Entitäten- und Relationsextraktion

Sophia Stotz, Valentina Stuß

	Entitäten	Relationen	Disambiguierung
Ansatz	• Regelbasierte Verfahren zur Erstellung der Lexika + Register + GND • Umformung bestehender Datenbanken	• Regelbasierte Verfahren zur Patternextraktion	• Themenähnlichkeit der Lebensartikel zu Personen
Durchführung	• Lokale Grammatiken • Unitex	• Lokale Grammatiken • Entitäten • Unitex	• TopicZoom • LDA
Ergebnis	**Lexika mit Entitäten:** • Personen • Orte • Organisationen • Fächer • Berufe	**Extraktion der Relationen:** • Studieren • Lehrer / Schüler • Besuchen • Emigrieren • Freunde **Ergebnisse für „studieren“:** Precision = 95% Recall = 67% F-Score = 78%	**Probleme:** • Mit 95%-tiger Genauigkeit können nur 3% der Fälle zugeordnet werden • Viele Parameter (Anzahl möglicher Topics) nicht ausgetestet • „Überblendung“ der Kandidaten mit Artikeln • Unterschiedliche „Sprachstufen“ der ndb und adb

Relationen

Extraktion von Personenbeziehungen

Relationstypen:

1) Hierarchische (über- / untergeordnet):

- Lehrer – Schüler
- Einfluss auf
- Einfluss von

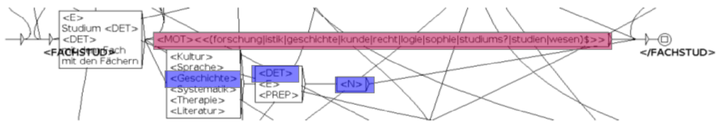
→

- promovieren;
- studieren / studiren;
- habilitieren;
- ...

2) Gleichberechtigte (positiv / negativ / neutral):

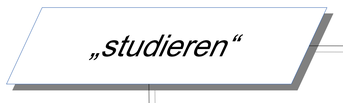
- Freund;
- Feind;
- Kollege.

→ befreundet mit; sein / <Name>'s Freund <Name>; Freundschaft schließen mit ...; … verbindet … Freundschaft; seine / <Name>'s Freundschaft mit …; u.ä.



SUCHMUSTER STUDIENFACH (Auszug)

Entitäten



Treffer:
...studierte 1909-13 Geographie, Geschichte der Volkswirtschaftslehre und Germanistik an den Universitäten <placeName type="study" key="nom:609246 osm:240058050" n="1385">Heidelberg</placeName> und <placeName type="study" key="nom:626076 osm:240111242" n="1409">Leipzig</placeName>

Dann studierte er Hüttenkunde an der Bergakademie <placeName type="study" key="nom:532485 osm:162012321" n="1810">Berlin</placeName>.

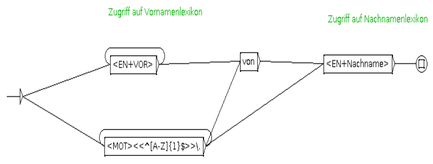
A. Erkennung von Personennamen

1. Bereits im Korpus markierte Personennamen

Stellung ein wie sein Vetter <persName key1="ppn116001127" n1="sfz69276" resp1="pa 1809 bis 1872preußischer Diplomat">Heinrich Abeken</persName> bei Bismarck. Seine statis
chters und Rhetorikprofessors <persName key1="ppn116713550" n1="sfz60445" resp1="pa 1508 bis 1560neulateinischer Dichter">Georg Sabinus</persName> an, so wie auch sein Lebe
wurde er als Nachfolger von <persName key1="ppn117185930" n1="sfz67334" resp1="pa 1839 bis 1912Physiologe">Hermann Munk</persName> auf den Lehrstuhl für gesamte Physiolog
M. Moscherosch</persName> und <persName key1="ppn11850021X" n1="sfz72" resp1="pa 1644 bis 1709Prediger">Abraham a Sancta Clara</persName> ein. Diese Sammlung fand Jahrhunde
. Nachfolger des alten Dekans <persName key1="ppn118517384" n1="sfz23420" resp1="#mr pa 1504 bis 1575Reformationstheologe">Heinrich Bullinger</persName>. Aber schon einige

