



WebLicht – Web-Based Linguistic Chaining Tool

Prof. Dr. Erhard Hinrichs
University of Tübingen

WebLicht - Motivation



- Many linguistic resources (corpora, dictionaries, ...) and tools (tokenizer, tagger, parser, ...) are available
- Most of them are implemented to run on local machines. This can be inconvenient and error-prone
- Requirements: go beyond “do-it-yourself” and “download-first” strategies
- The CLARIN solution: *Make tools and resources available as webservice*

WebLicht - Motivation



- For some kinds of LRT, its easy to put them online (make resources downloadable, offer search engines etc.)
- For other kinds, more effort is necessary (limiting access to resources, how to make tools online usable)
- Solution: a *Service Oriented Architecure (SOA)*
 - → The end user needs just a browser: no more installation, configuration etc. of software is necessary

The History of WebLicht



- Development started in November 2008 (Leipzig and Tübingen)
- Shortly after the first steps were done, more participants from the D-SPIN joined the project (Stuttgart, BBAW, and IDS)
- Presentation of a first version in February 2009 in Oxford
- International partners joined WebLicht (Helsinki, RACAI)
- Dual usage of web services in WebLicht and C4

- The processing format (TCF) evolved over time to version 0.4
- A grid-based version of several web services were recently used to build up automatically annotated text corpus (217 million words, 533 million constituents)

WebLicht - Architecture



Web Service
Workshop 2010

Stuttgart

aus der haue jagte, nachdem sie lange unet in der weit umgerirt, und da niemo eine
Person, die Schlingen und Kröten sprach, bei sich aufnehmen wollte, ging sie in wilden wäld
jämlich. zugrunde. r/leis-
-tokens
-tokens [id="42">Es-/tokens
-tokens [id="43">Wern-/tokens
-tokens [id="44">Hind-/tokens
-tokens [id="45">Kine-/tokens
-tokens [id="46">Witwe-/tokens
-tokens [id="47">v-/tokens
-tokens [id="48">de-/tokens
-tokens [id="49">Hette-/tokens
-tokens [id="50">Wen-/tokens
-tokens [id="51">Tochte-/tokens
-tokens [id="52">v-/tokens
-tokens [id="53">Oke-/tokens
-tokens [id="54">diere-/tokens
-tokens [id="55">chl-/tokens
-tokens [id="56">ov-/tokens
-tokens [id="57">Gescht-/tokens
-tokens [id="58">vnd-/tokens
-tokens [id="59">Charite-/tokens
-tokens [id="60">vnd-/tokens
-tokens [id="61">vnd-/tokens
-tokens [id="62">vnd-/tokens
-tokens [id="63">vnd-/tokens
-tokens [id="64">vnd-/tokens
-tokens [id="65">vnd-/tokens
-tokens [id="66">vnd-/tokens
-tokens [id="67">vnd-/tokens
-tokens [id="68">vnd-/tokens
-tokens [id="69">vnd-/tokens
-tokens [id="70">vnd-/tokens
-tokens [id="71">vnd-/tokens
-tokens [id="72">vnd-/tokens
-tokens [id="73">vnd-/tokens
-tokens [id="74">vnd-/tokens
-tokens [id="75">vnd-/tokens
-tokens [id="76">vnd-/tokens
-tokens [id="77">vnd-/tokens
-tokens [id="78">vnd-/tokens
-tokens [id="79">vnd-/tokens
-tokens [id="80">vnd-/tokens
-tokens [id="81">vnd-/tokens
-tokens [id="82">vnd-/tokens
-tokens [id="83">vnd-/tokens
-tokens [id="84">vnd-/tokens
-tokens [id="85">vnd-/tokens
-tokens [id="86">vnd-/tokens
-tokens [id="87">vnd-/tokens
-tokens [id="88">vnd-/tokens
-tokens [id="89">vnd-/tokens
-tokens [id="90">vnd-/tokens
-tokens [id="91">vnd-/tokens
-tokens [id="92">vnd-/tokens
-tokens [id="93">vnd-/tokens
-tokens [id="94">vnd-/tokens
-tokens [id="95">vnd-/tokens
-tokens [id="96">vnd-/tokens
-tokens [id="97">vnd-/tokens
-tokens [id="98">vnd-/tokens
-tokens [id="99">vnd-/tokens
-tokens [id="100">vnd-/tokens

