Frankenpost.- Bergakademie Freiberg.- Goethehaus Frankfurt/Main.- Frels, Elisabeth.- Ingenieurbüro Frenz und Hack.- Freudenberg, K.- Frey, W.- Fricke, Robert.- Fricke, Winfried.- Friedrich-Freksa.- Fritsch, Johann.- Fuchs, Hans Georg.- Fues, Erwin.- Fuss, Ferdinand Robert.- Futscher, Klaus.-

Gänserich, Dora.- Gatz, Konrad.- Gentner, W.- Gerhardt, R. M.- Gerlach, Walther.- Gesellschaft Deutscher Chemiker.- Gessner, Fritz.- Greiß, Franz.- Glodniok, Elisabeth.- Gmelin-Institut.- Gnatz.- Glöckner, Martin.- Goder, Franz.- Göller, A.- Görlich, Hedwig .- Goetsch, W.- Goette, Rudolf.- Gläß.- Goldstein, Hans.- Gottschalk.- Grabbe, Ingrid.- Graphot Service GmbH.- Greve.- Grosse.- Grüninger, Gertraud.- Grunert, Georg.- Gruppe sozialistischer Verleger.- Gschwender, Helene.</text>    <col3>1946-1947</col3>

<register art="Personen-Index">Abel, Kurt</register>

<register art="Personen-Index">Ablakos, Stefan</register>

<register art="Personen-Index">Achelis, Johann Daniel, Geb.: 07.06.1898</register>

<register art="Personen-Index">Arfsten, Arvest Newton</register>

<register art="Personen-Index">Artmann, Kurt</register>

<register art="Personen-Index">Auersbach, K.</register>

<register art="Personen-Index">Auvers</register>

</item>

Herausforderung

* Daten sind nicht valide
* Es sind Text Mining-Verfahren notwendig um die Personennamen aus den Daten zu extrahieren
* keine weiterführenden Informationen zu Personen, oftmals nur Nach- und Vorname
* Vermischung von Körperschaften, Personen und Freitext

Bewertung

* Bestand ist sehr schlecht für eine Anreicherung mit Normdaten geeignet

Ergebnis

|  |  |
| --- | --- |
| Einrichtung | Max-Planck-Archiv |
| Datei | III\_93\_10050\_10012\_DDBNORMXML.xml |

Gesamtergebnis

|  |  |
| --- | --- |
| Anzahl der Entitäten | 205.466 |
| Anzahl der Übereinstimmungen mit einem Personendatensatz in der GND | 9.210 (4,5%) |
| Davon individuelle Personen | 7.179 |