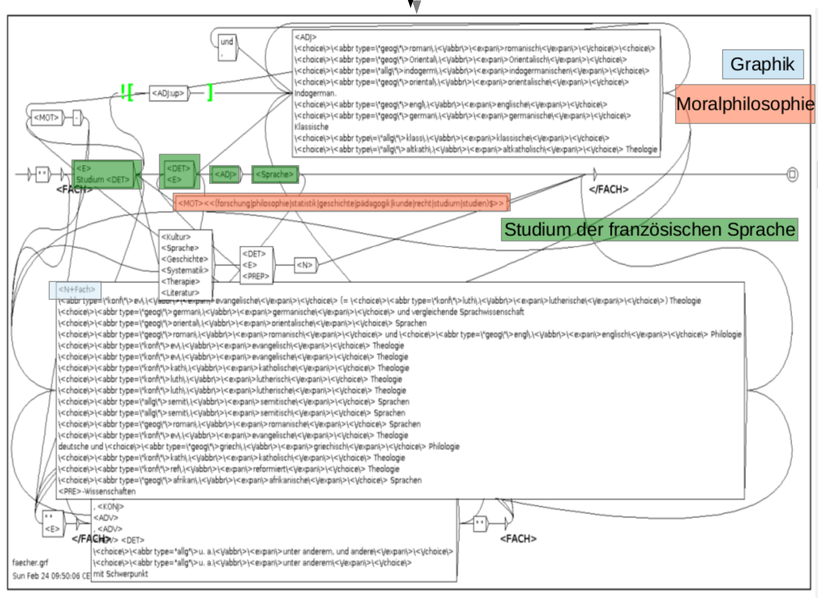
2. Im Korpus nicht markierte Personennamen: Erkennung mithilfe sprachlicher Muster und Lexika

und <persName>D. C. von Lohenstein</persName>), obwohl die Gesamtausgab
von <persName>D. Chodowiecki</persName>. Sein Werk, das zeitweise unter
Amte <persName>D. F. Loos</persName> ist er der eigentliche Repräsentant
mit <persName>Daniel Casper von Lohenstein</persName>. Sein Dichterruhm
von <persName>E. M. von Hornbostel</persName>, mit dem er ebenso wie mi
ung (<persName>E. Mörike</persName>, <persName>L. Uhland</persName>, <pe
me>, <persName>E. Schele</persName>, <persName>J. K. B.</persName> <GEO>
rer <persName>E. Sueß</persName>, 1899 <choice><abr type="allg">Dr. ph
ics"><persName>Eberhard I.</persName></hi> nachweisbaren edelfreien baye
ics"><persName>Eberhards III.</persName></hi> Schwester, die „heilige“ <
1919Chemiker">Emil Fischer</persName> fesselten den angehenden Forscher



