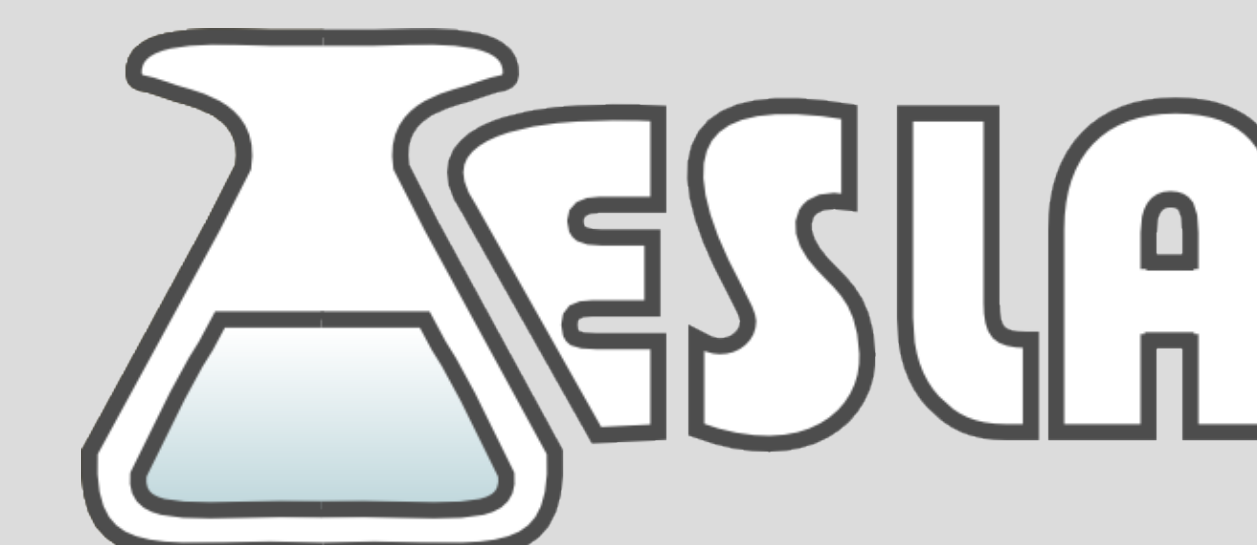


Tesla

Plattform für kooperative, experimentelle Textprozessierung

Stephan Schwiebert, Jürgen Hermes

Universität zu Köln, Institut für Linguistik – Sprachliche Informationsverarbeitung



Was ist Tesla?

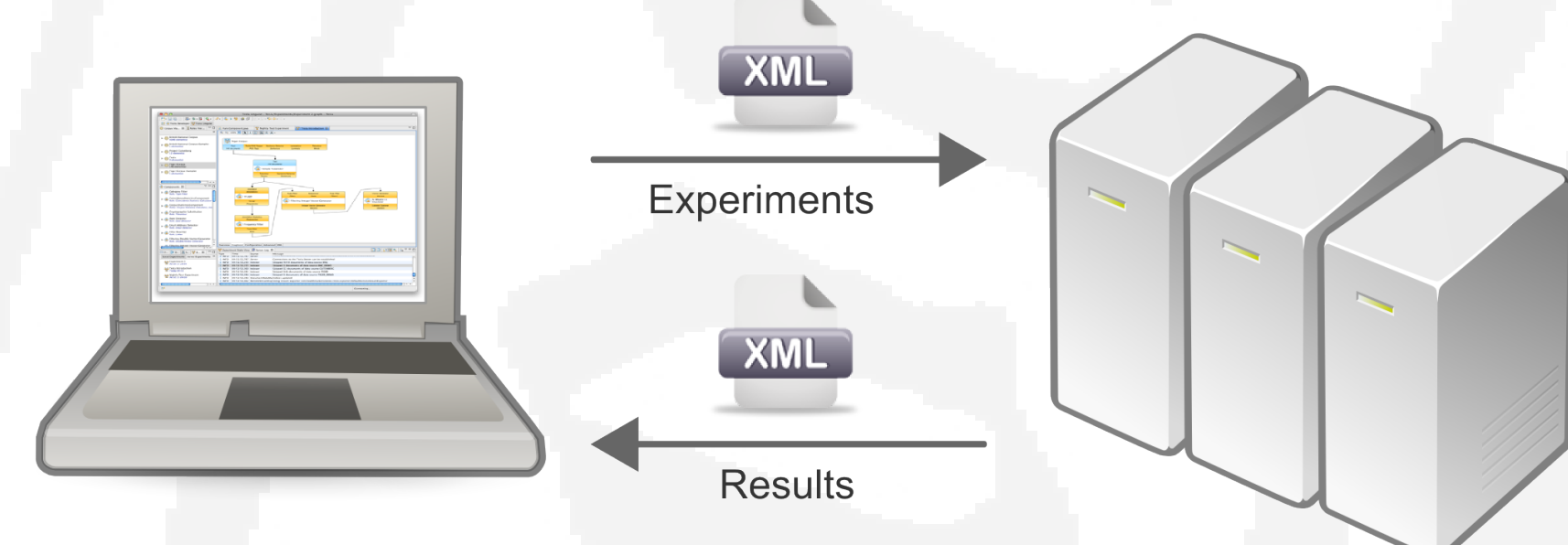
Tesla ist ein Akronym für *Text Engineering Software Laboratory* und bezeichnet ein Softwaresystem, das an der Sprachlichen Informationsverarbeitung (Institut für Linguistik, Universität zu Köln) entwickelt wurde.

