



WebLicht: Technical Background, SOA, and REST

Thomas Zastrow & Marie Hinrichs
University of Tübingen

WebLicht - Architecture



- WebLicht is a service-oriented architecture (SOA) for accessing and processing text corpora
- WebLicht consists of the following components:
 - **Distributed services:** offering functionality (resources & tools) over the internet. Implemented as web services (approximately 120 currently)
 - **Repository:** stores metadata and technical information about the services
 - **Web 2.0 based user interface:** interacts with the user and combines services and information from the repository. Access still possible via scripts / programming code

WebLight - Architecture



Web Service
Workshop 2010

Stuttgart

```
aus dem Hause jagte, nachdem sie lange unetert in der welt umhergeritt, und da nitern eine  
Person, die schlingen und kisten sprach, bei sich aufnahmen wollte, ging sie in wilder heide  
jämlich zugrunde.</text>  
<tokens>  
<token ID="127">Es</token>  
<token ID="128">ag</token>  
<token ID="147">el</token>  
<token ID="125">tel</token>  
<token ID="167">H</token>  
<token ID="175">y</token>  
<token ID="18">die</token>  
<token ID="129">h</token>  
<token ID="118">ze</token>  
<token ID="111">f</token>  
<token ID="112">.</token>  
<token ID="113">f</token>  
<token ID="116">d</token>  
<token ID="117">g</token>  
<token ID="118">v</token>  
<token ID="119">g</token>  
<token ID="121">C</token>  
<token ID="120">v</token>  
<token ID="122">h</token>  
<token ID="123">h</token>  
<token ID="124">h</token>  
<token ID="125">M</token>  
<token ID="126">.</token>  
<token ID="127">d</token>  
<token ID="128">v</token>  
<token ID="129">so</token>
```

Standard-conformant
Text Corpus Encoding

Tübingen

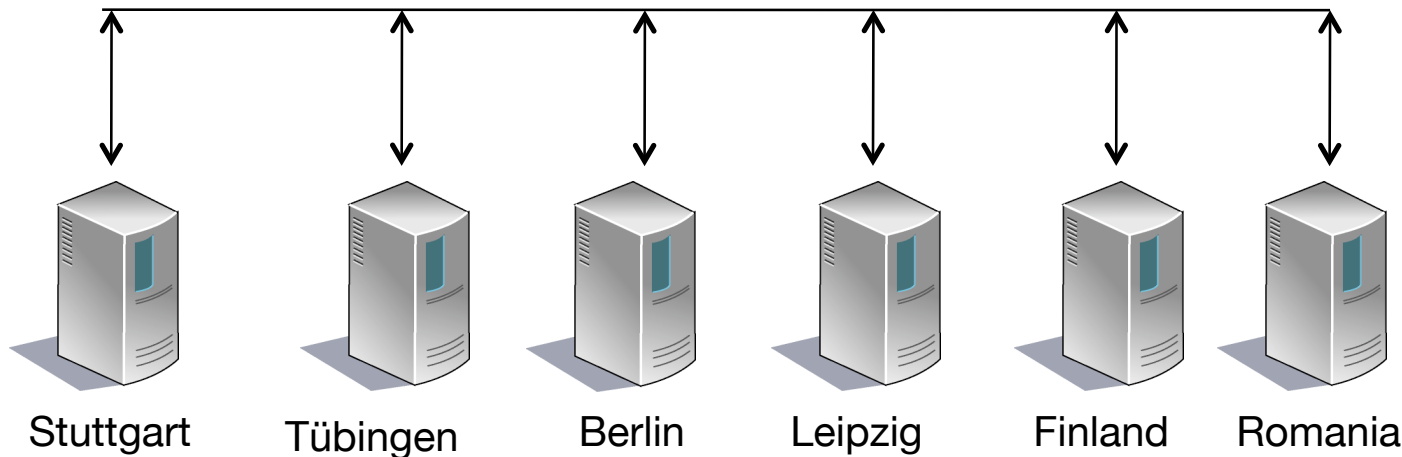


Web 2.0 Application for
Tool Chaining
and Execution

Leipzig



Repository



WebLicht – The Services



- Services are implemented as REST style webservice
- HTTPs POST method is used to send data from the UI to the services
- As client, anything which is able to use the HTTP protocol, can be used:
 - Browser
 - Commandline tools (wget, curl)
 - Programming Languages
- Anyone can implement their own interface to WebLicht

WebLicht - Features



- With RESTstyle webservices, everyone can implement a web service for WebLicht (short tutorials for Java and Perl)
- The SOA infrastructure is independent of programming languages or operating systems
- The chaining algorithm is independent of the used dataformat
- Form a legal point of view, the web services are still located in the institute where they were created



- Building a webservice for WebLicht consists of the following steps:
 - Create a RESTstyle webservice around the tool as wrapper
 - Make in- and output compatible with WebLicht's TCF format
 - Register the service in the repository



- *You can find online tutorials for Java and for Perl at:*
 - <http://weblicht.sfs.uni-tuebingen.de/englisch/weblicht.shtml>

WebLicht – The Repository



- Implemented at the ASV Leipzig

WebLicht – The User Interface



Web Service
Workshop 2010

- Web 2.0 Application for Tool Chaining and Execution
- Implemented at the SfS Tübingen
- Java application, deployed in Apache Tomcat
- Allows the user to:
 - upload a text (plain text, MS Word, RTF or PDF files)
 - construct a text from corpora in Leipzig, query corpora from the BBAW and the IDS
 - use some hardwired example texts
 - Build a chain of linguistic tools
 - Executes the tool chain with the uploaded text and presents the results
- During the chaining process, it queries the repository for available services

Test Case: Gutenberg Corpus



- On the basis of these structure, a part of the free available Gutenberg Project was annotated in Tübingen*
- Ca. 20.000 texts from 800 authors
- Runtime: ca. 3.5 weeks
- Result:
 - 217 million tokens (words), 533 million constituents, 110 GB data

* Using the BW Grid, not the web services: 3.5 weeks runtime

WebLicht Infrastructure



- Arguments for the WebLicht Infrastructure:
 - Fulfilling the needs of the linguistic community
 - Simplicity
 - Flexibility



Thank you for your attention