Klerikalseminar ein, wo er mit <PERS1 resp="Freund"><persName>Karl Adam</persName></PERS1>, dem späteren Tübinger Theologieprofessor Freundschaft schloß. Der Kirchenkomponist <persName><choi
d Mitarbeiter des Thomaskantors <PERS1 resp="Freund"><persName>Karl Straube</persName></PERS1>, mit dem ihn alsbald eine fruchtbare Freundschaft verband. Nach Tätigkeit als Chorleiter und Org
mit deren Vorstandsvorsitzenden <PERS1 resp="Freund"><persName>Martin Rehmer</persName></PERS1> ihn eine enge Freundschaft verband. <persName key1="ppn118595938" n1="sfz96971" resp1="#mr pa 1
er, vor allem den Zürcher Maler <PERS1 resp="Freund"><persName>Max Gubler</persName></PERS1>, mit dem ihn eine schicksalhafte Freundschaft verband. <choice><abbr type="lemma">J.</abbr><expan>
später Schalk und - vor allem - <PERS1 resp="Freund"><persName>Richard Strauss</persName></PERS1>, mit dem ihn eine jahrzehntelange Freundschaft verband. Die bei der Vorbereitung seiner Opern
r Begründer der Anthroposophie, <PERS1 resp="Freund"><persName>Rudolf Steiner</persName></PERS1>, mit dem ihn eine enge Freundschaft verband. <choice><abbr type="lemma">M.</abbr><expan>Molt</
nd mit deren Autoren, besonders <PERS1 resp="Freund"><persName>Sven Hedin</persName></PERS1>, ihn herzliche Freundschaft verband - zu beschränken und dadurch der Firma eine größere Beweglichk
hoice> Humboldts</persName> und <PERS1 resp="Freund"><persName>W. H. Riehls</persName></PERS1>, mit denen ihn enge Freundschaft verband, wie durch bedeutende, zum Teil von ihm angeregte Werke
Aufbau zu<pb n="35"/>fertigen. Mit <PERS1 resp="Freund"><persName key1="ppn118512374" n1="sfz4955" resp1="pa 1827 bis 1901Maler">Arnold Böcklin</persName></PERS1> verband ihn enge Freundscha
"1209">Paris</placeName>, wo er mit <PERS1 resp="Freund"><persName key1="ppn118540564" n1="sfz21631" resp1="pa 1891 bis 1950Dichter">Yvan Goll</persName></PERS1> Freundschaft schloß, dann nac
in</placeName> (1905-09), wo er mit <PERS1 resp="Freund"><persName key="ppn118764152" n="sfz69106" resp="#vs #pr">Karl Walser (<birth n="1877">1877</birth>-<death n="1943">1943</death></persN
1810">Berlin</placeName>, wo er mit <PERS1 resp="Freund"><persName key="ppn118773275" n="sfz60694" resp="#vs #pr">Leopold Zunz (<birth n="1794">1794</birth>-<death n="1886">1886</death></pers
1810">Berlin</placeName>, wo er mit <PERS1 resp="Freund"><persName key="ppn118777734" n="sfz70398" resp="#vs #pr">Georg Kolbe (<birth n="1877">1877</birth>-<death n="1947">1947</death></persN
nesungsurlaubes, in dem er auch mit <PERS1 resp="Freund"><persName><choice><abbr type="vorname">Heinr.</abbr><expan>Heinrich</expan></choice> Lersch</persName></PERS1> Freundschaft schloß, ve
istig anregenden Atmoshäre auf. Mit <PERS1 resp="Freund"><persName>E. T. A. Hoffmann</persName></PERS1> verband ihn lebenslange Freundschaft. Seit 1791 studierte er Rechts- und Staatswissensc

nden Freundschaft mit dem Tübinger Psychiater <PERS2 resp="Freund"><persName key1="ppn118715909" n1="sfz45965" resp1="pa 1888 bis 1964Psychiater">Ernst Kretschmer</pe
ihen Freundschaft mit dem um drei Jahre ältern <PERS2 resp="Freund"><persName>J. J. Bodmer</persName></PERS2> erfreute er sich dessen Mitarbeit, wie er anderseits des
ätten Freundschaft mit dem welterfahrenen <PERS2 resp="Freund"><persName>Moritz <choice><abbr type="allg">v.</abbr><expan>von</expan></choice> Schwind</persName></PERS
ice> Freundschaft mit den Kunsthistorikern <PERS2 resp="Freund"><persName key="ppn11873329X" n="sfz70708" resp="#vs #pr">Erich Meyer (<birth n="1897">1897</birth>-<de
hloß Freundschaft mit den Malern <PERS2 resp="Freund"><persName><choice><abbr type="vorname">Frdr.</abbr><expan>Friedrich</expan></choice> Müller</persName></PERS2> u
ange Freundschaft mit den Musikern <PERS2 resp="Freund"><persName key1="ppn118693948" n1="sfz17863" resp1="pa 1799 bis 1853Musiksammler" key2="ppn117575550" n2="sfz17
ange Freundschaft mit der dort als Pflegerin arbeitenden Schauspielerin <PERS2 resp="Freund"><persName><!--PRE+PRE-->Sarah Bernhardt</persName></PERS2>. 1872-76 lehrt

