MICROSERVICES ATYOURSERVICE

Bridging the gap between NLP Research and Industry

Co-financed by the Connecting Europe Facility of the European Union

The contents of this publication are the sole responsibility of the Microservices project and do not necessarily reflect the opinion of the European Union.

European Speech and Language Technology Landscape

- Technologies: speech recognition (ASR), machine translation (MT), speech synthesis (TTS), sentiment analysis, chatbots, proofing, …

- SaaS < PaaS < IaaS: Google Cloud, Azure, AWS, …

- So, all is good then, right?
  - For: Estonian MT, Finnish ASR, Icelandic TTS, Galician chatbots, …?

© Lingsoft Oy

European Speech and Language Technology Landscape, contd.

- Strong speech and language technology research community in almost all languages
  - Open source sharing of tools
- And, the European Language Grid (ELG)
  - "Our main objective is to address fragmentation in the European Language Technology business and research landscape by **establishing the ELG as the primary platform for Language Technology in Europe** and to strengthen European LT business with regard to the competition from other continents."
  - "The ELG will be **a platform for commercial and non-commercial Language Technologies**, both functional (running services and tools) and non-functional (data sets, resources, models)."

© Lingsoft Oy

Open Source Tools - Benefits and Challenges

Benefits

- Steaming hot research
- Established and trusted tools
- Free to use and adapt

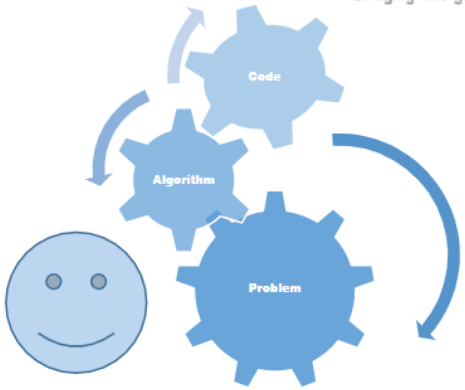
Challenges

- Difficult to find
- Poor documentation/lack of support
- Is it worth the trouble…?
- It works on my machine…

© Lingsoft Oy

MICROSERVICES ATYOURSERVICE

Bridging the gap between NLP Research and Industry



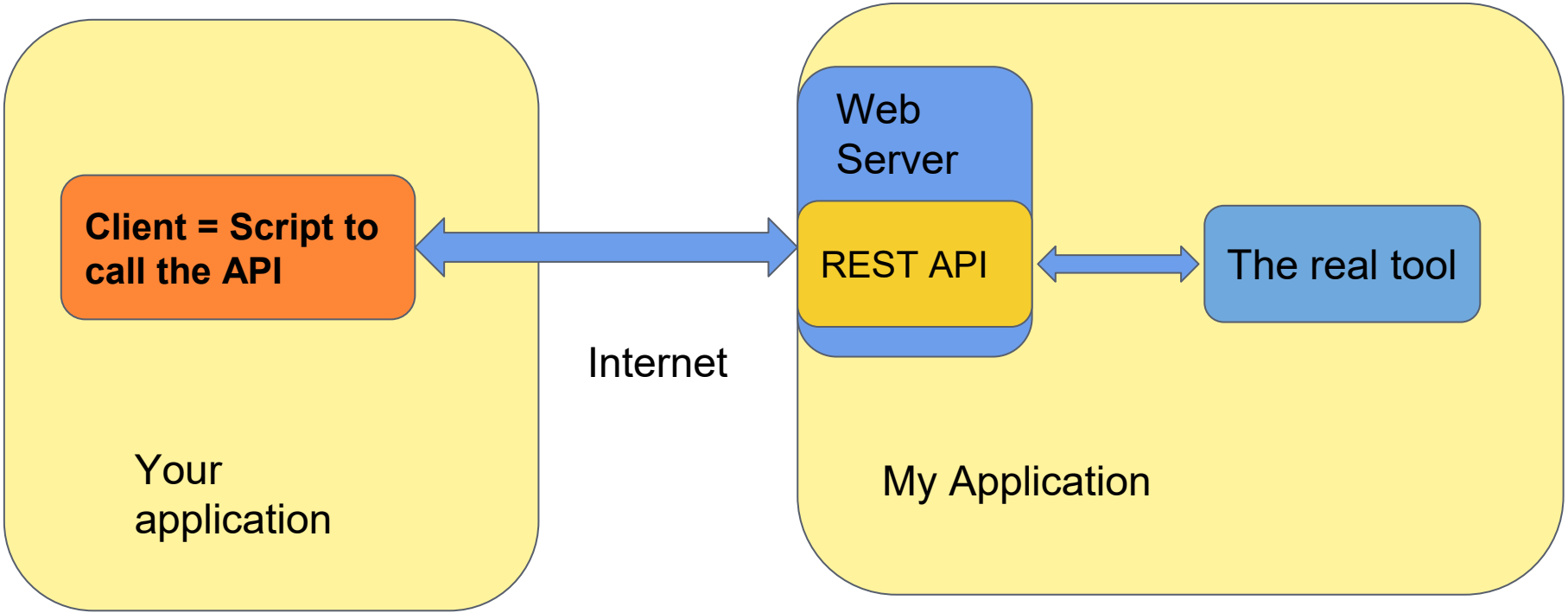
From Researchers and Their Githubs ...

