Chapter 20
Lingsoft Solutions as Distributable Containers

Sebastian Andersson and Michael Stormbom

Abstract Lingsoft is one of the leading language technology and language service providers in the Nordic countries. In the Lingsoft Solutions as Distributable Containers (LSDISCO) project, we packaged our language technology tools for distribution as containerised services via the European Language Grid (ELG). As a result, Lingsoft's speech recognition, machine translation, proofing, and morphological analysis was made available to users of the European Language Grid. The services primarily cover Finnish (general and healthcare domain), Swedish (also Finland Swedish), Danish, Norwegian bokmål and nynorsk, and English. The distribution as containerised services is a straightforward way of making our tools available and updated on ELG and we intend to continue to update our service offerings on ELG with new tools and languages as we develop them.

1 Overview and Objectives of the Pilot Project

Lingsoft is one of the leading providers of language technology solutions in the Nordic countries and one of the 100 largest language service providers in the world. The tools and models that Lingsoft contributed to ELG via the Lingsoft Solutions as Distributable Containers (LSDISCO) project already existed and in most cases they were already actively used in production by Lingsoft or our customers. The goal of the LSDISCO project was to make those tools and models available as ELG-compatible services for ELG users (Rehm et al. 2021). This included four types of services:

- *Speech recognition*, with the supported languages being Finnish (general and healthcare domain), Swedish and Norwegian bokmål
- *Machine translation*, for language pairs involving Finnish, Swedish, and English in any combination, as well as both directions of Finnish – German

Sebastian Andersson · Michael Stormbom
Lingsoft, Finland, sebastian.andersson@lingsoft.fi, michael.stormbom@lingsoft.fi

G. Rehm (ed.), *European Language Grid*, Cognitive Technologies,
https://doi.org/10.1007/978-3-031-17258-8_20

- *Proofing, entailing spelling and grammar error detection* for Finnish, Swedish, Danish, Norwegian bokmål, and spelling for Norwegian Nynorsk and English
- *Text analysis*, entailing morphological analysis (lemmatization and morphology) and named entity recognition (NER) for Finnish, Swedish, Danish, Norwegian Bokmål, Nynorsk, and English

The end result of the project was a set of high quality NLP tools for the Nordic languages available through ELG, for both commercial and non-commercial use, allowing companies and public organisations throughout Europe to efficiently incorporate Nordic language support in their solutions and services.

2 Methodology

The four types of tools and services in scope for the LSDISCO project – speech recognition, machine translation, proofing and text analysis – have been originally developed at Lingsoft in different periods in the company and software development history and for different primary use cases. The least common denominator was a need for refactoring the tools and service architecture to comply with the ELG requirements. Especially the machine translation tools needed conversion from an internally used tool to enable also external distribution as a service via ELG.

The LSDISCO project was divided into three phases per requirements in the ELG call outline: 1. Experiment; 2. Integration; 3. Dissemination. The Experiment phase consisted of refactoring Lingsoft's tools and architecture to comply with ELG's integration requirements. This phase also included enabling a licensing mechanism for the services and creation or upgrade of the terms of service documentation. For the Integration phase, we selected the option to integrate our services to ELG via a proxy container, as this was the most practical option for us requiring the least amount of additional maintenance. This means that all calls to the ELG service are forwarded to and processed by Lingsoft's back end. Upgrades to the services in Lingsoft's back end per our normal release update cycle, e. g., model improvements, are then immediately available also in ELG. The dissemination phase consisted of advertising Lingsoft's services and the ELG platform on Lingsoft's website and in suitable forums such as conferences and trade fairs.

3 Implementation

Lingsoft's proofing, text analysis and speech recognition services were already to a large extent ready for ELG integration. The improvements made for those largely followed the existing development roadmap. The biggest implementation and refactoring effort in the LSDISCO project was for enabling serving Lingsoft's neural machine translation (NMT) to external users, in this case ELG. The NMT engine and

models were migrated from a solution serving "only" Lingsoft's own translation production to the same Software as a Service infrastructure as our speech recognition. This gave us a scalable back end and the possibility to provide user credentials for NMT usage, thus making important improvements to commercialising Lingsoft's machine translation and serving also external organisations.

