OctaveMT: Putting Three Birds into One Cage

Juan A. Alonso
Lucy Software Ibérica (ULG Group)
Copèrnic 42-44 1r
08021 Barcelona, Spain
juan.alonso@ulgroup.com

Albert Llorens
Lucy Software Ibérica (ULG Group)
Copèrnic 42-44 1r
08021 Barcelona, Spain
albert.llorens@ulgroup.com

Abstract

This product presentation describes the integration of the three MT technologies currently used – rule-based (RBMT), Statistical (SMT) and Neural (NMT) – into one scalable single platform, OctaveMT. MT clients can access all three types of MT engines, whether on a user specified basis or depending on several translation parameters (language-direction, domain, etc.)

1 Introduction

Historically, Lucy Software and Services (a company of the United Language Group) has been focusing its development efforts on its RBMT system. However, during the last few years, we started to develop and use SMT technology and during the last months we have also been working on the NMT area. Our mid-term goal is to have an operational RBMT–NMT hybrid engine.
The aim of this presentation is to introduce and describe the integration of all three MT technologies into one single product platform, OctaveMT.

2 System Architecture

The system architecture is depicted in Figure 1. The platform keystone is the LT Task Scheduler component, a portable and scalable task distribution system offering high performance for many kinds of services. It accepts translation requests from one or more MT Clients through a RESTful API. These translation requests are stored in the Task Pool component of the Task Scheduler.

The translation tasks are then handled by one or more MT engines. Each engine has an eServant component that monitors its activity; when it is idle, it fetches one request from the Task Pool.

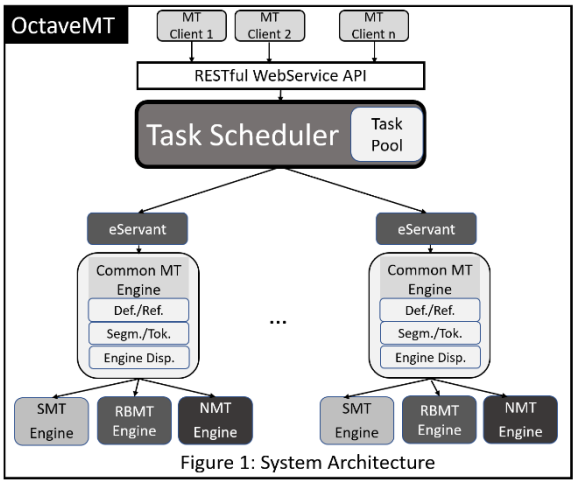

Figure 1: System Architecture

This task is then fed through the deformatter, the segmenter, the tokenizer and, finally, the engine dispatcher. The dispatcher sends the segmented text to the back-end engine type specified in the translation task (RBMT, SMT or NMT). After that, the translated text is sent back to the reformatter, and finally delivered to the originator MT Client through the Task Scheduler.

3 Advantages of this Approach

By re-using common sub-components for the three types of translation engines, tasks such as document format handling and conversion, which typically are a problem for raw SMT & NMT engines, can be properly handled. Additionally, this approach allows to use standard load-balancing techniques to build distributed high-performance MT infrastructures.

Pérez-Ortiz, Sánchez-Martínez, Esplà-Gomis, Popović, Rico, Martins, Van den Bogaert, Forcada (eds.)
Proceedings of the 21st Annual Conference of the European Association for Machine Translation, p. 339
Alacant, Spain, May 2018.