



CorpusExplorer v2.0

Seminartauglich in einem halben Tag

JAN OLIVER RÜDIGER, UNIVERSITÄT KASSEL

FREDERIK WALENDCZUS, UNIVERSITÄT KASSEL

DHd2018 – Kritik der digitalen Vernunft
Köln, 26 Februar – 2 März 2018

U N I K A S S E L
V E R S I T Ä T



Jan Oliver Rüdiger, M.A.

Studierte Germanistik, Informatik und Kunstwissenschaft an der Universität Kassel.
Aktuelles Disserationsprojekt: Korpushermeneutik - Ein korpuslinguistischer Ansatz zur Textanalyse
Arbeitet als IT-Freelancer/Projektberater in Düsseldorf.
Email: e-mail@jan-oliver-ruediger.de, Twitter: @notesjor.



Frederik Walenczus

Studiert Lehramt L3 – Deutsch und Geschichte an der Universität Kassel
Tutor/Wissenschaftliche Hilfskraft am Institut für Germanistik – Universität Kassel.
Email: frederik.walenczus@uni-kassel.de, Twitter: @FWalenczus



Workshop-Materialien

Online-Kurs (Video-Tutorials, Anmerkungen, OpenAccess)

<https://lernen-mit.jan-oliver-ruediger.de>

CorpusExplorer (Download, Handbuch, Korpora / Addons)