Tesla übernimmt die Rolle eines virtuellen Arbeitsplatzes für Wissenschaftler, die mit textuellen Daten arbeiten. Anwendungsgebiete sind dabei:

- Überprüfung von Hypothesen anhand existierender oder selbst zusammengestellter Korpora
- Entwicklung neuer Werkzeuge und Verfahren für die Korpuslinguistik
- Kontrastive Evaluation unterschiedlicher korpuslinguistischer Verfahren
- Publikation von Verfahren und Ergebnissen, die eine Reproduktion der Resultate erlaubt.

Architektur

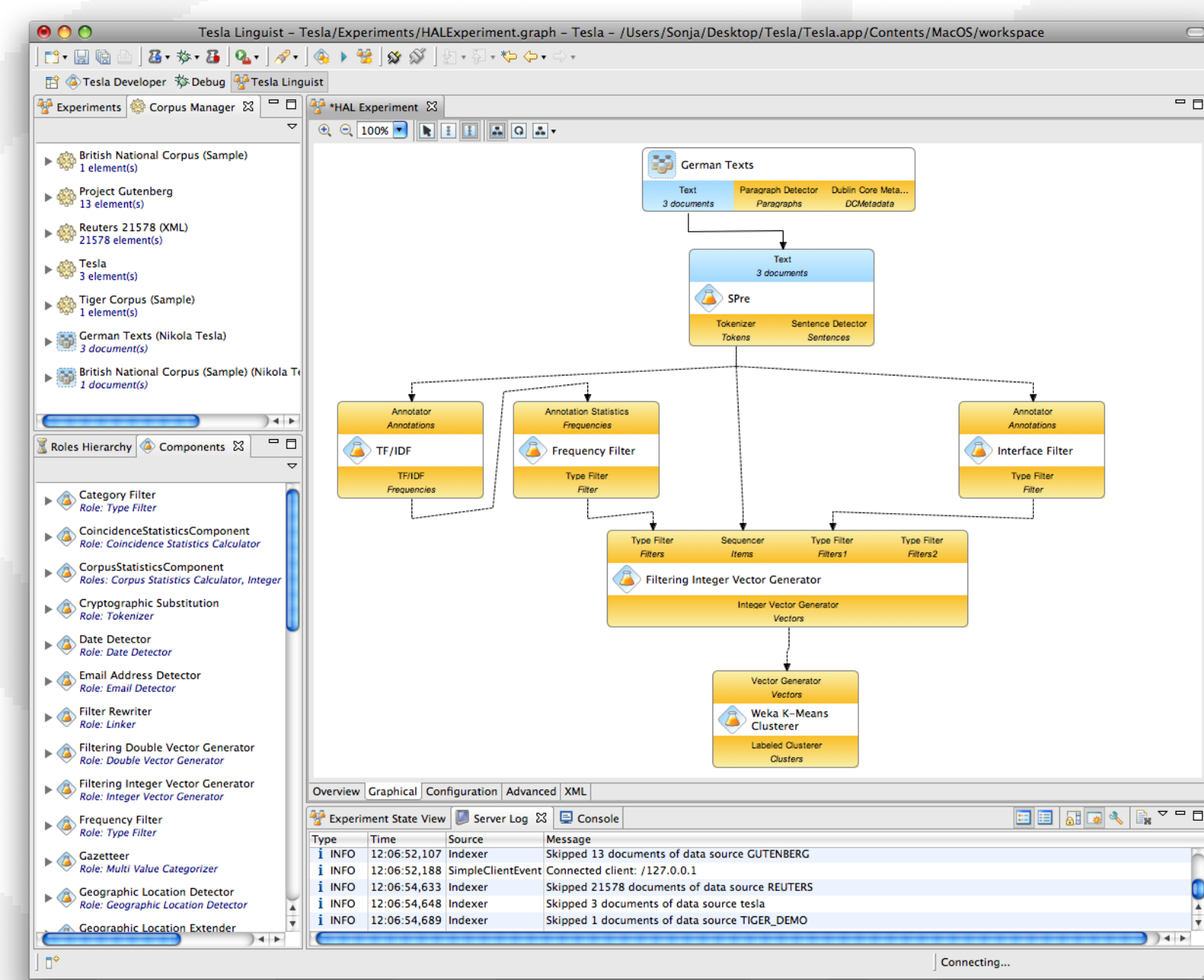
Tesla basiert auf einer Client-Server-Architektur, bei welcher der Eclipse-basierte Client für die Verwaltung von Korpora, die Zusammenstellung und die Konfiguration von Experimenten sowie die Visualisierung der Ergebnisse genutzt wird, während der Server für die Datenhaltung und die Prozessierung zuständig ist.