Standard-conformant
Text Corpus Encoding

Tübingen

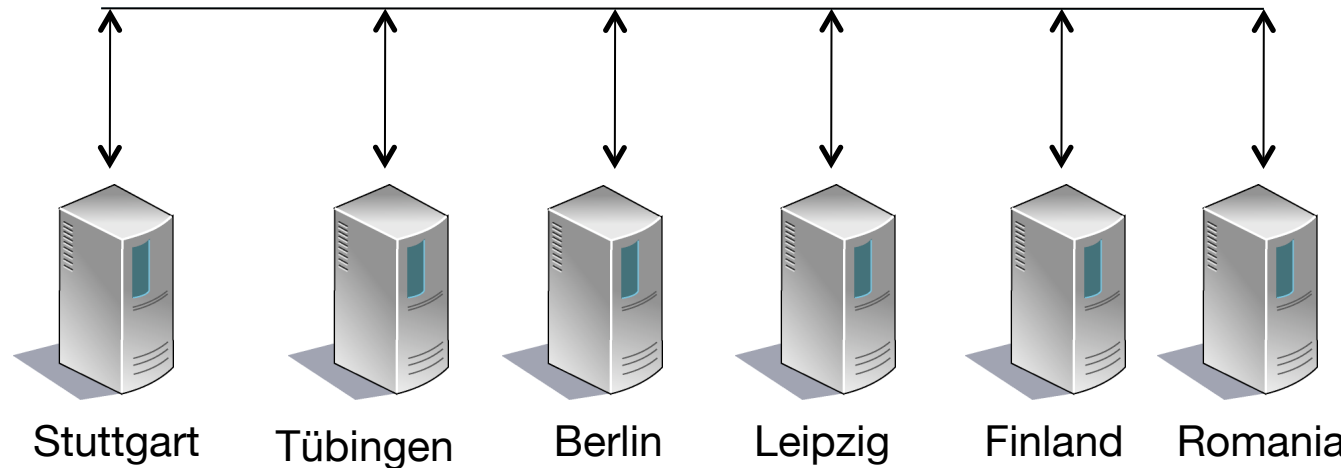


Web 2.0 Application for
Tool Chaining
and Execution

Leipzig



Repository



Available Services



Web Service
Workshop 2010

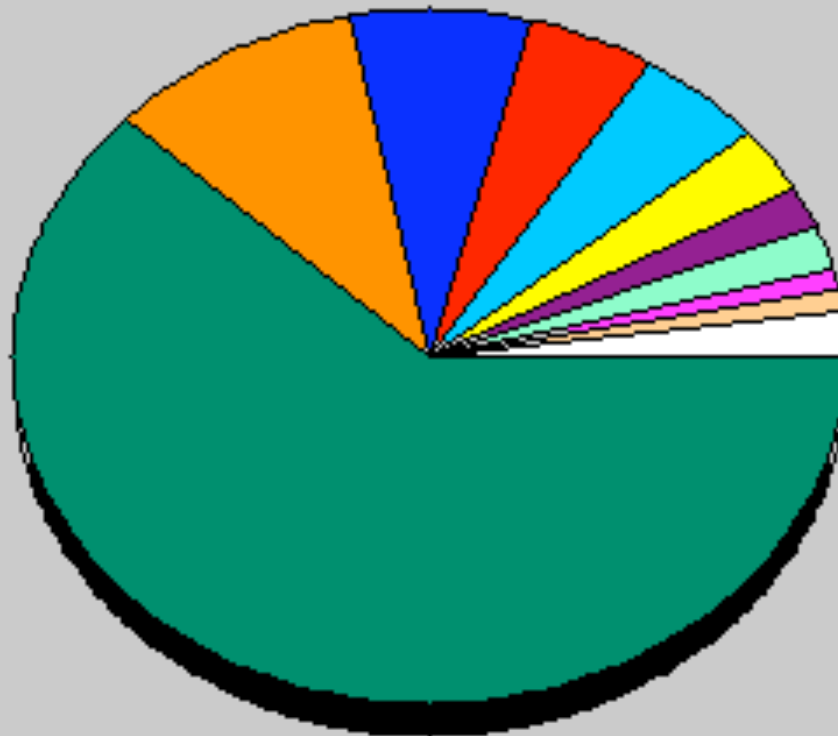
- Approximately 120 web services are currently available
 - 3 countries (Germany, Finland, Romania)
 - 8 partners
 - 10 languages (German, English, Finnish, Italian, French, Spanish, Czech, Hungarian, Romanian, Slovenian)
 - More than 10 different tools (converter, tokenizer, part-of-speech tagger, lemmatizer, parser, etc.)

