

Korrektur und Anreicherung von TEI-Daten durch einen Gamification-Ansatz

Frank Fischer¹ & Mathias Göbel²

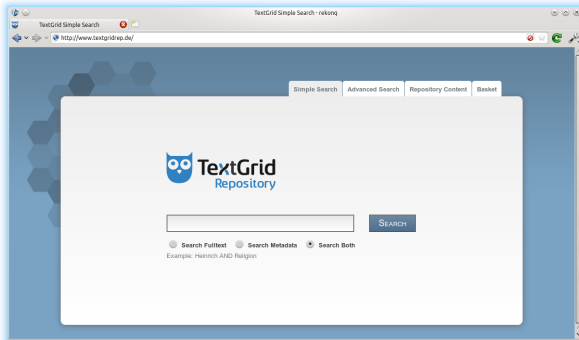
¹Göttingen Centre for Digital Humanities

²Göttingen State and University Library

¹frank.fischer@zentr.uni-goettingen.de, ²goebel@sub.uni-goettingen.de

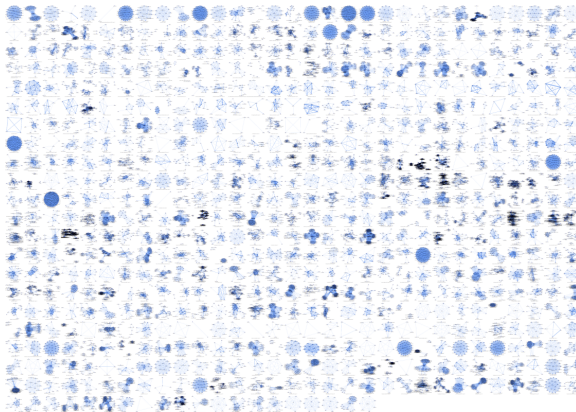
11. Februar 2015

Im Zentrum des Interesses



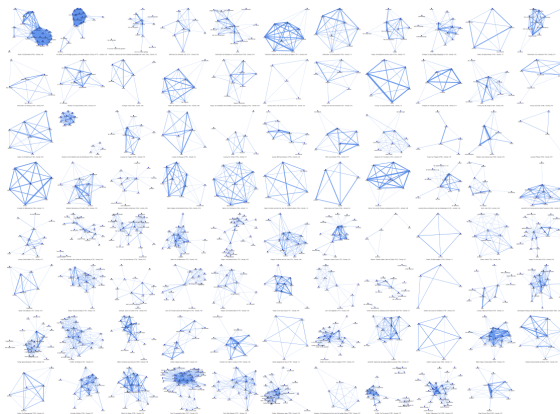
`textgridrep.de / genre:drama / tei:speaker/string()`

Anwendungsfall Netzwerkanalyse (1)



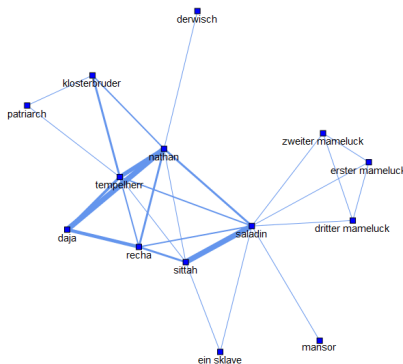
TextGrid Repository, 671 Dramen (5. Jh. v. Chr. bis 1930)

Anwendungsfall Netzwerkanalyse (2)



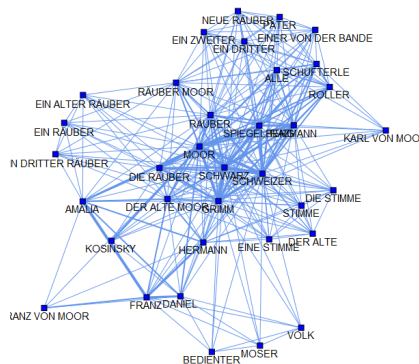
TextGrid Repository, 87 Dramen aus dem 18. Jahrhundert

Anwendungsfall Netzwerkanalyse (3)



Lessing: *Nathan der Weise* (1779), Netzwerk gemäß ausgelesenem
tei:speaker/string(), nachbearbeitet

Anwendungsfall Netzwerkanalyse (4)



Schiller: *Die Räuber* (1781), Netzwerk gemäß ausgelesenem
`tei:speaker/string()`, unbehandelt

Trouble

Probleme bei Personennamen, wenn `@who/xml:id` nicht vergeben sind:

- übernommene Druckfehler
- OCR-Fehler
- Abkürzungen aus editorischen Gründen
- Inkonsistenzen bei der Sprecherangabe (“Odoardo” vs. “Odoardo Galotti”)