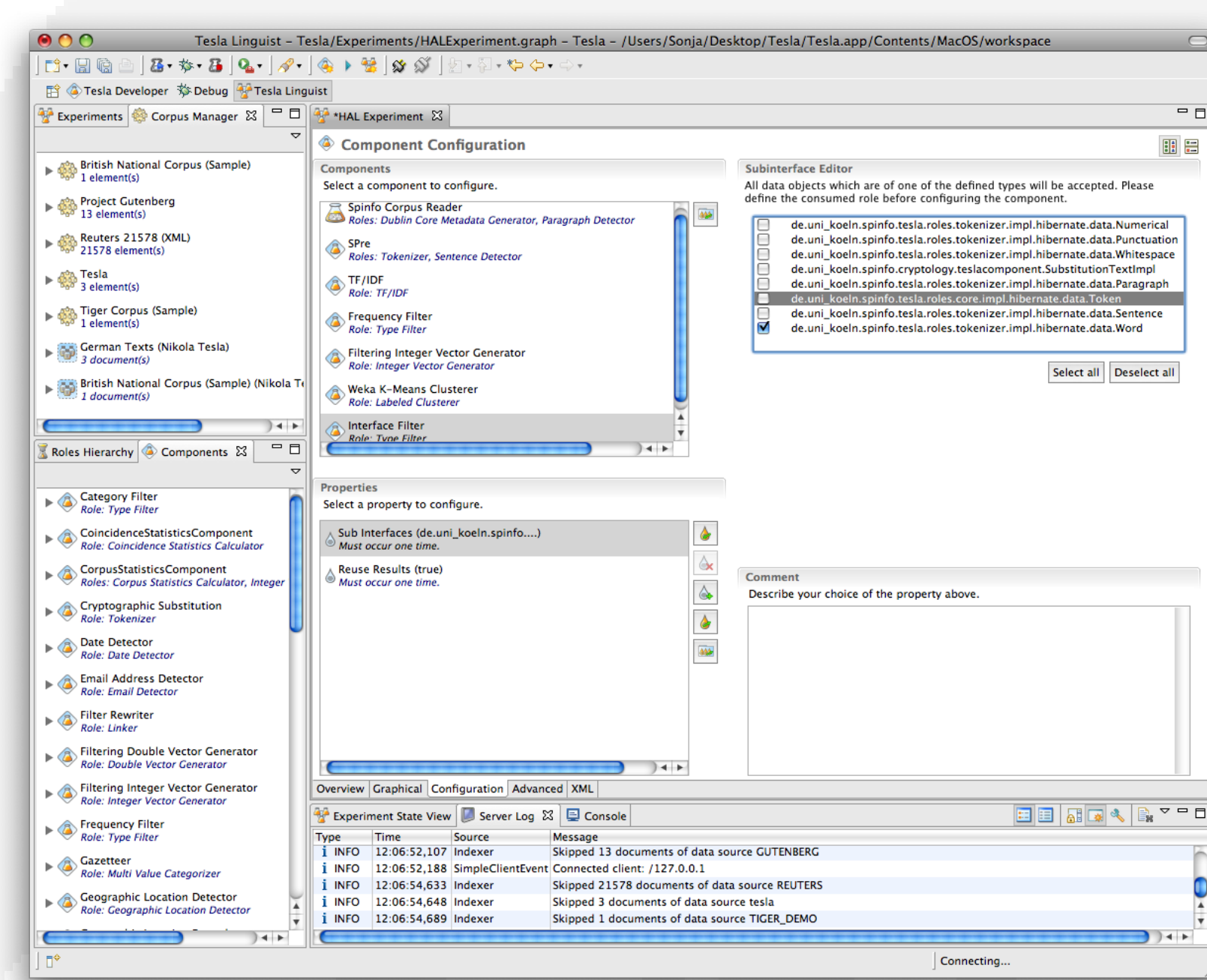
Die Client-Server-Architektur ermöglicht einen direkten wissenschaftlichen Austausch innerhalb von Arbeitsgruppen ebenso wie eine optimierte Ausnutzung von Rechenkapazität.