To integrate our services with ELG, we implemented the Lingsoft ELG adapter. The Lingsoft ELG adapter is an API proxy container, illustrated in Figure 1. It exposes the ELG platform's internal LT Service API specification compatible endpoints and acts as a proxy to the Lingsoft APIs:

ASR API Lingsoft Speech Recognition API
NMT API Lingsoft Machine Translation API
LMC API Lingsoft Language Management Central API (text analysis)

In the proxy container, we implemented the conversion between the ELG and the Lingsoft API specifications. The proxy container also includes the mechanism for forwarding authentication via ELG for Lingsoft's back end service.

The Lingsoft ELG Adapter was packaged into a Docker image and submitted to DockerHub. Lingsoft then filled in the ELG XML metadata specifications for Lingsoft's services on the ELG platform, and the ELG technical team could proceed with the actual integration. The DockerHub image of the Lingsoft ELG Adapter was created for ELG, but it can be deployed by other organisations in a Docker environment and integrated with the organisation's own solutions. All that another

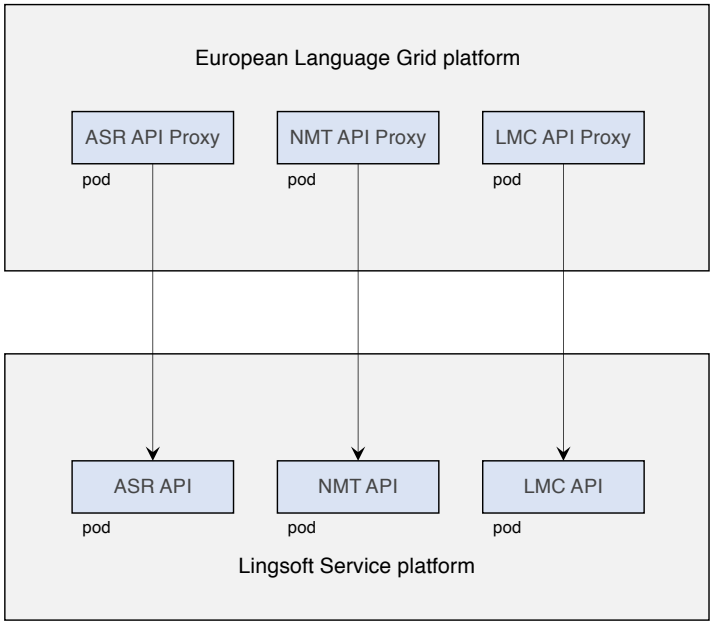

Fig. 1 API proxy containers relay Lingsoft's services to ELG

organisation would need to deploy the same Docker image into their environment are credentials from Lingsoft that allows calling the Lingsoft back end services.

As the ELG technical team preferred one service per functionality and language. This meant that Lingsoft provided a total of 35 services for ELG integration. The full set of services is presented in Table 1.

Service	Supported Languages/Domains
Speech recognition	Finnish, Finnish Healthcare, Swedish, Norwegian bokmål
Machine translation	Finnish ↔ English, English ↔ Swedish, Finnish ↔ Swedish, German ↔ Finnish
Proofing	Finnish, Finnish Healthcare, Swedish, Finland Swedish, Danish, Norwegian bokmål and nynorsk, English
Morphological analysis (incl. Lemmatization)	Finnish, Swedish, Danish, Norwegian bokmål and nynorsk, English
Named Entity Recognition (NER)	Finnish, Finnish Wikidata, Finnish YSO, Swedish, Danish, Norwegian bokmål and nynorsk, English

Table 1 Lingsoft services and languages

4 Evaluation

Generally, online guidelines and human integration support from ELG were clear and sufficiently detailed throughout the course of the project. The integrated services work per expectation in the "try out" user interface on the ELG platform.