Wie es ist: DTA vs. TextGrid Rep

```
<sp who="#FAU">
  <speaker>
    <hi rendition="#g">Faust</hi>.
  </speaker><lb/>
  <p>Habe nun, ach! Philosophie,<lb/>
  Juristerey und Medicin,<lb/>
  Und leider auch Theologie!<lb/>
  Durchaus studirt, mit heißem Bemühn.<lb/>
  [...]
```

```
<sp>
  <speaker xml:id="tg1386.2.7">FAUST.</speaker><lg>
  <l rend="zenoP1m4n4" xml:id="tg1386.2.8">Habe nun, ach! Philosophie,</l>
  <l rend="zenoP1m4n4" xml:id="tg1386.2.9">Juristerei und Medizin,</l>
  <l rend="zenoP1m4n4" xml:id="tg1386.2.10">Und leider auch Theologie</l>
  <l rend="zenoP1m4n4" xml:id="tg1386.2.11">Durchaus studiert, mit heißem Bemühn.</l>
  [...]
```

Beginn des *Faust I* (1808), oben: DTA, unten: TextGrid Rep

Entstehung des TextGrid Rep

Beispiel: Adelung, Artikel *Aufreisen*

zeno-xml:

```
<article>
<lem>Aufreisen</lem>
<text><p>* <lemfloat> Aufreisen</lemfloat>, <i>verb. reg. neutr.</i> mit dem Hülfs Worte
    seyn, aufwärts reisen, in die Höhe fahren; ein im Hochdeutschen unbekanntes
    Zeitwort, S. Reisen.</p><br/>
<p class="zenoLY">Nachdem Messias war zum Vater aufgereist,</p>
<p class="zenoPR">Opitz.</p><br/></text>
</article>
```

TEI P5:

```
<entry>
<form>
<usg type="time">* </usg> <form type="headword,, norm="Aufreisen">Aufreisen</form>
</form>
<sense>, <hi rend="roman">verb. reg. neutr.</hi> mit dem Hülfs Worte seyn, aufwärts
    reisen, in die Höhe fahren; ein im Hochdeutschen unbekanntes Zeitwort, S.
    Reisen.<cit>
<quote><seg rend="zenoLY">Nachdem Messias war zum Vater aufgereist,</seg></quote>
<bibl><seg rend="zenoPR">Opitz.</seg></bibl></cit>
</sense>
</entry>
```

Transformation des zeno-xml in TEI

Was tun?

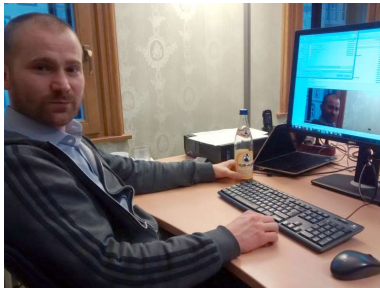
Typen von Bestimmungsmöglichkeiten:

- *is* (Identität von Personen)
- *aggregates* (Gruppen von Personen)
- *@who* (IDs à la DTA)
- perspektivisch: Auszeichnung qualifizierender Beziehungen von Personen (Verwandtschaftsbeziehungen etc.)

Erste Möglichkeit

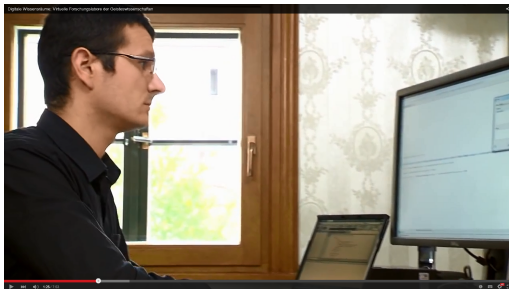
Also: eigentlichTM simpel zu lösende Aufgabe, die aber vom Computer nicht zufriedenstellend gelöst werden kann (Messung und Auswertung der Levenshtein-Distanz zwischen verschiedenen Personennamen hier z. B. ungeeignet).

Zweite Möglichkeit (1)



So ...

Zweite Möglichkeit (2)



... oder so ...

Dritte Möglichkeit (1)

Wir lösen die Aufgabe mit einer GUI, die für uns leicht zu nutzen ist:

- im Browser
- auf verschiedenen Geräten
- eine einfache Abfolge durchlaufend

Dritte Möglichkeit (2)

Gamification!



`https://personae.gcdh.de`

Technischer Aufbau des Spiels

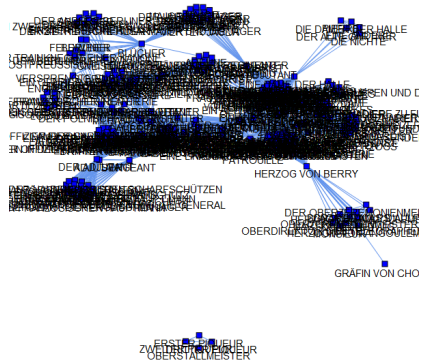


Punktesystem

Punkte werden vergeben für:

- markierte *is*-Relationen
- für ausgewählte Aggregationen (“Alle”, “Beide” usw.)
- das Verifizieren der Eingaben anderer
- wenn ein Dokument als ‘done’ markiert wird
- ...

Final Encounter



C. D. Grabbe: *Napoleon oder Die hundert Tage* (1831)