

Integration von KobRA Services in die CLARIN Forschungsinfrastruktur

Prof. Dr. Erhard Hinrichs

Wissenschaftlicher Koordinator von CLARIN-D

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

- Europäische Forschungsinfrastruktur für Sprachressourcen in den Geistes- und Sozialwissenschaften
- Eines von 5 Infrastrukturprojekten im Rahmen der *European Research Infrastructures Roadmap* (ESFRI – European Strategy Forum on Research Infrastructures)
- Ziele:
 - Aufbau einer integrierten, interoperablen und skalierbaren Forschungsinfrastruktur über einen Verbund von Zentren
 - Bereitstellung von linguistischen Daten, Werkzeugen und Diensten
 - Zielgruppe: WissenschaftlerInnen in den Geistes- und Sozialwissenschaften

CLARIN in Germany

- Centres (B):
 - Berlin: BBAW
 - Nijmegen: MPI
 - Mannheim: IDS
 - University of Hamburg
 - University of Leipzig
 - University of Munich (LMU)
 - University of the Saarland
 - University of Stuttgart
 - University of Tübingen
- Computing Centres (E):
 - Garching: RZG
 - Göttingen: GWDG
 - Jülich: RZJ
- Discipline specific working groups (F-AGs)
 - 10 F-AGs each with 10-20 members (not displayed)
 - SSH communities



- Members: (recently joined)

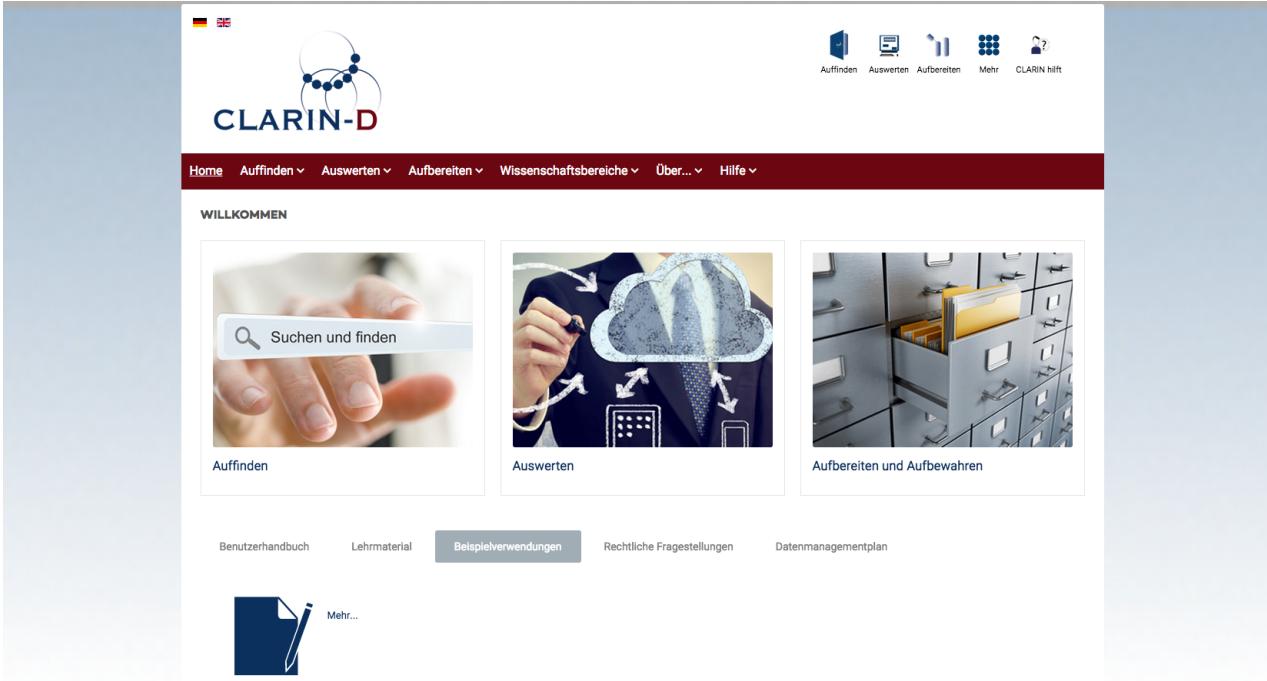
- Austria
- Bulgaria
- Czech Republic
- Denmark
- Dutch Language Union (Belgium, NL)
- Estonia
- Germany
- Greece
- Lithuania
- Netherlands
- Poland
- Portugal
- Slovenia
- Sweden
- Norway

