

MT4ALL

UNSUPERVISED MT FOR LOW-RESOURCED LANGUAGES

Expected outputs:

- bilingual dictionaries &
- MT engines for under-resourced language pairs

Methodology:

- Unsupervised training (Artetxe et al. 2019)

USE CASES

A) Financial domain

Finnish, Norwegian, Latvian
→ English

B) Legal domain

Ukrainian, Georgian and
Kazakh → English

C) Customer support

Norwegian, Spanish and
German → English

D) General domain

Basque, Catalan → English

PROJECT STATUS

- Compilation and cleaning of monolingual training corpora for use cases A & D (general and in-domain)
- Preparation of validation sets for use cases A & D
- Training of initial unsupervised models and dictionaries for use cases A & D

