

NEC TM DATA PROJECT

Alexandre Helle

Pangeanic / B.I. Europe
PangeaMT Technologies Division
Valencia, Spain
a.helle@pangeanic.com

Manuel Herranz

Pangeanic / B.I. Europe
PangeaMT Technologies Division
Valencia, Spain
m.herranz@pangeanic.com

Abstract

The objective of project NEC TM Data¹ is to organise unexploited national bilingual assets that can be used as open data and general data for machine learning, in order to lower translation costs at a national level and across member states. It runs a study on the expenditure at the national level on translation contracts, as well as at the regional and municipal levels. The software will help member states centralise these language assets with the *NEC TM* database, following industry best practices.

NEC TM is based on the ElasticSearch² (Gormley, 2015) *ActivaTM* server which is a centralised translation memory (TM) server independent of any computer-assisted translation (CAT) tool for efficient data sharing, TM matching, TM retrieval, and domain categorisation of resources. In short, *ActivaTM* separates the need of every CAT tool to have its own TM server. It is possible to store bilingual assets and later retrieve them through any CAT tool using the API calls to *NEC TM* Translators can translate and access each other's work simultaneously from different CAT tools.

ActivaTM is the basis for the *NEC TM* (fork-out) for the scope of this project. It has been selected as a CEF (Connecting Europe Facility) project by the European Commission as the database of choice to provide unified translation memory

services to EU public administrations and collect and build bilingual big data from public translation contracts.

Each European country will be able to install their own *NEC TM* and new translation contracts from translation companies will benefit from fuzzy matching analysis and will be able to work online and connect to each national *NEC TM* server.

Translation data will be categorised in *NEC TM*, and a connection provided to eTranslation and ELRC.

The consortium for the project is composed by Pangeanic, Tilde, Ciklopea and Secretary of State for Digital Progress (SEAD) of Spain.

The NEC TM Data project consortium advocates for the facilitation of a single digital market. It will act as a meeting point for European data gathering efforts and the collection of national digital big data. By building a data bridge between public administrations and translation vendors, NEC TM Data project will promote the free flow of data between Public Administrations and translation professionals.

References

Gormley, Clinton, and Zachary Tong. 2015. *Elasticsearch: The definitive guide: A distributed real-time search and analytics engine*. O'Reilly Media, Inc.

© 2019 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CCBY-ND.

¹ Action No. 2017-EU-IA-0149 shall run from 01/09/2018 until 28/02/2020

² <https://www.elastic.co/es/products/elasticsearch>