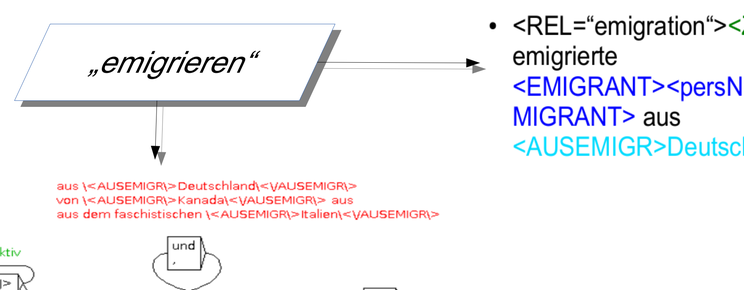


<ZEIT>1867-69</ZEIT> und <ZEIT>1871-72</ZEIT> studierte er in <ORT><placeName type="study" key="nom:532485 osm:162012321" n="1810">Berlin</placeName></ORT>, <choice><abbr type="allg">u. a.</abbr><expan>unter anderem</expan></choice> bei <BEI><persName>E. E. Kummer</persName></BEI>, <BEI><persName>K. Weierstraß</persName></BEI> und <BEI><persName>H. Helmholtz</persName></BEI>

Zeit: 1867-69, 1871-72

Ort: Berlin

Bei: E. E. Kummer, K. Weierstraß, H. Helmholtz



- <REL="emigration"><ZEITEMIGR>1933</ZEITEMIGR> emigrierte <EMIGRANT><persName>Lichtenstein</persName></EMIGRANT> aus <AUSEMIGR>Deutschland</AUSEMIGR></REL>.