- Candidates (Observers):

- Finland
- United Kingdom



Angebote für die Geistes- und Sozialwissenschaften



The screenshot shows the CLARIN-D homepage. At the top, there are language selection buttons (German, English), the CLARIN-D logo, and a navigation bar with links: Home, Auffinden, Auswerten, Aufbereiten, Wissenschaftsbereiche, Über..., Hilfe. Below the navigation is a section titled "WILLKOMMEN" with three cards: "Auffinden" (searching), "Auswerten" (evaluating), and "Aufbereiten und Aufbewahren" (preparing and storing). Below these cards are links to "Benutzerhandbuch", "Lehrmaterial", "Beispielverwendungen" (highlighted in grey), "Rechtliche Fragestellungen", and "Datenmanagementplan". A "Mehr..." button with a pencil icon is also present. At the bottom left, there's a "Willkommen bei CLARIN-D!" section with text about research data management and a photo of a man. On the right, there's a "NEWS" section with a list of recent updates.

Willkommen bei CLARIN-D!

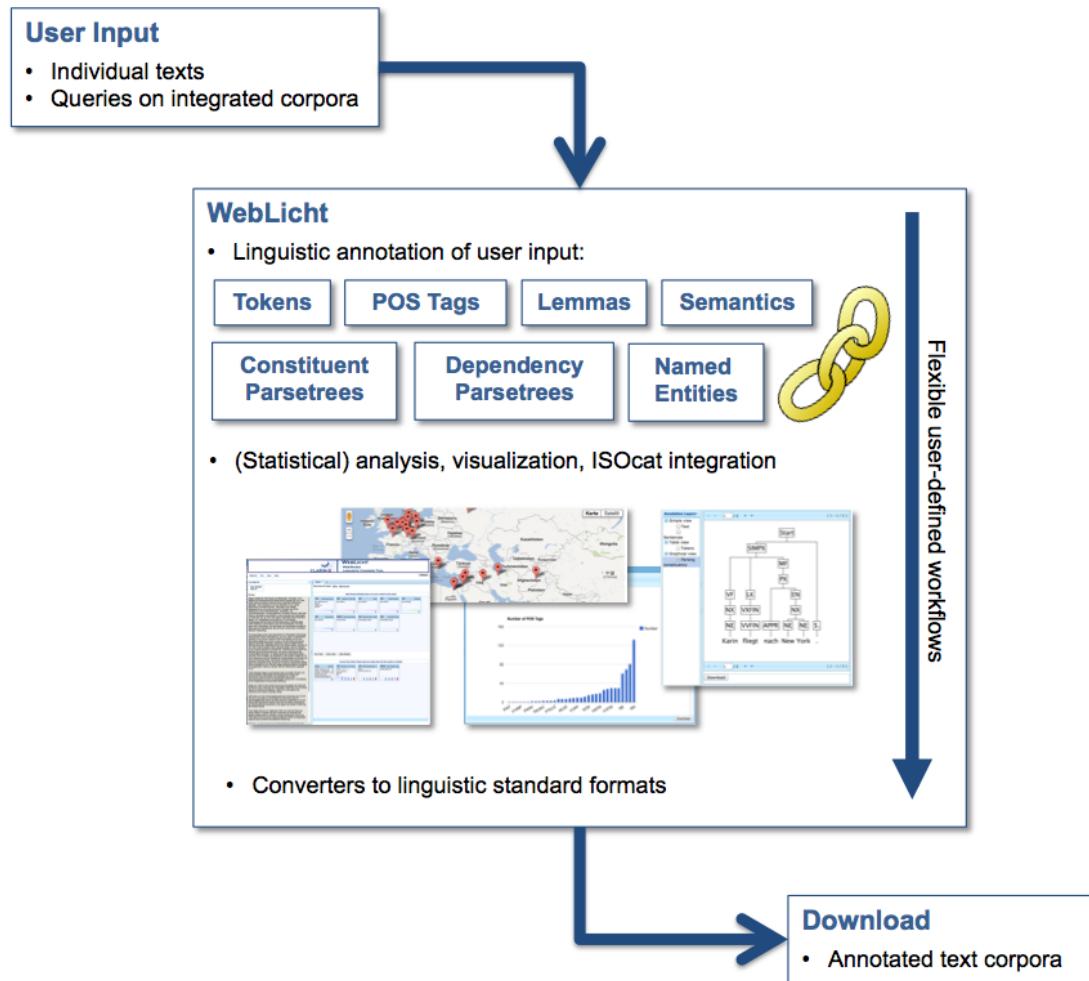
In den Geistes- und Sozialwissenschaften werden Quellen verwendet, die entweder in Forschungsprojekten selbst hergestellt, gesammelt oder aus Archivbeständen übernommen werden. CLARIN-D stellt eine Forschungsdateninfrastruktur für die Geistes- und Sozialwissenschaften dar, die Forschende beim Auffinden, Aufbewahren und Auswerten von Forschungsdaten unterstützt. Darum herum bietet die Infrastruktur Informationen, Lehrmaterial, Hinweise zu Datenmanagementplänen und fachbezogene Unterstützung.