Lingsoft also provided the ELG project with feedback from a commercial perspective regarding the integration process and platform functionality. For example, the demonstration services available in the "try out" box are quite slow. Lingsoft's speech recognition supports near real-time "live" subtitling/dictation, but this is not yet possible to demonstrate via the ELG platform. The commercial aspects of the platform are also work-in-progress at the time of writing, with no working solution for billing an ELG end user for the use of, e. g., Lingsoft's services. At present, we provide our solutions through ELG mainly for demonstration purposes, as a marketing channel, and for non-commercial use.

5 Conclusions and Results of the Pilot Project

The ELG project allowed us to upgrade our service infrastructure for easier distribution via ELG as well as through other channels. We believe that we will continue to utilise other providers' ELG resources and services for our benefit, especially open source tools and resources. From our experience with trying to utilise open source tools from the academic community, the ELG approach of researchers (and other developers) providing their open source tools as shareable docker containers with an exposed API is a great improvement over the current situation.

For Lingsoft, ELG can be seen as an additional distribution channel for tools and services we already provide. As an SME from Finland, it is expected that an official EU platform will increase the findability of our services and raise the credibility of our solutions outside of Finland, where we are well known. ELG is therefore expected to facilitate reaching customers outside of Finland and the Nordics.

We provide our tools both for commercial usage (on a Software as a Service subscription model) by companies and organisations, and for research purposes (free of charge for non-commercial use). In our internal work processes, e. g., subtitling and translation, the dockerised tools and API access is ideal, as this facilitates keeping our technology pipeline modular, and the core language technology tools easily replaceable and/or upgradable.

A centralised catalogue of European language technology, if widely adopted, will be beneficial to private providers of language technology, such as Lingsoft, for reaching new customers with our tools and services offerings. Conversely, we hope our contribution to the platform with our services benefit ELG in becoming widely adopted by providing more quality items for the ELG catalogue. Our solutions are robust and widely used with a proven track record. Our spelling and grammar tools have been distributed with the Microsoft Office suite and are used by the Finnish Digital and Population Data Agency, as well as several of the largest newspapers in Sweden; we have collaborated with the Swedish Post and Telecom Authority and the public service broadcaster SVT in creating speech-to-text for Swedish and our Finnish speech-to-text is in use for transcription in a number of Finnish organisations, including the Finnish parliament.

As ELG grows, we believe we will get good exposure for our services by having them on display at ELG. The service adapter ELG integration allows us to continuously improve the content of our ELG services with a minimum of additional maintenance effort. We also intend to continue to release new tools and covered languages in line with our general development roadmap.

Lingsoft is proud to have been one of the selected organisations for the ELG integration projects. We look forward to being part of the continued development of the ELG platform and hope that a substantial part of the ELG visions are fulfilled in the near future.

Lingsoft's services can be found in the European Language Grid.[1]

[1] https://live.european-language-grid.eu/catalogue/search/Lingsoft

Acknowledgements The work described in this article has received funding from the EU project European Language Grid as one of its pilot projects.

References

Rehm, Georg, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Victoria Arranz, Andrejs Vasiļjevs, Gerhard Backfried, José Manuel Gómez Pérez, Ulrich Germann, Rémi Calizzano, Nils Feldhus, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Julian Moreno-Schneider, Dimitris Galanis, Penny Labropoulou, Miltos Deligiannis, Katerina Gkirtzou, Athanasia Kolovou, Dimitris Gkoumas, Leon Voukoutis, Ian Roberts, Jana Hamrlová, Dusan Varis, Lukáš Kačena, Khalid Choukri, Valérie Mapelli, Mickaël Rigault, Jūlija Meļņika, Miro Janosik, Katja Prinz, Andres Garcia-Silva, Cristian Berrio, Ondrej Klejch, and Steve Renals (2021). "European Language Grid: A Joint Platform for the European Language Technology Community". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations (EACL 2021)*. Kyiv, Ukraine: ACL, pp. 221–230. URL: https://www.aclweb.org/anthology/2021.eacl-demos.26.pdf.