Experimente

"Experiment" ist in Tesla der Oberbegriff für alle Verfahren, die auf der Prozessierung textueller Daten beruhen.



Experimente bestehen aus Text-Selektionen und Komponenten, die spezifische Tasks auf den Daten ausführen.



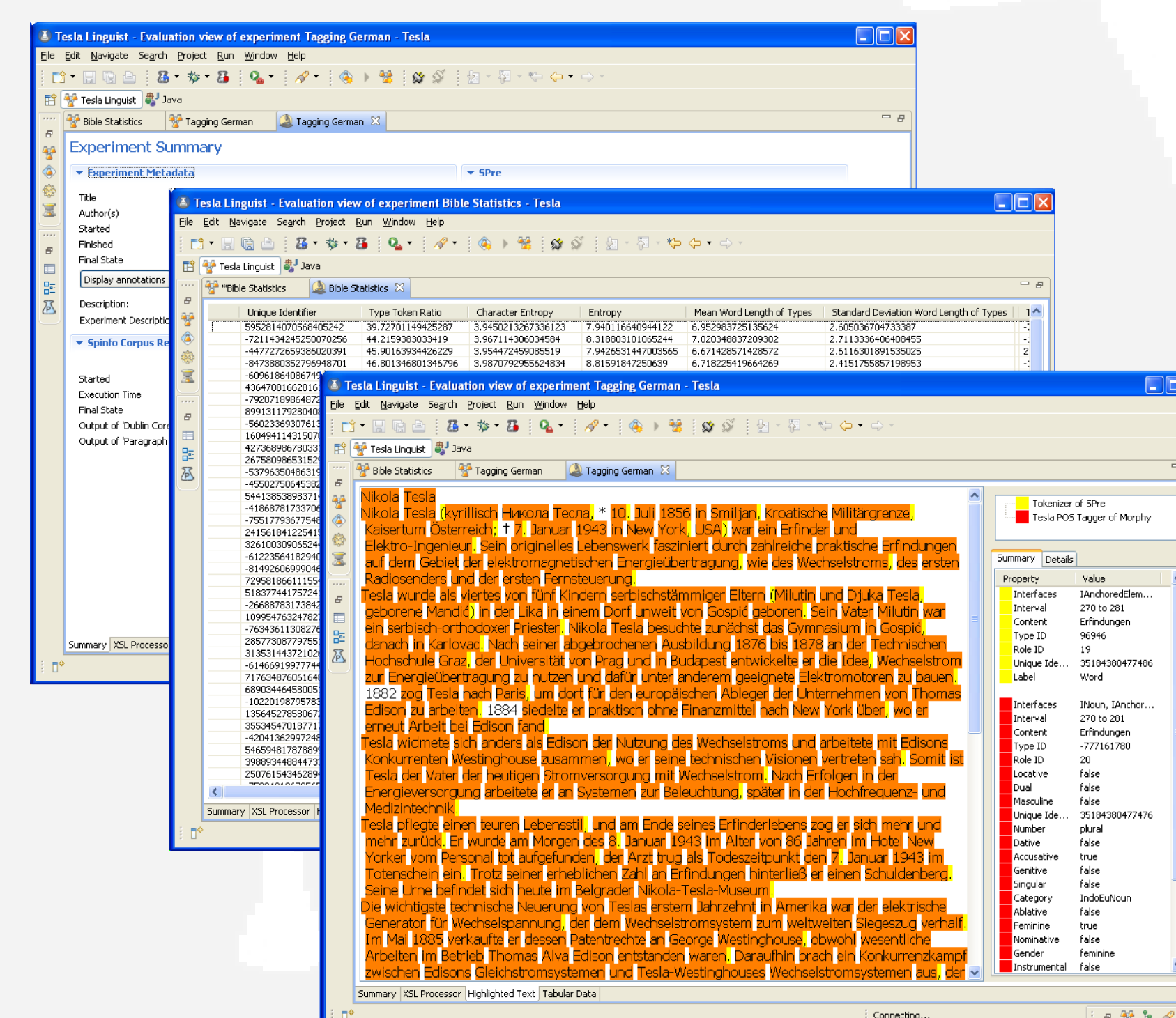
Die Arbeitsweise von Komponenten wird über Konfigurationsschnittstellen modifiziert. Komponenten interagieren miteinander, indem sie die Ergebnisse vorangehender Komponenten weiterverarbeiten und ihre Resultate nachfolgenden Komponenten zur Verfügung stellen. Die Interaktionsschnittstelle ist unrestrictiv gehalten,