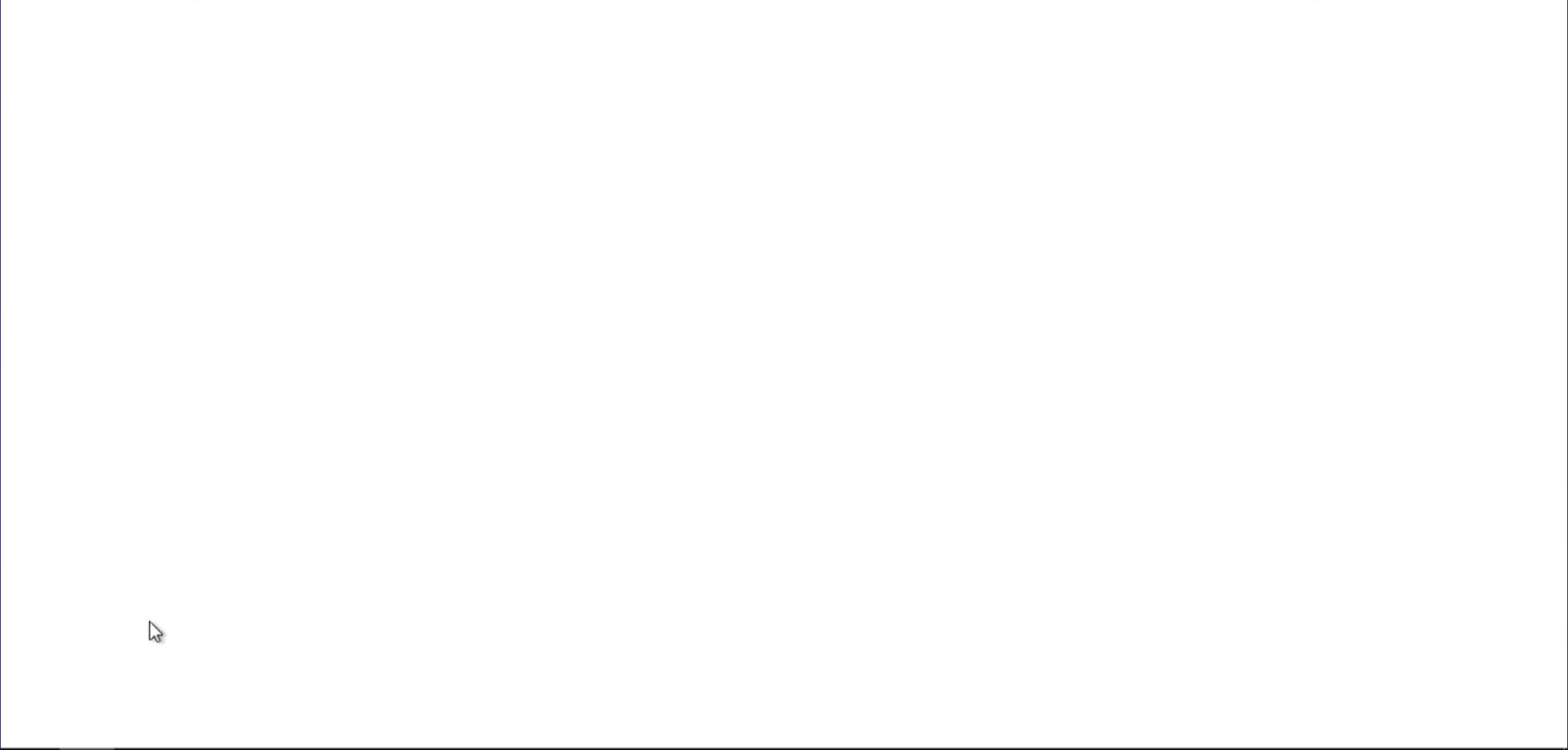
<http://www.CorpusExplorer.de>

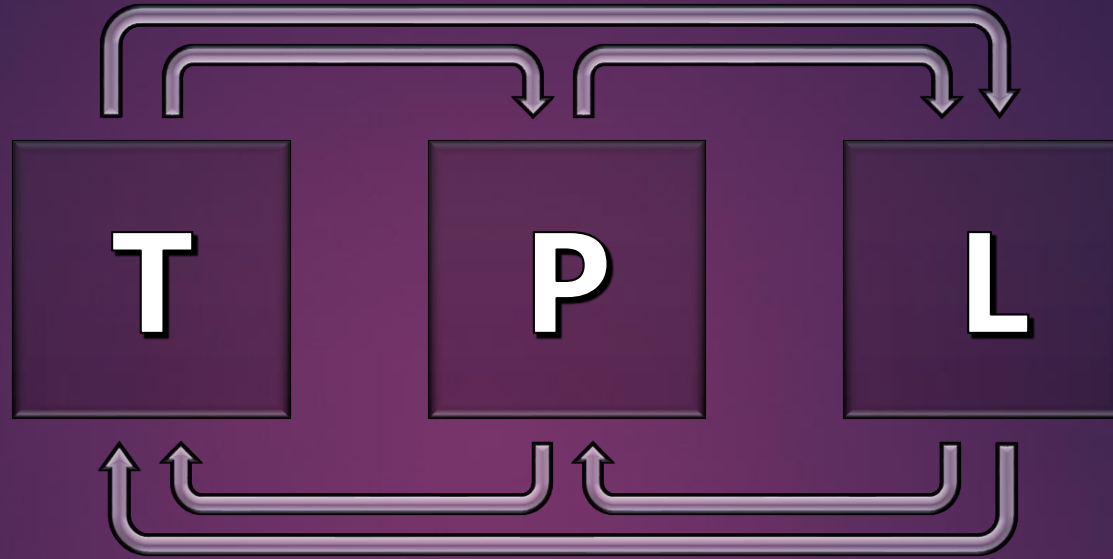


Installation

- Läuft auf Windows ab Vista (setup.exe übernimmt die Gesamtinstallation)
- Linux / MacOS folgende Lösung:
 - Virtualisierung mittels VirtualBox – oder ähnlichen (setzt Windows-Lizenz voraus).
 - Nativ mit CrossOver (kostet ca. 40 Euro – 30 Tage Testversion verfügbar).
 - **Heute:** Das Hochschulrechenzentrum Kassel stellt eine VDI mittels VMWare Horizon bereit. Vorteil: Kann auch auf schwachen Windows-Notebooks oder auf iPad genutzt werden. (*Hinweis:* Hat durch einen Update-Fehler aktuell einige Einschränkungen)

<http://www.CorpusExplorer.de>





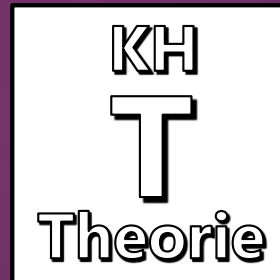
Promotionsprojekt:
„Korpushermeneutik – Ein korpuslinguistischer Ansatz zur Textanalyse“

Literaturwissenschaft
... / Deduktiv / Based / Theorie

Sprachwissenschaft
Empirie / Driven / Induktiv / ...

Hermeneutik + Korpuslinguistik

W. Teubert (2006)
Korpuslinguistik,
Hermeneutik und die soziale
Konstruktion der Wirklichkeit.

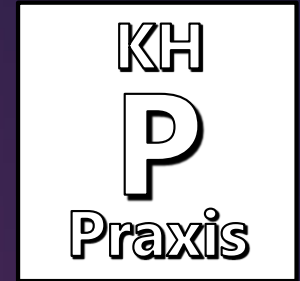


U. Haß (2007)
Korpus-Hermeneutik: zur
hermeneutischen Methodik
in der lexikalischen Semantik

1. Die Analyse muss mehrere abwechselnde Zyklen durchlaufen
2. Die Entwicklung des Wissens geschieht durch Falsifikation
3. Aufzeigen (Visualisieren) von Handlungsmöglichkeiten

Aufbereitung / Auswertung

- Entweder: Viele kleine Einzelprogramme



Vorteil:

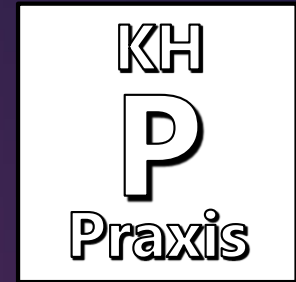
- Flexibel / Individuell

Nachteil:

- Erfordert viel Know-How

Aufbereitung / Auswertung

- Entweder: Viele kleine Einzelprogramme
- Oder: Ready-To-Use Plattform



Beispiele:

- GATE
- Apache UIMA
- CorpusExplorer

Vorteile:

- Großer Funktionsumfang
- Flexibel bei Bedarf
- Einfach in der Nutzung*

Nachteil:

- Komplex in der Entwicklung
- * Alleinstellungsmerkmal CorpusExplorer



- ▶ Unterstützt über 100 unterschiedliche Datei-/Textformate für Im-/Export.
- ▶ Integrierter Webcrawler zum Sammeln eigener Webkorpora.
- ▶ Sehr einfache Programmoberfläche / Korpus mit wenigen Mausklicks analysefertig.
- ▶ Anbindung unterschiedlicher Tagger - z. B. TreeTagger, OpenNLP, Stanford POS, uvm.
- ▶ Eigene In-Memory Datenbank – austauschbar gegen SQL- (MySQL, SQLite) und NoSQL-Datenbanken (ElasticSearch). >>> Unbegrenzte Korpusgröße
- ▶ Über 45 Visualisierungen - z. B.: Frequenzanalyse, N-Gramme, Phrasen, Kookkurrenzen...
- ▶ Analysen werden vom Analysematerial mittels Schnappschüssen isoliert.
- ▶ Per Shell/Konsole steuerbar – nützlich für eigene Skripte (R, python, etc.).
- ▶ Flexibles SDK (Software Development Kit) – CorpusExplorer erweitern oder integrieren.