Service Oriented Infrastructure

What is it?

Many experts in universities, research institutions and industry are developing smart language technology processing tools for text, audio and video material. The current paradigm of using these tools for research work is mostly characterized by the "download first principle", i.e. all resources and tools that are required are first downloaded and installed on the own notebook or department server before to really start the actual research work. A Service Oriented Infrastructure is meant to avoid this overload by making use of the advanced capabilities of the Internet and deploying these tools as web services each of which is being executed on some server or in a cloud or grid of strong computer clusters.

Besides, a Service Oriented Infrastructure offers a framework that supports the combination and still distributed execution of such processing components. The idea is that users simply see functional descriptions of these components in metadata registries, select and combine them with a so-called workflow tool and then execute the chain of processes without technical knowledge of technicalities not even knowing on which computers they are actually being executed.

What is it for?

From the user perspective, a Service Oriented Infrastructure reduces the complexities of using advanced technological analysis and exploitation tools. The user is only confronted with the functional aspects of the components and a web client with a nice user interface.

Who can use it?

All researchers that are members of CLARIN identity federations should be able to use such web-accessible components, to combine them to achieve new, more advanced and tuned to their necessities, virtual applications and execute them in the Internet.

When can it be used?

The area of web services and service oriented infrastructures are part of the future internet and it is still in a continuous evolutionary stage, i.e. not all aspects have been settled yet. However, first excellent examples of such distributed processing scenarios have been launched and in CLARIN a number of teams have already created web services and simple to use tools to combine them (see Short Guide on Web Service Interoperability).

The area of service oriented infrastructures is now a hot topic in the IT domain and many computer science teams are working on generic solutions which will be adopted in CLARIN as well.

CLARIN
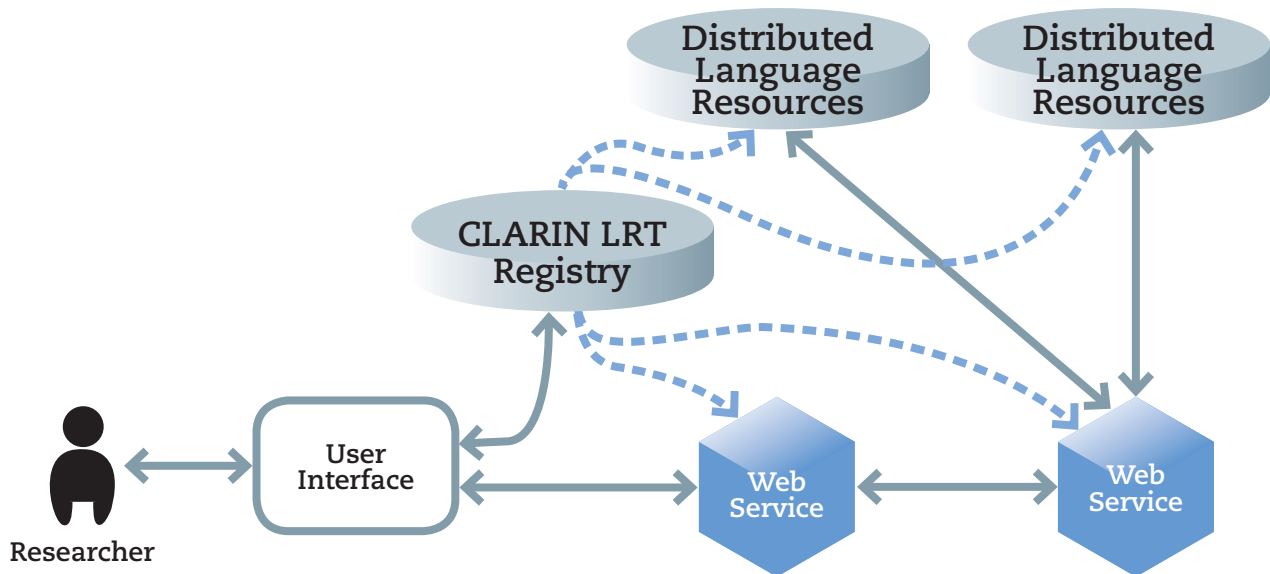
How does it work?

Analysis tools of interest for the analysis and exploitation of text, audio and video materials need to be deployed as accessible web services.

Each of these services (data and tools) needs to be described by standardized metadata which for CLARIN is CMDI (component metadata initiative). To achieve interoperability the metadata descriptions defined by the components used need to make use of those elements that have already been registered in ISOcat and therefore have a declared semantics. If resources and services are described using the same semantics it will allow us to automatically check which tools can operate on a certain data resource which also can be intermediary files generated during a particular process.

Also the programming interfaces of each of these services must be formally specified so that they can be invoked automatically. Two standards are currently being used: WSDL/SOAP and REST. Many prefer to use REST due to its simplicity, however it is much less standardized and does not enforce rigidity, i.e. in a large open market place scenario of web services compatibility problems may be expected.

Summarizing it can be said that the functioning of an open web services scenario will widely depend on strict adherence to standards and the availability of converters and wrappers to include legacy data and tools. CLARIN currently is building and registering all these components.

Who is responsible?

In CLARIN the whole domain of web services is tackled from two sides: (1) this ShortGuide represents the work about infrastructure aspects and (2) there will be another Shortguide that represents the concrete projects that have been carried out so far.

CLARIN will push this work ahead, but will interact with IT driven initiatives about generic solutions.

Whom to contact?

For the CLARIN infrastructure initiative the official web-site gives most recent information: http://www.clarin.eu

For all aspects mentioned in this note you can contact: Nuria Bel (UPF Barcelona) or Marc Kemps-Snijders (MPI): mailto: nuria.bel@upf.edu; marc.kemps-snijders@mpi.nl

For all matters of linguistic applications you can contact Thomas Zastrow (U Tübingen): thomas.zastrow@uni-tuebingen.de

Where to find more information?

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

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

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

CLARIN Metadata Infrastructure Shortguide:
http://www.clarin.eu/documents/short-guides

ISOcat Information: http://www.isocat.org