iEMS – Interactive Experiment Management System for Machine Translation

Pēteris ŅIKIFOROVS

p@mak.id.lv

Abstract. Interactive Experiment Management System (Interactive EMS or iEMS) is an experiment management system with a graphical user interface for designing and running statistical machine translation experiments. It is written in JavaScript and runs in all modern desktop browsers with no installation. iEMS produces a script that can be used to train a complete machine translation engine from scratch. There is an optional backend that can be used to launch and configure virtual servers on the Amazon Web Services cloud for running experiments. It is an open-source project licensed under the Apache 2.0 license. The development of iEMS is supported by the European Association for Machine Translation (EAMT). Project website: https://github.com/pdonald/iems

Description

Training a statistical machine translation engine consists of many steps. There are experiment management systems like Moses EMS and eman that help manage all these steps. However, their configuration is stored in text files which makes it is difficult to visualize the order of execution and adding or modifying steps may require diving into application code.

iEMS is an interactive experiment management system with a graphical user interface that aims to make it easy to design and run machine translation experiments from scratch. The main objective of the project is to have very little friction to get started. It is a single-page application written in JavaScript intended to run in modern desktop browsers without any setup.

In the application, there are several predefined tools that represent various steps in machine translation training that can be dragged and dropped onto a design surface. They can be linked together to set the order of execution. Tools can be grouped which allows quick swapping one set of tools with another.

There is also a backend that allows the user to run their experiments. The experiments can be run on a local host via the secure shell SSH or a new virtual server can be launched to run them on the Amazon Web Services cloud with a single click. Both regular (only pay for what you use) and spot (cheaper but can be terminated at any time) types of virtual servers are supported. Docker, a tool for automating the deployment of software inside Linux containers, is used for provisioning which means it is not necessary to compile or install dependencies for machine translation tools and setting up the server for use takes very little time.

The project is still in development. Its development is supported by the European Association for Machine Translation.