Shaping AI speech and text technologies for media and the newsroom

For whom?

It will help media monitors and journalists make sense of huge content streams (big data analysis) – and also enable them to enrich audiovisual (AV) output.

The challenge

Collecting, processing and visualising a huge amount of multilingual information (“up to ten million news items a day”) is a big (data) task.

How?

SELMA aims to build a continuous deep-learning multilingual media platform using extreme analytics. The platform will be open source.

Visit us

www.selma-project.eu

@SELMA_project

company/selma-project-eu/

Contact

Kay Macquarrie
SELMA project coordinator Deutsche Welle

info@selma-project.eu

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957017.

priberam

Fraunhofer IAIS

LABORATOIRE INFORMATIQUE D'AVIGNON

Institute of Mathematics and Computer Science University of Latvia

European Commission | Horizon 2020 European Union funding for Research & Innovation

Languages

01 **Albanian** Shqip
02 **Amharic** አማርኛ
03 **Arabic** العربية
04 **Bengali** বাংলা
05 **Bosnian** B/H/S
06 **Bulgarian** Български
07 **Chinese** (Simplified) 简
08 **Chinese** (Traditional) 繁
09 **Croatian** Hrvatski
10 **Dari** دری
11 **English** English
12 **French** Français
13 **German** Deutsch
14 **Greek** Ελληνικά
15 **Hausa** Hausa
16 **Hindi** हिन्दी
17 **Hungary** Hungarian
18 **India** Tamil
19 **Indonesian** Indonesia
20 **Kiswahili** Kiswahili
21 **Latvian**
22 **Macedonian** Македонски
23 **Pashto** وټښپ
24 **Persian** ﻯﺳﺭﺍﻑ
25 **Polish** Polski
26 **Portuguese** Português do Brasil
27 **Portuguese** Português para África
28 **Romanian** Română
29 **Russian** Русский
30 **Serbian** Српски/Srpski
31 **Spanish** Español
32 **Tamil** தமிழ்
33 **Turkish** Türkçe
34 **Ukrainian** Українська
35 **Urdu** ودرا

Project

The aim of SELMA is to build a continuous deep-learning multilingual media platform using extreme analytics.

Technologies

The projects aims to work on technology that can process (very) large volumes of content and will feature a (self) learning AI system.

The system will be able to share information about data streams – and keep the added value of each language through a novel approach: a crosslingual common space.

That means: The system will always collect and analyse data in the original language and subsequently translate it into another language upon request.

Use Cases

Two use cases are created:

Media Monitoring - it helps journalists and media monitors to make sense of huge content streams (Big Data Analysis).

Multilingual Content Production - it helps newsrooms to enrich audiovisual output through entity recognition, translation and voiceover and thus makes it more accessible.