die Aussagekraft von Komponenten ist damit sehr frei. Ein Rollensystem für Komponenten sichert die Interoperabilität.

Experimente (Selektion, Komponentenauswahl, Komponentenkonfiguration) werden vollständig dokumentiert und können über einen XML-Export weitergegeben werden.

Evaluation

Mit Ausführung des Experiments werden die von den Komponenten erzeugten Daten in möglicherweise unterschiedliche Datenbanken geschrieben. Über eine Aggregation im XML-Format können sie in eine einheitliche Form gebracht werden, die sich auch exportieren lässt. In diesem Zuge wird auch ein Experiment-Report generiert, in dem die wichtigsten Informationen übersichtlich dargestellt werden.

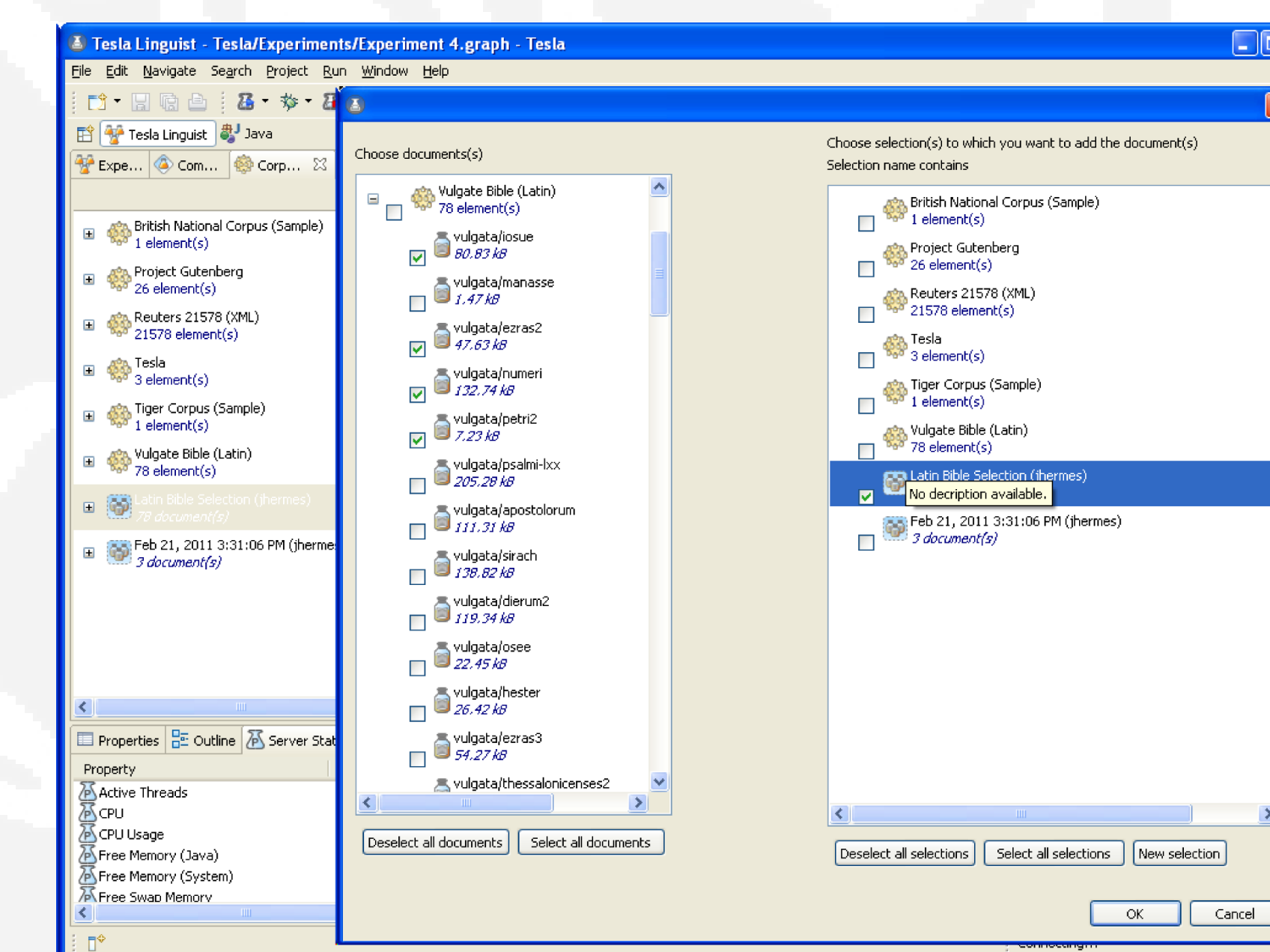


Die exportierten Daten lassen sich anschließend unterschiedlich visualisieren, z.B. als farblich unterlegter Text, als Tabelle oder als Graph.

Korpus-Management

Als Daten für die Verarbeitung können alle Arten von Textdaten genutzt werden. Beim Upload kann definiert werden, auf welche Art das Dokumentenformat interpretiert werden soll, indem ein

