How can I integrate my tools and services with the CLARIN infrastructure

Thorsten Trippel
• easy, single sign-on access to language data
• distributed advanced services for research
• supporting archiving and sustainability for language data
• easy, single sign-on access to language data
• distributed advanced services for research
• supporting archiving and sustainability for language data
The scholar's DREAM

• Brilliant idea
  – Using existing data D
  – Perform complex analysis A
  – Gets results R
  – Required time \( t \sim 0 \)
  – Research Miracle:
    \[ A(D)_t = R \]
  – Knows of analysis tool
• Performs the Analysis
• Publishes a paper
The scholar's Nightmare

- Brilliant idea
  - Using existing data $D$
  - Perform complex analysis $A$
  - Gets results $R$
  - Required time $t \sim 0$
  - Research Miracle: $A(D)_t = R$

- Knows of analysis tool
  - Cannot find it
  - Cannot access it
  - Cannot install it
  - Cannot contact developer
  - Cannot run it
  - Tool crashes with her/his data

- No Analysis
- No paper
- Frustration
• Seeks for an industry job
Challenges for Running the Tools

• Installation
  – Operating system
  – System environment
  – Configuration (!)
  – Maintenance/Security (!)

• Computing power
  – Demands in working memory
  – Long time running
  – Network communication

• Licenses
More Challenges

- Preprocessing by other tools
  - Requirements before using the tools on data
  - Preanalysis
- Interface between different tools
  - Communication
  - Data format
- Availability of tools
  - Download options
  - Contact developers
Bob Hope

From Wikipedia, the free encyclopedia

For other uses, see Bob Hope (disambiguation).

Bob Hope, KBE, KCSG, KSS, (born Leslie Townes Hope; May 29, 1903 – July 27, 2003), was an English-born American comedian, vaudevillian, actor, singer, dancer, author, and athlete who appeared on Broadway, in vaudeville, movies, television, and on the radio. He was noted for his numerous United Service Organizations (USO) shows entertaining American military personnel—he made 57 tours for the USO between 1941 and 1991. Throughout his long career, he was honored for this work. In 1997, the U.S. Congress declared him the "first and only honorary veteran of the U.S. armed forces."[2]
• service orchestration and execution environment for incremental processing
• distributed services for data processing
• web application offering user-friendly graphical interface
  – for building chains of services and
  – executing chains of services
• Tokenizer
• Parser
• Lemmatizer
• Format converter
• Tagger
• Named entity recognition
• Morphological analyser
• Visualization tools
• ...
• More Tools
  – More languages
  – More services
  – More analysis
  – More distribution
• ...

TOOLS WANTED!

DH2014 CLARIN
Requirements for providing a Tool as a service

• Available on a server
• RESTful interface
• Metadata description of tool in specific format
• Distribution of metadata description
• Registration of service infrastructure as a centre
Availability on a server

- Robust service
- Enough bandwidth
- Enough storage space
- Security
- Error handling
Availability on a server

- Robust service
- Enough bandwidth
- Enough storage space
- Security
- Error handling
• Communications of services via http
• Wrapper possible for
  – Import
  – Export
• Support data format:
  – TCF
  – Various formats for input conversion
• Provided in CMDI
  – WebLichtWebService

Profile (Profile ID: clarin.eu:cr1:p_1320657629644)

– Specifying
  • Input
  • Output
  • Parameters
• Provided via OAI-PMH
  – Open Archive Protocol for Metadata Harvesting
  – Based on http
• Data format: CMDI
• CLARIN Center Registry
• No self registration possible
• OAI-PMH server needs to be provided
  – May also serve other (CMDI) metadata records
  – Harvested regularly
• Contact via Email the CLARIN developers
• Short time after harvesting of metadata
• Metadata harvested daily – only
• Change of services controlled by service provider
Available documentation and tools

• Tutorials for
  – TCF
  – Creating a webservice wrapper (Java)
  – CMDI

• Documentation for
  – Setting up OAI-PMH server
  – CMDI

• Tools
  – Editing CMDI
  – I/O library for TCF
"But I don't have a server...."

- If....
  - The service is ready for use by the research community
  - Developers want to publish it
  - Developers provide all code and documentation
  - Providers contact CLARIN

- Then....
  - CLARIN might find someone willing to host the service
  - Depending on the requirements and amount of work involved to install and configure (and debug) the service
  - Provided enough rights are granted
• 5 "simple" steps:
  – Creating service (based on TCF)
  – Making it restful
  – Describing it in CMDI
  – Publishing the CMDI via OAI-PMH
  – Registering with the CLARIN centre registry
• Increase visibility and use of your service and...
Make the scholar's dream come true