Global Under-Resourced Media Translation (GoURMET)

Alexandra Birch,[1] **Barry Haddow,**[1] **Ivan Titov,**[1]**, Antonio Valerio Miceli Barone,**[1] **Rachel Bawden,**[1]
Felipe Sánchez-Martínez,[2] **Mikel L. Forcada,**[2] **Miquel Esplà-Gomis,**[2] **Víctor M. Sánchez-Cartagena,**[2]
Juan Antonio Pérez-Ortiz,[2] **Wilker Aziz,**[3] **Andrew Secker,**[4] **Peggy van der Kreeft**[5]

[1]University of Edinburgh (GBR), [2]Universitat d'Alacant (ESP)
[3]Universiteit van Amsterdam (NDL), [4]BBC (GBR), [5]Deutsche Welle (DEU)

Abstract

We present the EU H2020 GoURMET project (2019-2021) which aims to tackle the challenge of low-resource machine translation for our media partners. This will help them to both monitor news in a wider range of languages, and also more efficiently produce content especially for languages from Africa and India.

Overview

Machine translation (MT) is an increasingly important technology for supporting communication in a globalised world. Over the last few years neural machine translation methods have led to significant improvements in translation quality. However, they rely on large parallel corpora for training and are not able to deliver usable translations for the vast majority of language pairs in the world. The aim of GoURMET is to significantly improve the robustness and applicability of neural machine translation for low-resource language pairs and domains. The project is co-ordinated by the University of Edinburgh, and the other partners are: the Universitat d'Alacant, the Universiteit van Amsterdam, the British Broadcasting Corporation, and Deutsche Welle. So far we have delivered 8 translation models to our user partners (BBC and DW) for evaluation. These models cover the language pairs of English to and from Turkish, Gujarathi, Swahili and Bulgarian.

Objectives

1. Advancing low-resource deep learning for natural language applications;

2. High-quality machine translation for low-resource and diverse language pairs and domains;

3. Development of tools for media analysts and journalists;

4. Sustainable, maintainable platform and services;

5. Dissemination and communication of project results to stakeholders.

Use Cases

1. Global content creation – managing content creation in several languages efficiently by providing machine translations for correction by humans;

2. Media monitoring for low-resource language pairs – tools to address the challenge of monitoring media in strategically important languages;

3. International business news analysis – reliably translating and analysing news in the highly specialised financial domain.

Impact

The outputs of the project will be field-tested at partners BBC and DW, and the platform will be further validated through innovation intensives such as the workshops centred around our user group and BBC NewsHacks.

Contact

info@gourmet-project.eu
www.gourmet-project.eu