entsprechender Reader ausgewählt wird. Eine Reihe von allgemeinen Readern für die Interpretation gängiger (Txt, Html, Pdf, Xml, MS-Office, Odf) und spezifischerer Formate (BNC, Tiger) stehen bereits zur Verfügung. Weitere Reader können auf einfache Art und Weise implementiert werden. Tesla schreibt die Codierung der Daten nicht vor, die Reader können über 100 verschiedene Encodings verarbeiten.



Über eine Suchmaske lassen sich nutzerdefinierte Textselektionen zusammenstellen.

Download / Kontakt

Der Tesla-Client steht unter <http://tesla.spinfo.uni-koeln.de/> für Windows, Mac OS und Linux zur Verfügung - dort findet sich auch eine umfangreiche und aktualisierte Dokumentation zum Projekt Tesla.

Mit Download des Clients erhält man auch einen Tesla-Server sowie eine interne Datenbank, so dass man Tesla ohne die Einrichtung zusätzlicher Software nutzen kann. Zur Einführung stehen eine Reihe von Tutorials zur Verfügung.

Wenn Sie Tesla nutzen oder auch nur weiter verfolgen wollen, wenn Sie Fragen haben, so kontaktieren Sie uns gerne per Mail: sschwieb@spinfo.uni-koeln.de jhermes@spinfo.uni-koeln.de