Disambiguierungsverfahren

B. Erkennung von Toponymen

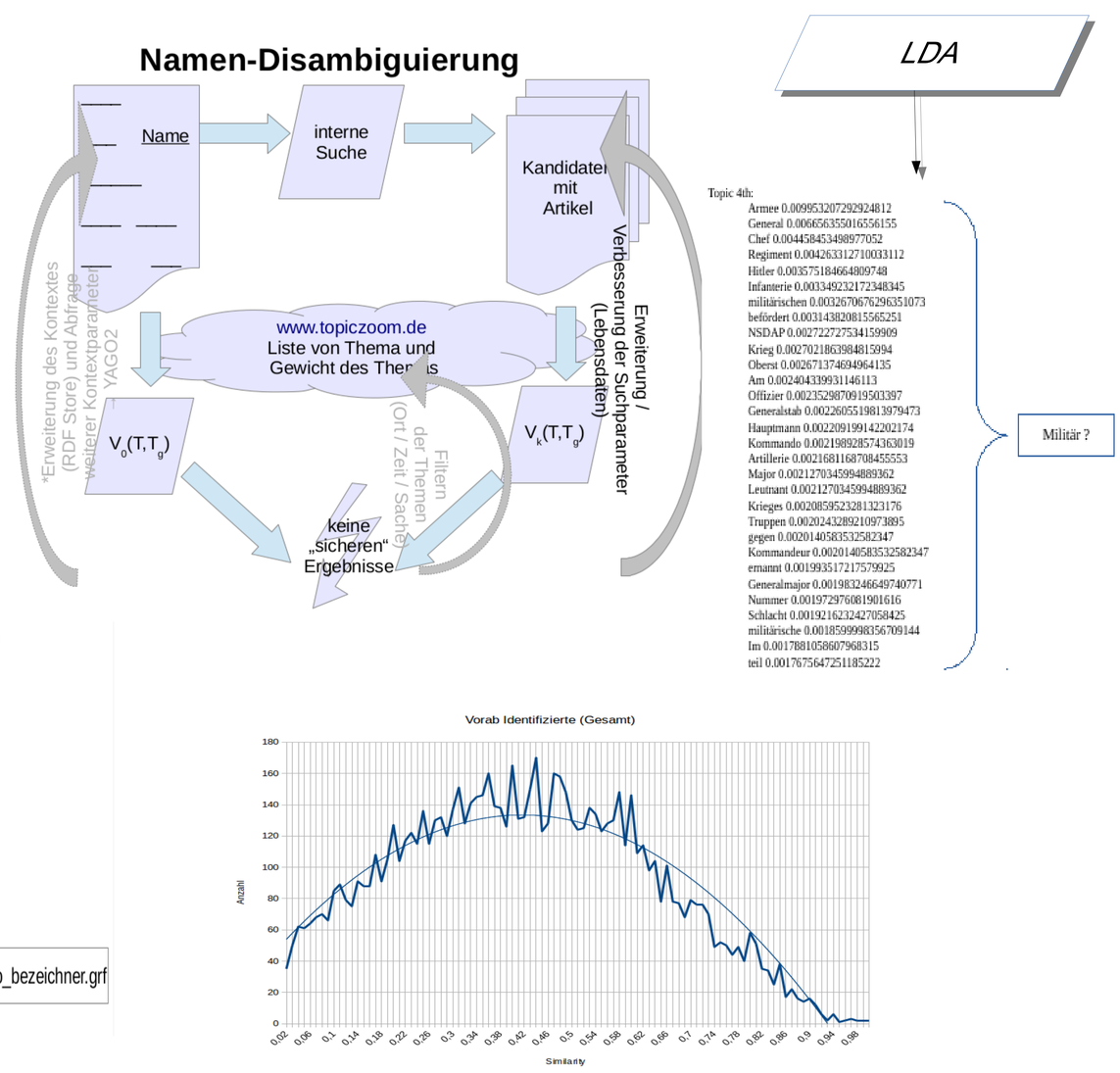
→ anhand der Suffixe (vgl. Nagel, 2008);
→ Himmelsrichtungen (vgl. Geierhos, 2010);
→ Toponymbezeichner (vgl. ebd.);
→ Nähere Bestimmungen der Form:
- Alaska (Nordamerika);
- Amerika (Nord- und Südamerika);
- Nordost-Afrika;
- Brooklyn/New York;
- Dortmund-Hörde-Aplerbeck-Unna;

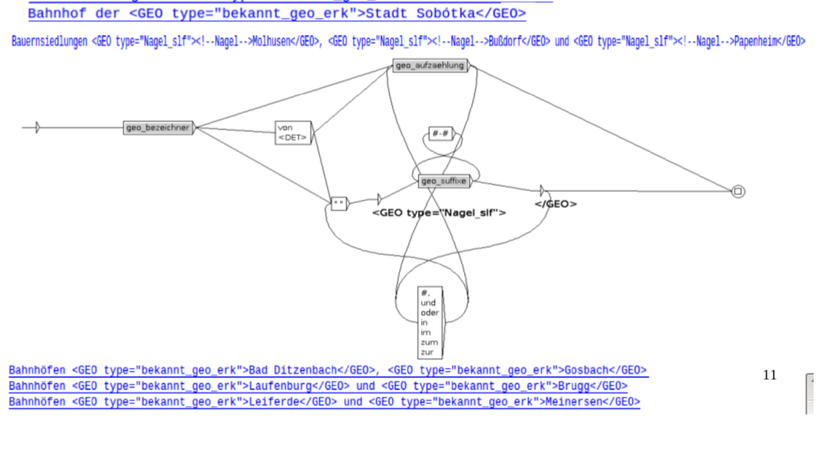
→ Aufzählungen (,; und; sowie);
→ Bewegungsverben und Präpositionen (kam aus …, ging über … nach …, begab sich nach …).

2 Schritte:

1. Lokalisierung (5 Hauptgraphen, Subgraphen eingebunden)
2. Ergänzung und Vervollständigung (2 Graphen)

- Anwendung auf das Korpus in zwei Kaskaden





Subgraph: geo_bezeichner.grf