

EU-BRIDGE – the Goal

Bridges Across the Language Divide (EU-BRIDGE) aims to develop speech and machine translation capabilities that exceed the state-of-the-art in new and more challenging business cases. It will advance spoken language technologies to permit transmission of human information content from one language to another in situations that cannot be handled by automatic techniques to date.

Another key challenge is to reduce the cost of data collection and the depen-

dence on large, manually collected and annotated data for the development of the core recognition and translation technologies. The team of partners proposes to build one of the largest combined language capabilities in speech recognition and translation.

EU-BRIDGE seeks to achieve rapid technology transition and market insertion by building distributed services instead of transferring software and by making deployment part of the project.

EU-BRIDGE – Objectives

Effective, innovative and independent alternatives must be provided for Europe to serve its translation, communication, content processing, and publishing needs. The joint expertise of the consortium will drive the technological development forward, ensure best performance of core

capabilities, and ensure that insertion in actual business operations is achieved. At the core of EU-BRIDGE's objectives is the development of a speech translation service infrastructure upon which several use cases will be built.



EU-BRIDGE – Research

In order to build the service infrastructure for speech translation, engines from the academic partners will be integrated into a service framework that offers an easy-to-use infrastructure for the use case developers. Therefore, the academic partners perform research at their laboratories on the following:

Performance:

We will advance spoken language technologies so they process and transmit human information content from one language to another in situations that can so far not be handled by automatic techniques. This includes specialized but varied topics (lectures, seminars, presentations), as well as highly disfluent, conversational, accented, and noisy speech (meetings, telephone calls). We will perform research in the areas of robustness, rapid adaptation in speech and translation, semantic modelling, and content summarization. We will also develop personalization schemes that adapt systems to individual users and groups of users for more specific high performance operation that will address business needs better than a web-based one-size-fits-all solution.

Language Portability for Europe:

EU-BRIDGE will provide speech and translation capability for languages of main interest to Europe. Building on key efforts and expertise, it will build one of the largest combined repertoires of languages available both in speech recognition and translation and will stretch to do so robustly for all communication channels. We will include core European languages and under-resourced European languages and will reach out to languages of the BRIC economies. Focused research efforts to improve portability itself play an important role. These efforts will lower the cost of moving capabilities effectively from one language to another.

Reduction of the Dependency on and Cost of Data:

Since data is the “crude oil” of information processing, solutions must make production cheaper and reduce our dependency on it. By making speech and machine translation components adaptive and language and style independent, and by streamlining the process, EU-BRIDGE will significantly reduce data needs. By involving the users themselves in correcting and building the systems

implicitly, i.e. by crowd sourcing, the cost of data acquisition will be reduced, thus, building and improving the systems. By taking better advantage of available but not well-prepared data, the cost of data preparation will be reduced and the effective usable data increased. This includes comparable data, monolingual data, spoken and textual data, noisy data, and automatic methods for judging the quality of data.

EU-BRIDGE – Evaluation and Dissemination

The consortium will monitor the progress of our activities in this project and thoroughly evaluate the success. EU-BRIDGE will evaluate the improvements from the core research activities on in-house test sets using automatic measures. The test data for the in-house tests will be closely related to the use cases developed in this project and will be collected with the help of the partners heading the use cases. The engines developed for the integration into the speech translation service infrastructure will be evaluated on an international stage against worldwide participants in external evaluation campaigns.

Rapid Technology Transition and Market Insertion:

The program will strive to transition research development to commercial deployment more rapidly. This will be done by building distributed services instead of transferring software and by making deployment part of the project. The systems will be applied to real-world data, and we will carry out pilot experiments and field studies around four business opportunities.