WebLicht Access Statistics



Web Service
Workshop 2010

Usage by Location for October 2010

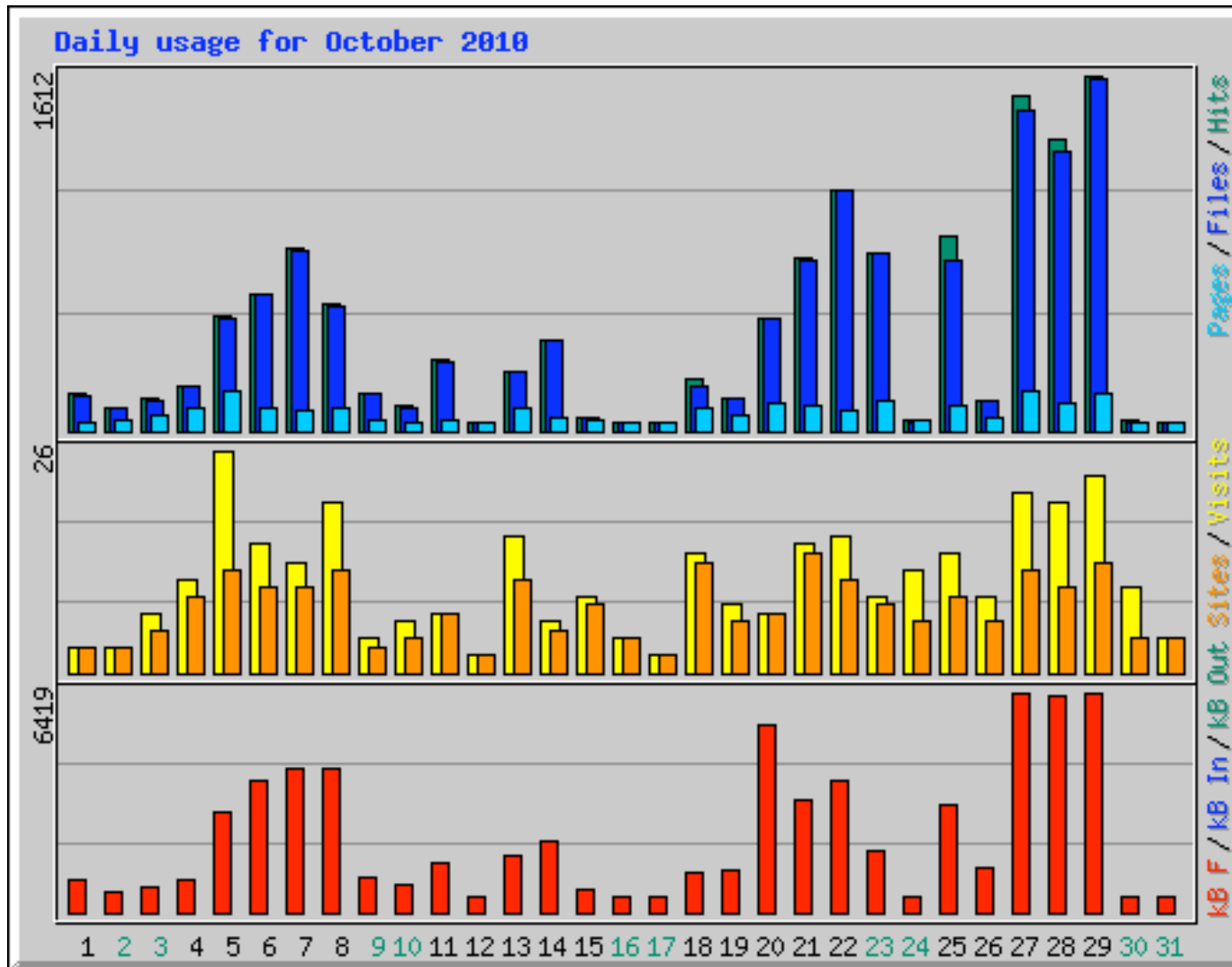


Germany	(62%)
Austria	(10%)
Latvia	(7%)
Spain	(5%)
France	(5%)
Bulgaria	(3%)
Portugal	(2%)
Unresolved/Unknown	(2%)
Finland	(1%)
Italy	(1%)
Other	(2%)

WebLicht Access Statistics



Web Service
Workshop 2010



WebLicht Access Statistics



Web Service
Workshop 2010