Beim Auffinden von Ressourcen geht es darum, Zugang zu Daten zu erhalten, die der wissenschaftlichen Gemeinschaft zur Verfügung gestellt wurden. Diese Daten werden zitiert und können so gefunden und in anderen Forschungskontexten und zur

NEWS

- Neue Direktorin im CLARIN ERIC
- Europäische Sommeruniversität in Digitalen Geisteswissenschaften
- 3. CLARIN-D Fach-AG Workshop

WebLicht: Web-Based Linguistic Chaining Tool

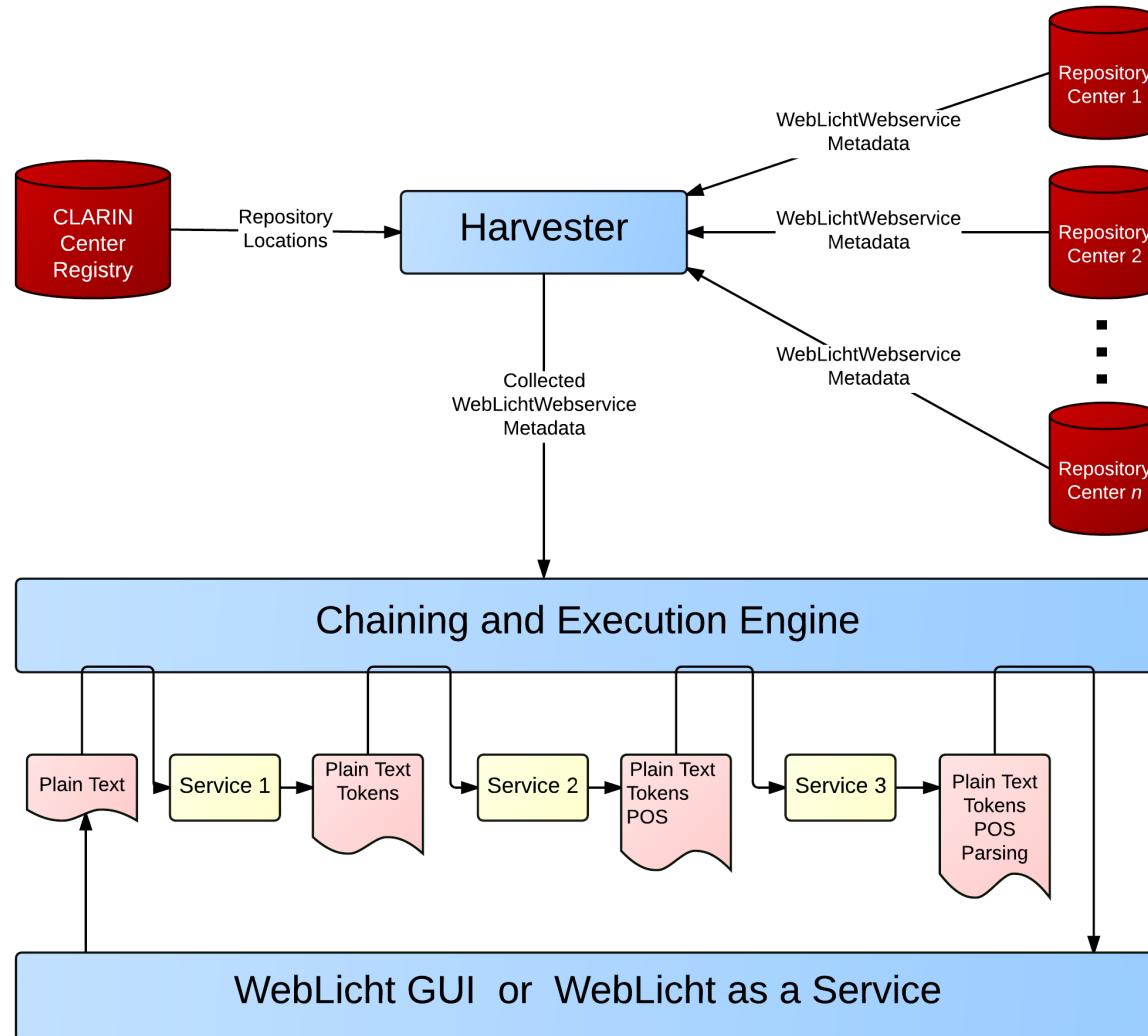


Werkzeuge zur Verarbeitung von natürlicher Sprache müssen häufig von der Kommandozeile / mithilfe von Programmcode bedient werden, aber:

- schwer zu installieren
- nicht immer für alle Betriebssysteme verfügbar
- schwer bedienbar für diejenigen mit wenig Programmier- bzw. Shell-Erfahrung
- Verkettung von Software schwer
 - Ein- Ausgabeformate unterschiedlich
 - verschiedene Plattformen

WebLicht ist eine web-basierte, verteilte Infrastruktur zur Ausführung von NLP Werkzeugen:

- Serviceorientierte Architektur (SOA)
- *Web services* können miteinander kombiniert werden
- Ketten werden ausgeführt indem jedes Kettenelement (web service) via HTTP POST angefordert wird
- Ausgabe des n-ten Kettenglieds ist Eingabe des n+1 Kettenglieds
- *Web service* muss einheitliches Datenformat unterstützen: *Text Corpus Format (TCF)*
- Jeder *web service* reichert die Eingabe an (zusätzliche Annotationsebene(n))



Consequences of the architecture

Es ist einfach, Webservices an WebLicht hinzuzufügen:

- RESTful Webservice entwickeln
- CMDI Metadaten erstellen
- Service und Metadaten zur Verfügung stellen

Anleitungen zum Erstellen eines WebLicht Webservices stehen zur Verfügung.

Es gibt eine Online-Anwendung zum Erstellen und Testen von CMDI Metadaten.

Vielen Dank für Ihre Aufmerksamkeit

Erhard Hinrichs

Email: erhard.hinrichs@uni-tuebingen.de