ELG Tutorial 2020: How to ingest your Language Technology service or data set into the European Language Grid

09:00-09:40 Part 1 – How to use the European Language Grid (ELG)
- Short overview: ELG – introduction, overview, history, concept
- Exploring the ELG catalogue (hands-on):
    - Search for Language Technology services (LT) and resources
    - Run different types of ELG services in the grid platform
      (e.g. Information Extraction, TTS, ASR, Machine Translation)
    - Browse profile pages from companies, academic organisations and researchers

09:40-10:30 Part 2 – How to use the European Language Grid
- Introduction: ELG as the "yellow pages" of the European LT community
- Presentation of ELG's metadata schema
- Use ELG's metadata editor for creating/registering entities, e.g., for datasets, lexical conceptual resources, LT services, organizations, SMEs

10:30-11:00 Coffee Break

11:00-12:00 Part 3 – How to integrate an LT service/application into the ELG platform
- Introduction to the ELG LT Service integration API
- Follow along "Hello World" example (hands-on)
    - Clone a Gitlab repo that contains a simple LT tool
    - Add code that makes it compatible with the ELG LT Service API
    - Dockerization of a service, e.g., create the respective Dockerfile
    - Test the Docker image locally on your laptop
    - Register the LT service in the ELG platform, e.g., create the required configuration files and a metadata record
- Brief presentation of custom helper libraries/tools provided by ELG that facilitate the process of creating an ELG compatible LT service (e.g., Spring Boot-based for java)

12:00-13:00 Part 4 – Install the ELG platform locally
- Present the ELG platform architecture
    - Main software components
    - Kubernetes/LT Service auto-scaling etc.
- Install the ELG platform locally
    - Git clone a repository that contains the installation scripts
    - Configuration options, which scripts should be used, how to check if it works
    - Load test/seed data from a database dump to populate the catalogue

Tutorial Description

This tutorial is organised under the umbrella of the EU project European Language Grid (ELG; 2019-2022). ELG develops a scalable cloud platform, providing, in an easy-to-integrate way, access to hundreds of commercial and non-commercial Language Technologies (LTs) for all European languages, including running tools and services as well as data sets and resources. At META-FORUM 2019 (Oct. 2019), a [first prototype](#) version of the ELG was demonstrated. Currently, about 100 functional services are up and running in the ELG; in addition, more than 200 LT resources in over 50 languages are available. By May 2020, more services, data sets and features such as user management as well as a more detailed catalogue will be available. The ELG aims at listing all relevant stakeholders, from technology development to research centres, from SMEs to large enterprises. All stakeholders can register themselves in the ELG – the "Yellow Pages" of European Language Technology. Through the ELG, companies, academic organisations and individual researchers can gain visibility, provide services, data sets and resources and make use of those services, data sets and resources made available through the platform by others. The online catalogue is browsable and searchable: users can filter and search for domains, sectors, regions, countries, languages, service types, data sets and more. The ELG is the first large-scale LT platform aiming at production use, based on containerisation technology (i.e., Docker) and the Kubernetes container orchestration platform. The ELG enables the commercial and non-commercial European LT community to deposit and upload their technologies and data sets into the platform and to deploy them through the grid. This tutorial explains how to integrate an application (= LT service) into the ELG platform and how to set up a profile page. For more details: https://www.european-language-grid.eu

Motivation and Topics of Interest

Language barriers affect cross-lingual communication and the free flow of knowledge and thought. Monolingual and multilingual Language Technologies can help overcome these omnipresent language barriers and significantly improve trade, administration, politics, culture, intercultural communication and understanding. The vision of technology-enabled multilingualism applies both to the whole planet and also to several multilingual regions – ELG's focus is Europe with its 24 official EU Member State languages, 60 additional unofficial or minority languages as well as languages of immigrants and trade partners. To date, a general LT platform that attempts to bring together all offerings (in terms of functional processing services and data sets) does not exist. Instead, there are multiple activities that typically have a specific focus on a concrete domain. Language resources and data sets are distributed over different heterogeneous repositories and organisations. A similar situation exists for basic and also more sophisticated processing and generation tools including ASR and TTS, tokenisers, NER, morphological analysers, lexical resources, ontologies, sentence and word aligners, chunkers, parsers, generators, translation systems, IR and IE systems, QA systems, dialogue systems, sentiment and emotion analysers, multi-modal systems processing speech, text, images and video, etc. Hundreds, if not thousands, of functional services exist but are extremely difficult to find, which is not only highly inefficient but also frustrating both for the user and also

for the provider of the respective service. This situation is true both for the commercial and also for the academic area. Most European Language Technology developing companies are SMEs that operate in highly focused domains and/or geographic niches. ELG brings together the expertise of nine project partners from academia and industry. ELG is a project organised *by* the European LT community *for* the European LT community.

Tutorial Speakers

- Georg Rehm, DFKI, Germany, georg.rehm@dfki.de
- Dimitris Galanis, R.C. "Athena", ILSP, Greece, galanisd@athenarc.gr
- Ian Roberts, The University of Sheffield, UK, i.roberts@sheffield.ac.uk
- Andrejs Vasiljevs, Tilde, Latvia, Andrejs@tilde.lv

Additional Information

The tutorial is relevant for participants from industry or research/academia who develop Language Technology services in the form of application-ready prototypes (including written and spoken). Accordingly, general familiarity with NLP, NLG or speech applications, first experience with containerization-related technology (Docker, Kubernetes, Helm etc.) as well as general programming and software development experience is helpful.

Participants are requested to bring their own laptops, ideally with Docker and Python installed. If available but not necessary, participants can bring their own LT services which can then be made available through the ELG platform.