The *MELIA Observatory* Project

Media Literacy Observatory for Active Citizenship and Sustainable Democracy

Output T2.2 Instrument for media contents scrutiny

Project co-funded by European Union funds (ERDF, IPA, ENI)

Document Reference

Project Acronym	Melia Observatory			
Project Code	**DTP3-657-4.1**			
Project URL	http://www.interreg-danube.eu/approved-projects/melia-observatory			
EU Project Officer	Johannes Gabriel			
Project Manager	**Name**	prof. dr. Matevž Tomšič	**Affiliation**	SASS
	Email	matevz.tomsic@fuds.si	**Phone**	+386 40 300-292
Output Name	Instrument for media contents scrutiny			
Output Number	T2.2			
Type	Report, Data	**Distribution Level**	Internal	
Responsible Author(s)	Leo Mršić, PhD (AUC)			
Contractual Date of Delivery	August 2021.			
Status	Final			
Quality assurance reader	Tamara Besednjak Valič			

Change History

Version	Date	Status	Author (Unit)	Description
0.1	November 30th, 2021	Final	Leo Mršić, PhD (AUC)	Instrument for media contents scrutiny
0.2				
0.3				

Table of Contents

Introduction

Media data was be collected using sampling and selected media crawlers and it will be used to develop instrument for scrutinising media contents. The instrument rests on principles of AI. We will, in close cooperation of experts and practitioners from different fields, develop selected complex and sophisticated algorithms for detection of various types of media messages. We discern structure of information, generated by different media outlets. Special focus was be on detection of disinformation in media contents commonly known as the 'hate-speech', i.e. contents that spread disinformation and hatred among people.

Media audit models are the basis for development of capacity building tools and educational methods aimed at upgrading the level of media literacy. It is challenging to detect hate speeches well since they are context and domain dependent. Furthermore, trolls try to evade or even corrupt such machine learning classifiers. Even if progress is made in detecting hate speech and disinformation and raising awareness about it with concrete examples from our region, the purveyors of hate and misinformation could adjust to some of the same tools in order to evade detection. Therefore, special attention was be paid to methodological approach to use technology to create solutions to avoid such attacks. Focus was be on semantic, technological and managerial solutions of combining technological and human capacities in controlling the functioning of the algorithm. The instrument is available for both analytical and learning purposes. It contributes not only to higher level of media literacy, but also to higher sensibility for misuse of media messaging for aims that are in opposition to the principles of democratic society.

Programme Output Indicator

Number of tools for strengthening institutional capacities and supporting transnational multilevel governance developed and/or implemented.
Target value 1.0 delivered.

Development of selected algorithms

The algorithms are the regulations needed to perform the steps necessary to process information, and the tools for representing information are data structures. The choice of an algorithm to solve a given problem depends on the data structures used, and vice versa - the selected data structure affects the way the algorithm is developed. Therefore, algorithms are a list of steps or a sequence of unambiguous instructions which are understandable. The algorithm should be written at such a level of detail that it is possible to estimate its effectiveness.

When designing the architecture of the future platform, as well as thinking about the methods of its complex activities, we had to demonstrate not only substantive knowledge in a given field, but also:

- → the ability of abstract, logical, analytical and algorithmic thinking,
- → the ability to precisely present your thoughts and ideas,
- → the ability to predict the occurrence of possible events,
- → imagination and the ability to associate various facts,
- → the ability to correctly, legibly and unambiguously formulate problems,
- → and common-sense knowledge.

As a result, the algorithms we choose are closely matched to the task they need to do. The algorithms we choose are the best possible solutions and will contribute to the smooth functioning of the platform. They will also contribute to the development platform by tracking changes in languages and their adaptation in the technical environment.

Objectives

Utilizing the developed sentiment dictionaries, language resources, and annotated corpora, we employed machine learning methods in order to build and evaluate models for automatic sentiment analysis for each of the target languages. This includes empirical tests on the target domain (news publications and social media posts w/ an emphasis on hate speech detection). The utilized methods range from linear models to deep learning neural network architectures. The methods work on both news texts and social media posts.

Algorithms used: Machine learning methods

Classic approaches

In order to achieve a sentiment analysis system, we used various machine learning approaches.