Dockerized solutions + APIs

ELG
ELRC-SHARE

... Into Easily Adopted Solutions for Industry

SaaS > Web Service API > REST API



© Lingsoft Oy

What is Docker?

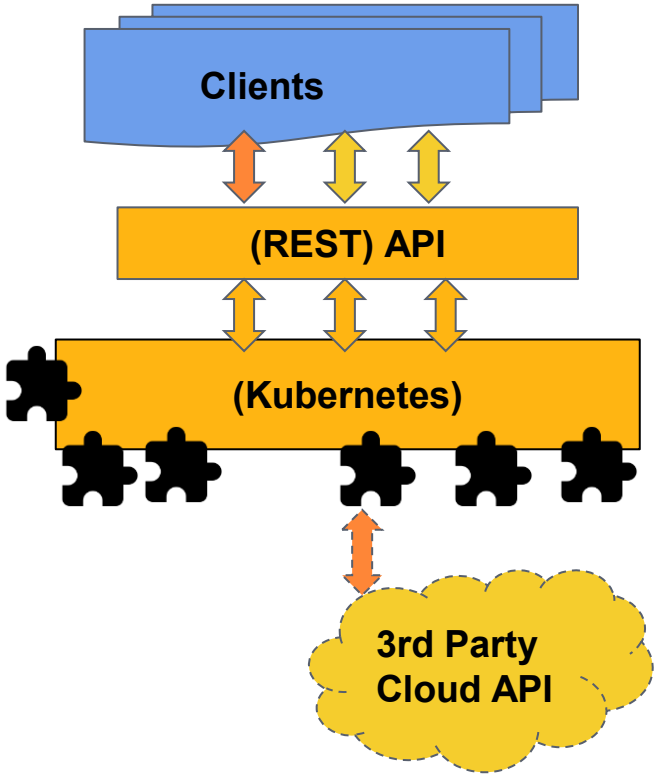
- Wikipedia says: "[Docker is] a set of platform as a service (PaaS) products that use OS-level virtualization to deliver software in packages called containers."
- **Image** (the shareable software package) - **Container** (the running software package)

- Makes it easy to install tools with different dependencies/OS on same **host** machine
- Makes it easy for others to install & use the tool
  - Previous years machine translation/parsing/whatnot systems for scientific challenges
  - Previous years software development examples for students
  - …
- Makes it easy to create microservice architecture

© Lingsoft Oy

Microservice Software Architecture

- = NLP component in Docker container
- Easy integration of open source and 3rd party components
- Easy extension to new languages and functionality
- Easy replacement of components



© Lingsoft Oy

Progress so Far (Lingsoft Example)

Tool name	Description	Language	Original Creator	DockerHub	ELG
HeLI OTS	Language identification	Multilingual	University of Helsinki	link	TBD
Finto AI	Topic classification	Finnish	National Library of Finland	link	link
KB BERT NER SV	Named entity recognition	Swedish	National Library of Sweden (KBLab)	link	TBD
NB BERT NER NO	Named entity recognition	Norwegian	National Library of Norway (NbAiLab)	link	TBD
…	…	…	…	…	…

Project Summary

- **Project period**: March 2021 - February 2023
- Co-financed by Connecting Europe Facility (CEF)

Project number 2020-EU-IA-0046

- **Objective**: Make at least **40 tools** available as easily integratable microservices covering a total of **11 languages**: Finnish, Swedish, Norwegian, Northern Sami, Estonian, Latvian, Lithuanian, Spanish, Portuguese, Icelandic, and Faroese

- **Project webpage** contains a continuously updated list of tools and previous workshops

Lingsoft®

www.lingsoft.fi

contact: sebastian.andersson@lingsoft.fi

Eteläranta 10, 00130 Helsinki

Kauppiaskatu 5 A, 20100 Turku

+358 2 2793 300