short Guide

Common Language Resources and Technology Infrastructure

CLARIN

February 2010

Relevance

CLARIN community

for all communities

Web Services Interoperability

What is it?

Interoperability of webservices refers to the seamless flow of data between web-based applications and services. To be interoperable, these webservices must first agree on protocols defining the interaction between the services (WSDL/SOAP, REST, XML-RPC). They must also use a shared and standardized data exchange format, which is preferably based on widely accepted formats already in use (UTF-8, XML).

What is it for?

Interoperability ensures that linguistic resources and tools can be combined into a common processing pipeline. Web services encapsulate these tools and combine them in a common service-oriented architecture. Metadata for the individual webservices are stored in a centralized repository. By the use of Web 2.0 techniques, the user interface allows the user to choose between various processing pipelines.

Who can use it?

Web services and web service chains can be used by all researchers in the humanities and social sciences who need to automatically or semi-automatically annotate, analyze and query their data sets. Easy-to-use interfaces ensure that the technical expertise required to configure and execute customized annotation and query chains is kept at a minimum.

Web services and web service chains will typically draw upon corpora, tools, and other services from different CLARIN member institutions or affiliated partners. As a result a broad set of services will be available to users without the need to download and customize such services on their own local machines.

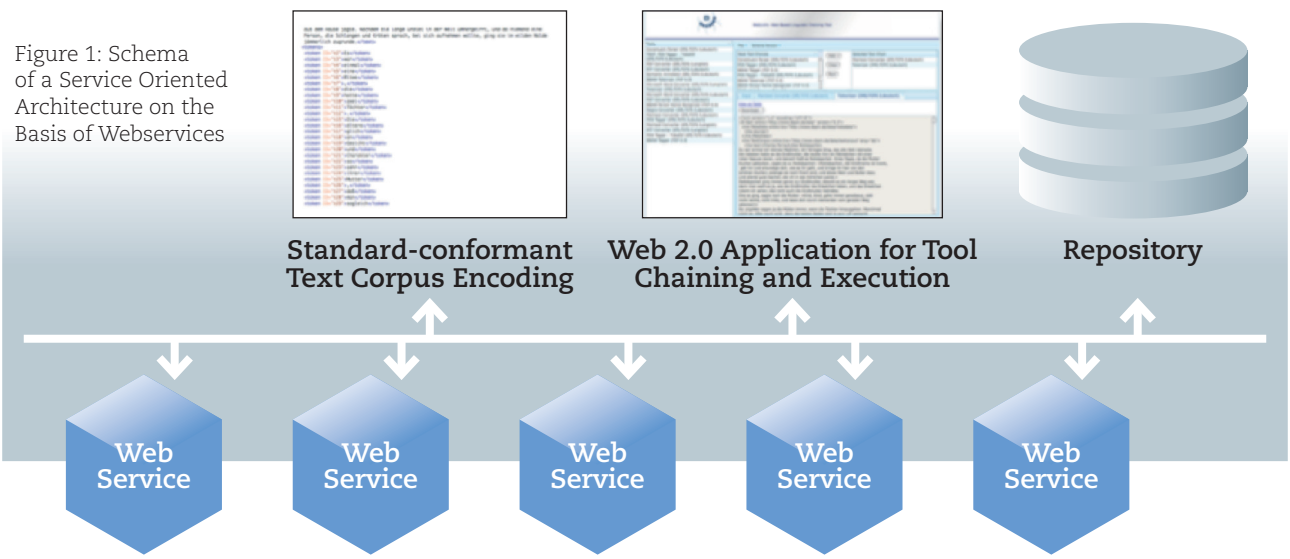

Figure 1: Schema of a Service Oriented Architecture on the Basis of Webservices

CLARIN

Short Guide

When can it be used?

CLARIN member institutions have already produced a number of web services and web services chains that are currently available. The following list summarizes these initiatives (for detailed information we refer to the deliverables):

- WebLicht: Web Based Linguistic Chaining Tool. Contact: Seminar für Sprachwissenschaft, Universität Tübingen, Germany; Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Germany; Abteilung Automatische Sprachverarbeitung, Universität Leipzig, Germany; Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, Germany. Access/Documentation: http://www.d-spin.org
- GATE: General Architecture for Text Engineering. Contact: GATE group, Department of Computer Science, University of Sheffield, UK. Access/Documentation: http://gate.ac.uk/science.html
- IULA Web Services. Contact: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona, Spain. Access/Documentation: http://gilmere.upf.edu/WS/
- ILSP Text Processing Chain (ILSP TPC). Contact: Institute for Language and Speech Processing, Athens, Greece.
- RACAI Services. Contact: Research Institute for Artificial Intelligence, Romanian Academy of Sciences, Bucharest, Romania. Access/Documentation: http://www.racai.ro/webservices/Default.aspx
- WS-LexicalPlatform. Contact: Consiglio Nazionale delle Ricerche, Istituto di Linguistica Computazionale, Pisa, Italy. Access/Documentation: http://www.clarin-it.it/Simple/SimpleGUI.html
- LXService: A web service for language technology of Portuguese. Contact: University of Lisbon, Department of Informatics, Natural Language and Speech Group, Lisbon, Portugal. Access/Documentation: http://nlx.di.fc.ul.pt/; to get access contact: sara.silveira@di.fc.ul.pt
- WROCUT/ICS PAS services. Contact: Institute of Informatics, Wrocław University of Technology, Wrocław, Poland; Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland. Access/Documentation: http://plwordnet.pwr.wroc.pl/clarin/ws/

CLARIN is open for all interested people to add their services into the CLARIN web services framework. Currently, there is a tutorial for developing WebLicht services using Java, available for download at: http://weblicht.sfs.uni-tuebingen.de/englisch/weblichttutorial.shtml. More tutorials are being written for integrating services into WebLicht using other programming languages and servlet containers. These will be made available as soon as possible.

How does it work?

Web services can easily be invoked by a web application that is accessible via any web browser (user interface, see figure 1). This avoids many problems that can occur when downloading and installing desktop tools.

Who is responsible?

In CLARIN the whole domain of web services is tackled from two sides: (1) This short guide describes the essentials of a Service Oriented Architecture for linguistic resources and tools, as well as the concrete projects that have been carried out in CLARIN so far and (2) there is another short-guide that summarizes the infrastructure aspects of webservices.

Whom to contact?

For inquiries regarding specific web services listed above, please contact the respective institutions directly. For the CLARIN infrastructure initiative the official web-site gives most recent information: http://www.clarin.eu

For all SOA infrastructure aspects you can contact:
Núria Bel or Marc Kemps-Snijders: nuria.bel@upf.edu, marc.kemps-snijders@mpi.nl

For all SOA interoperability aspects you can contact:
Erhard and Marie Hinrichs or Thomas Zastrow Email: weblicht@d-spin.org

Where to find more information?

The official CLARIN web-site is the source of all information:
CLARIN: http://www.clarin.eu

CLARIN Web Services Requirements Document:
http://www.clarin.eu/deliverables

CLARIN Service Oriented Infrastructure:
http://www.clarin.eu/documents/short-guides

Responsible for the content:
Erhard Hinrichs, Marie Hinrichs, Iris Vogel, Thomas Zastrow
Email: weblicht@d-spin.org