At first, we tried the following feature-based approaches:

1. linear models (including SVM, Logistic Regression, Random Forests, and others) w/ various features
2. non-linear models (including SVM, FFNN neural networks, and others) w/ various features

The sets include bag-of-words, as well as sentiment clues using various sentiment lexicons and also using an emotion lexicon. Finally, we used vector embedding (such as those stemming from BERT).

Deep learning approaches

At the second we utilized the deep neural networks and various transformers to solve our problem. This includes the possibility of using BERT-based and LSTM-based models. Within the framework of our experiments, we tried the following techniques: transfer learning, multitask learning, and cross-language learning.

During the deep learning experiments, we heavily relied on produced sentiment corpora within our project. The knowledge base was also unitized.

Process of learning

Due to the specificities of our corpora to be produced, the final architecture and a set of techniques cannot be predicted. The best chosen model in terms of measured quality, performance, and explainability will be chosen and implemented into the web service.

Results

When working on the development of the application, we took into account a large number of known algorithms. The design team selected the ones that best appropriate the design of the MELIA Observatory platform, these are:

1. linear algorisms model
2. nonlinear algorithms model
3. LSTM model
4. BERT model

The principles of their functionality are presented below.

Linear models

Machine Learning is dynamical system meaning that the system which changes overtime changes such as Neural Network, Support Vector Machine, etc. In Machine Learning Linear Models uses convex optimization techniques such as (Linear Regression, Linear Support vector Machine, Expectation Maximization, Hidden Markov models, etc.) are the Linear models. Linear Models have a global minima in it or global maxima in it, therefore it deals with convex optimization. well, it also depends on your problem too, suppose you want to predict a model by definition the relation between input and output is directly proportion or inversely proportion meaning that changing one thing will directly effect the output of your model. Linear Regression is a simpler but a powerful model in econometrics, as it is mostly used for finding trends. it mostly depends on the co-variance of your feature inputs.

Classification problem if two classes can be separated by linear boundary then it is a Linear problem (Male/Female) problem.

Non-Linear models

Directly uses the Non-Linear Functions one thing to note here is that Non-Linear models are hard to deal suppose a neural network which is non-linear dynamical system, it has more than one minima so, it should be used convex optimization technique. Non-linear boundaries in data such as you are classifying sentences, when data is plotted on graph it shows that the classes are merge together and they form a circle and it is hard to separate a circle with Linear separator boundary. It needs a non-linear function to deal with it. Most of the time the physical systems are non-linear and we can not compute the exact value, that is why e need to use approximations which is easier rather than doing full analysis.

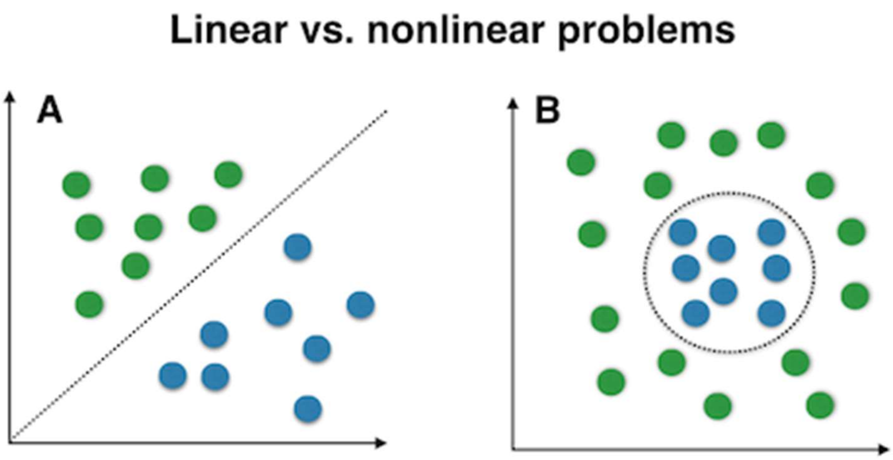


Figure 1: Linear vs. nonlinear problems

LSTM

LSTM is special kind of recurrent neural network that is capable of learning long term dependencies in data. This is achieved because the recurring module of the model has a combination of four layers interacting with each other.

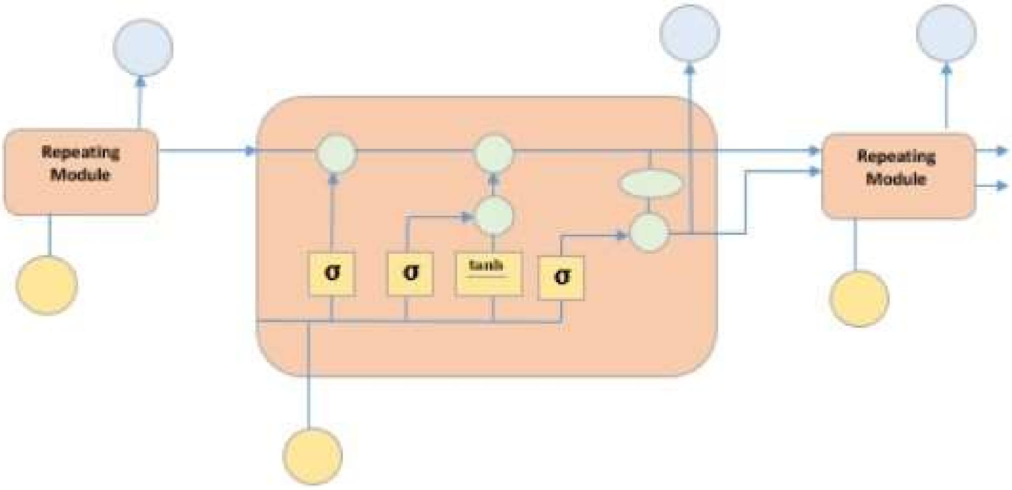


Figure 2: LSTM model

The picture above depicts four neural network layers in yellow boxes, point wise operators in green circles, input in yellow circles and cell state in blue circles. An LSTM module has a cell state and three gates which provides them with the power to selectively learn, unlearn or retain information from each of the units. The cell state in LSTM helps the information to flow through the units without being altered by allowing only a few linear interactions. Each unit has an input, output and a forget gate which can add or remove the information to the cell state. The forget gate decides which information from the previous cell state should be forgotten for which it uses a sigmoid function. The input gate controls the information flow to the current cell state using a point-wise multiplication operation of 'sigmoid' and 'tanh' respectively. Finally, the output gate decides which information should be passed on to the next hidden state.

BERT model

BERT model makes use of Transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text. In its vanilla form, Transformer includes two separate mechanisms — an encoder that reads the text input and a decoder that produces a prediction for the task. Since BERT's goal is to generate a language model, only the encoder mechanism is necessary.

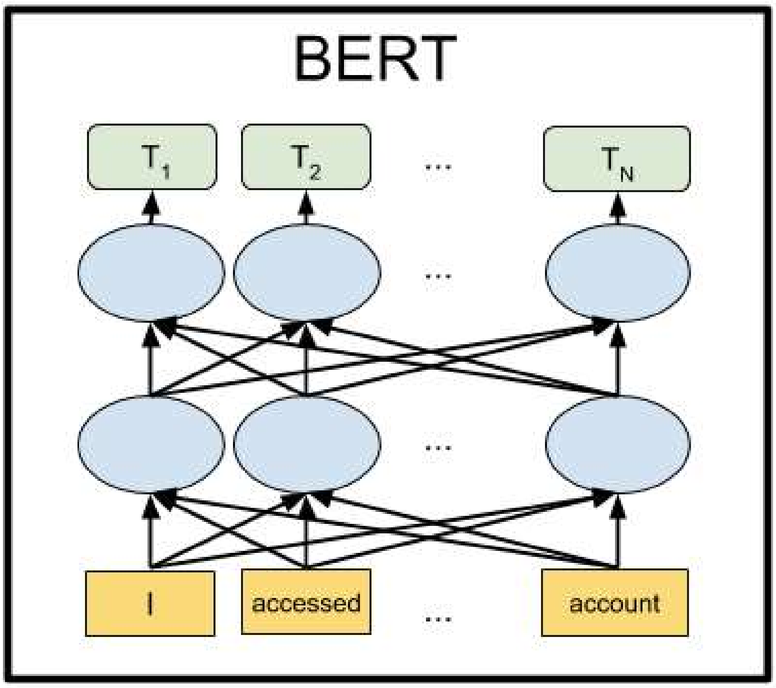


Figure 3: BERT model

BERT is undoubtedly a breakthrough in the use of Machine Learning for Natural Language Processing. The fact that it's approachable and allows fast fine-tuning will likely allow a wide range of practical applications in the future.

Platform architecture

User interface wireframes

The sketches presented in this section are wireframes for the web application. Although they are somewhat graphically developed, they may be slightly changed and improved during the implementation.

On the home page of the web application, the users are given a text box in which they may paste (1) a media text, (2) a link to a news portal media post, or (3) a list of links to news portal media posts. The system will automatically detect which of these modes of operation the user is requesting.

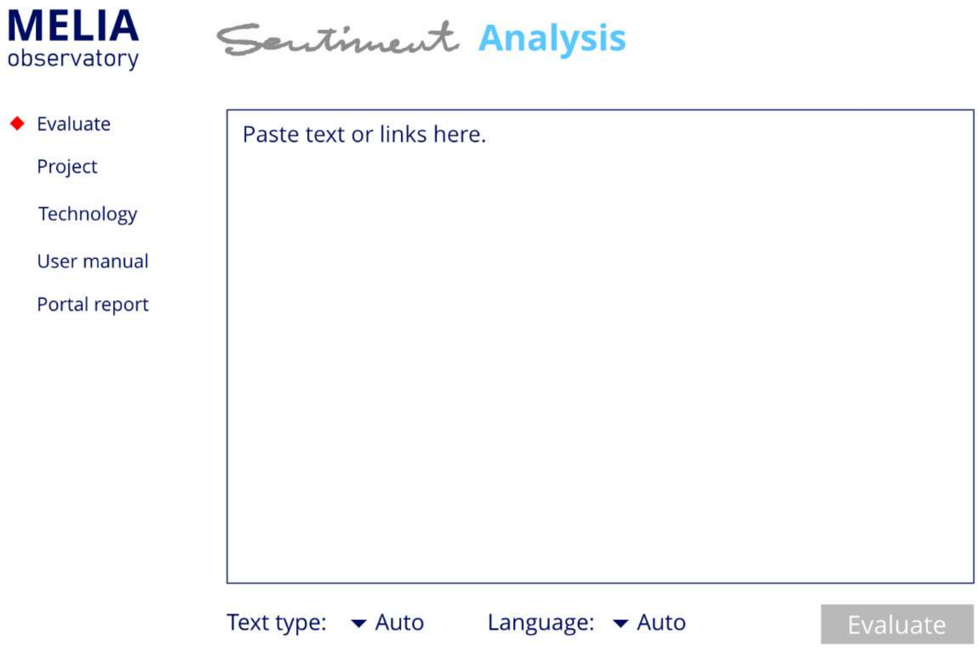


Figure 4. Home page of the web application

After pasting the users may select the button "Evaluate" which will run the main function of sentiment analysis on the text, as shown on Figure 2. Also, the users may let the system automatically detect the language of the document or may select one of the implemented languages to force sentiment evaluation in that particular language.

The same goes with text type – the users may select "news portal text" or "social media post" since the internal workings of the sentiment analysis service are such that for each language there are two models for sentiment analysis. Also, the users may let the system automatically detect the type of text it is dealing with.

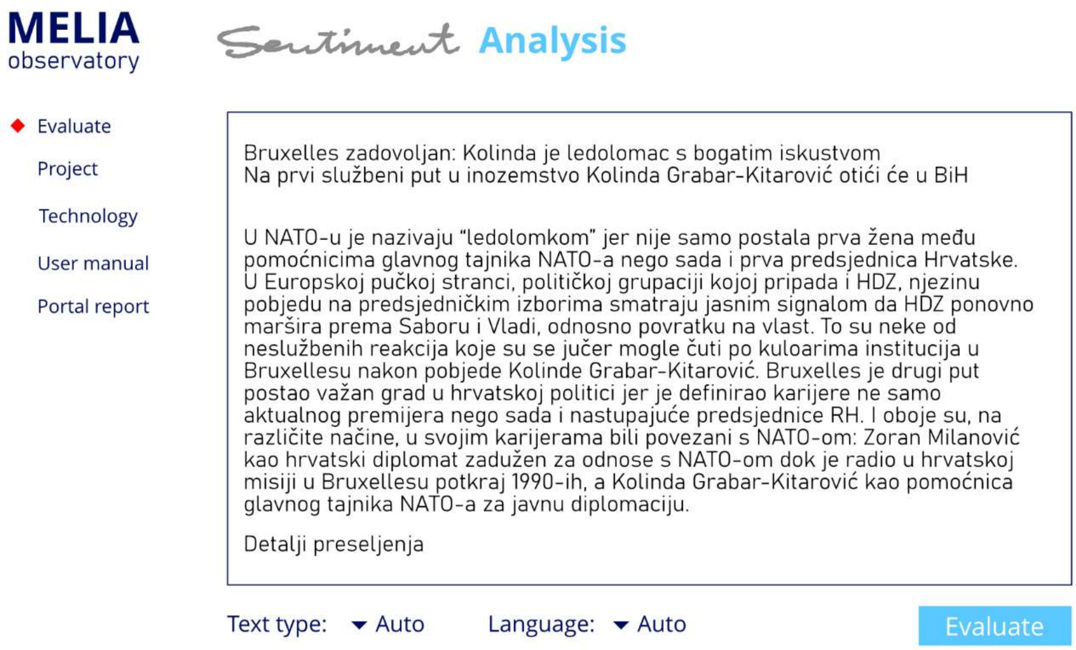


Figure 5. Evaluate option is enabled after entering text or link(s) in the main text box

If the user engages the mode of operation (1) or (2), a screen akin to the one shown on Figure 3. will be rendered. The user is shown the overall sentiment score for the text, and the most important words that led to this decision are highlighted.

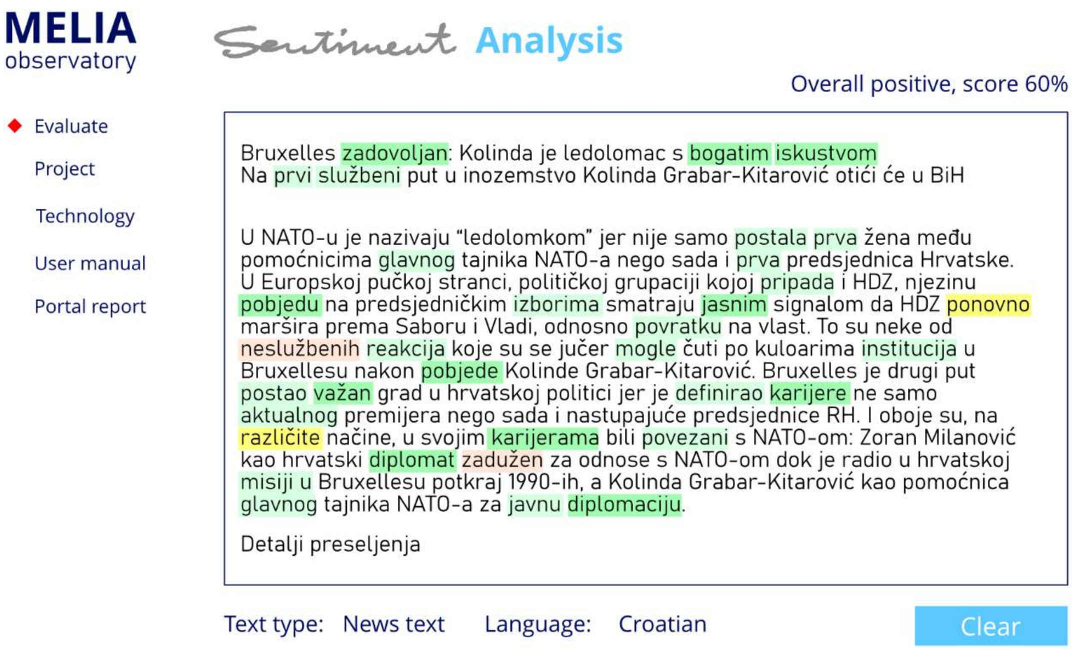


Figure 6. Automatically labeled sentiment

As shown on Figure 4., the user may see when each of the preselected news portals were crawled. The service crawls a list of news portals periodically (at least daily), finds new content, labels its sentiment, and saves it to the local database. So the users may select any of the news portals, and are led to the interface show on Figure 5.

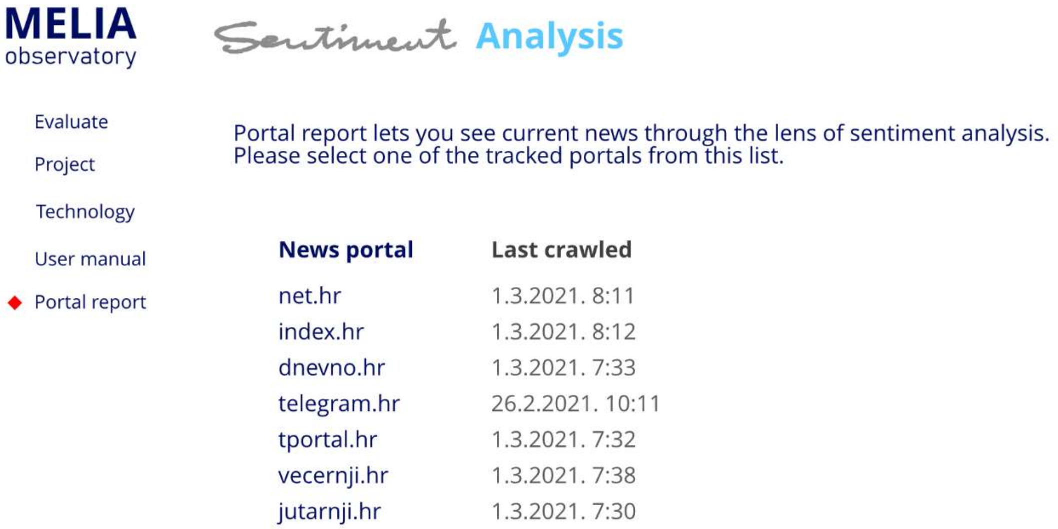


Figure 7. List of preselected and crawled news portals

Figure 8. Multi-document sentiment analysis list

Figure 5. shows an interface with a list of crawled articles which are automatically labeled with sentiment models. This list is helpful to gain a quick insight into the general sentiment orientation of a news portal title page. There is a number of scenarios where this kind of listing can be useful for media literacy education.

If the user selects any article from the list, an interface similar to the one on Figure 3. is being shown.

Also, the multi-document list shown on Figure 5. may be achieved by entering a list of links (for pre-selected news portal) into the main text box on Figure 1.

Modes of operation

To sum up the aforedescribed interfaces, here is a brief list of modes of operation enabled by our web application:

— sentiment decision on one text (news portal text or social media text)

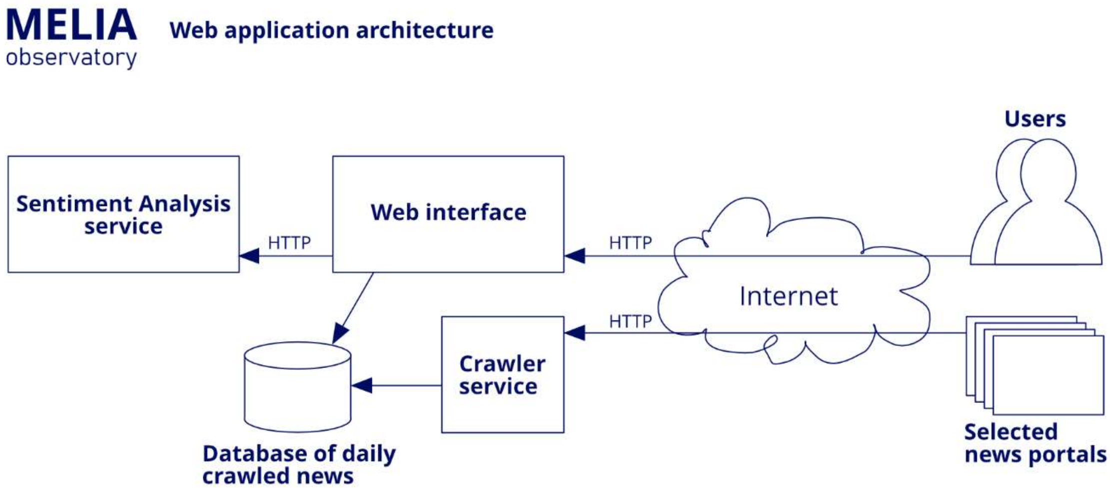
— sentiment decision on one link (news portal link)

— sentiment decision on many links (news portal links)

— sentiment decision on one portal home page (predefined news portals)

System architecture

Figure 6. shows components of our web application. The users use their browsers to fetch the web interface. The interface allows them to seamlessly call a separate module – the sentiment analysis service which for a given text (and some parameters) returns its sentiment and phrase labels.



3.

Figure 9. Web application architecture

Separately from this, there is also a Crawler service which works independently and periodically crawls news text from a list of chosen news portals. This news are being saved into a database of daily crawled news which can be accessed through the web interface as well.

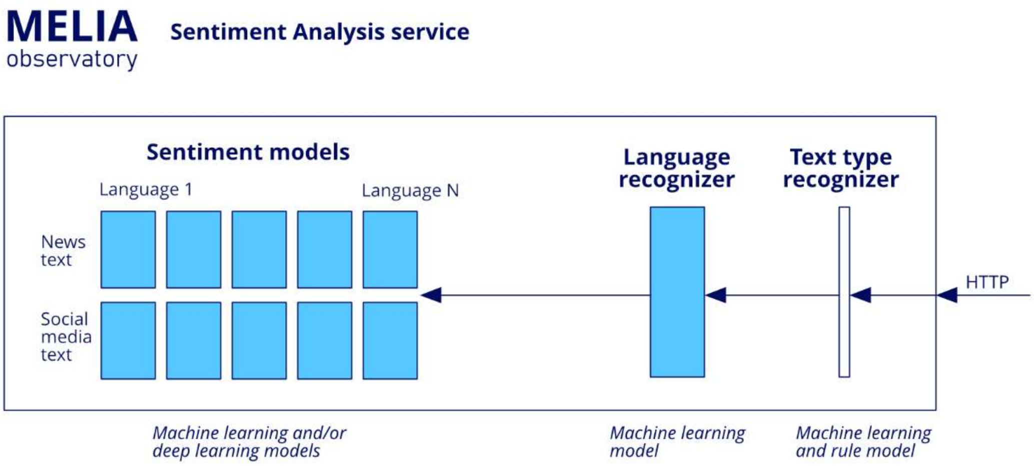


Figure 10. Sentiment analysis service architecture

The sentiment analysis service has a straightforward architecture. Once a request has been made with a text to be processed, first the text type recognizer decides if it is a news article or a social media post by using a simple machine learning model and some rules. After that, the request is being fed to the language recognizer which uses a machine learning model to find out the language of the text. If the language of the text is not matching any of the modeled languages, then either the most similar language is chosen, or an error is returned.

Of course, a request may specify the text type and the language of the target text so one or both of these steps may be skipped.

Since at this point, the information on language and text type are known, the service chooses the appropriate model and feeds the text from the request to it, in order to return the overall sentiment score, label, and important words to the interface.

The sentiment models will be done using machine learning and/or deep learning methods.

Results

The result of our activities is a user-friendly web application that will serve as a tool for practical education in the field of media literacy. Educators can use a web application to show and analyze examples of harmful news texts. He will also be able to show, using examples from everyday life, the distribution of moods among the articles of the internet portal, and thus analyze its daily channel.

The application is intuitive, easy to use and has basic functions that enable the analysis of source texts. It is adapted to users of all ages, although it is especially dedicated to young people.

Thanks to the fact that it will be possible to use the platform in 5 languages – Croatian, Slovenian, Hungarian, Romanian and English, a very large part of the European Union society will be able to use it. We hope that this will significantly increase the awareness of European Union citizenship about the issues related to the conscious use of media literacy.