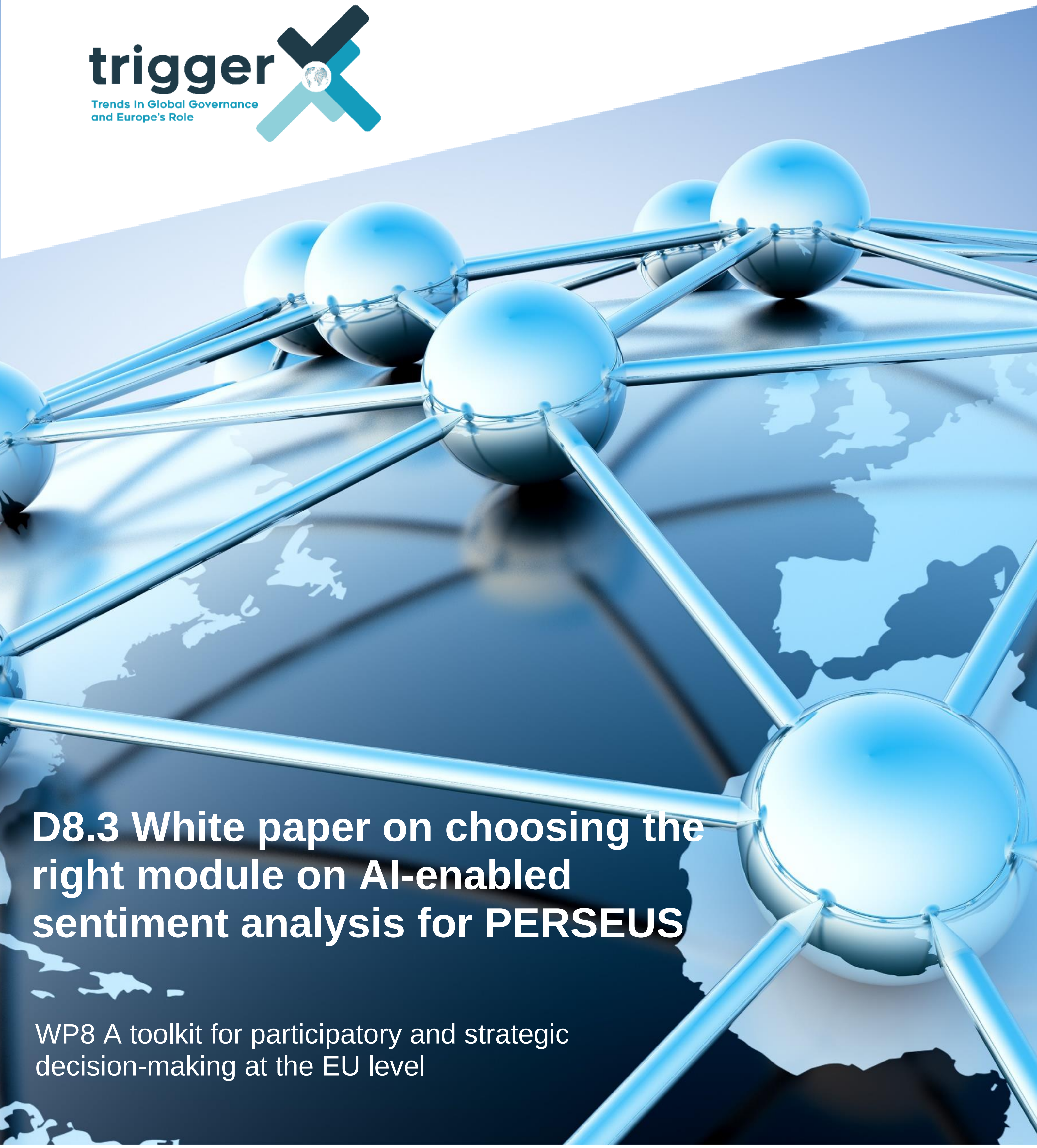

D8.3 White paper on choosing the right module on AI-enabled sentiment analysis for PERSEUS

WP8 A toolkit for participatory and strategic decision-making at the EU level

Grant Agreement n° 822735, Research and Innovation Action

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 822735. This document reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

TRIGGER

TRends in Global Governance and Europe's Role

Deliverable number: D8.3	
Deliverable name:	D8.3 White Paper on choosing the right module on AI-enabled sentiment analysis for PERSEUS (working paper)
WP / WP number:	WP8 A toolkit for participatory and strategic decision-making at the EU level
Delivery due date:	30.11.2019
Actual date of submission:	4.03.2021
Dissemination level:	Public
Lead beneficiary:	POLIMI
Contributor(s):	Marco Brambilla, Diletta Di Marco, Lorenzo Corti, Andrea Tocchetti, Tommaso Buganza

Changes with respect to the DoA

Not applicable

Dissemination and uptake

Public

Evidence of accomplishment

Report

Content

Index of Tables

Index of Figures

1. Introduction

This document is developed as part of the TRIGGER project (Trends in Global Governance and Europe's Role), which has received funding from the European Union's Horizon 2020 Research and Innovation programme, under the Grant Agreement number 822735.

This document aims at describing the tools among which to choose the right module on AI-enabled Sentiment Analysis for PERSEUS. The chosen format is represented by a white paper, an authoritative guide aimed at informing readers concisely about the ongoing methods on AI-enabled tools for Sentiment Analysis. The main goal is to describe the methodology and best practices and tools in an attempt to help readers understand the reasons behind the application of an AI-enabled module in PERSEUS.

The AI-enabled tool allows policymakers to formulate hypotheses and run SWOT analyses of problems in an unprecedented way. Exploiting the most recent machine learning algorithms, the aforementioned tool applies content analysis to extract valuable information from all kinds of data generated in PERSEUS.

PERSEUS, thanks to the use of digital technologies, combines public engagement, strategic foresight and the development of accurate, detailed policy decisions over the home and global governance. Among all, the AI-enabled tool has the most important impact in the process, enabling the toolkit in PERSEUS to understand and capture all the produced information in an intuitive manner.

The deliverable is structured as follows: Chapter 2 illustrates an overview of PERSEUS and the actors involved, Chapter 3 provides a complete and exhaustive explanation of Sentiment Analysis alongside some concrete use cases, Chapter 4 provides a practical overview of the available tools and libraries for Sentiment Analysis.

2. PERSEUS Overview

This chapter provides an overview of PERSEUS, a brief introduction about what it is, who are the actors involved, and what are the components of the toolkit.

PERSEUS aims at integrating different components to provide a comprehensive and organized toolkit to address policy-making strategies and challenges.

Figure 1 illustrates the components currently envisioned, which include:

- AGGREGATOR: A database containing two groups of datasets: datasets on global governance, contained in WP1, and datasets on EU governance, contained in WP2. These datasets were collected in the first year of the project and are currently being updated and refined. The final AGGREGATOR website will be composed of different elements from e.g. the four deep dives (an explanation is provided in deliverables 7.1 and 7.2), links to the GlobalStat database, the EurLex dataset, etc.
- COCTEAU: an experimental co-creation tool for public engagement that aims to engage European citizens in challenges related to contemporary societal problems, created by policymakers and European organizations.
- Foresight: a component involving the creation of scenarios, which will be published on the COCTEAU platform to collect data, relevant deep dives and technological papers.

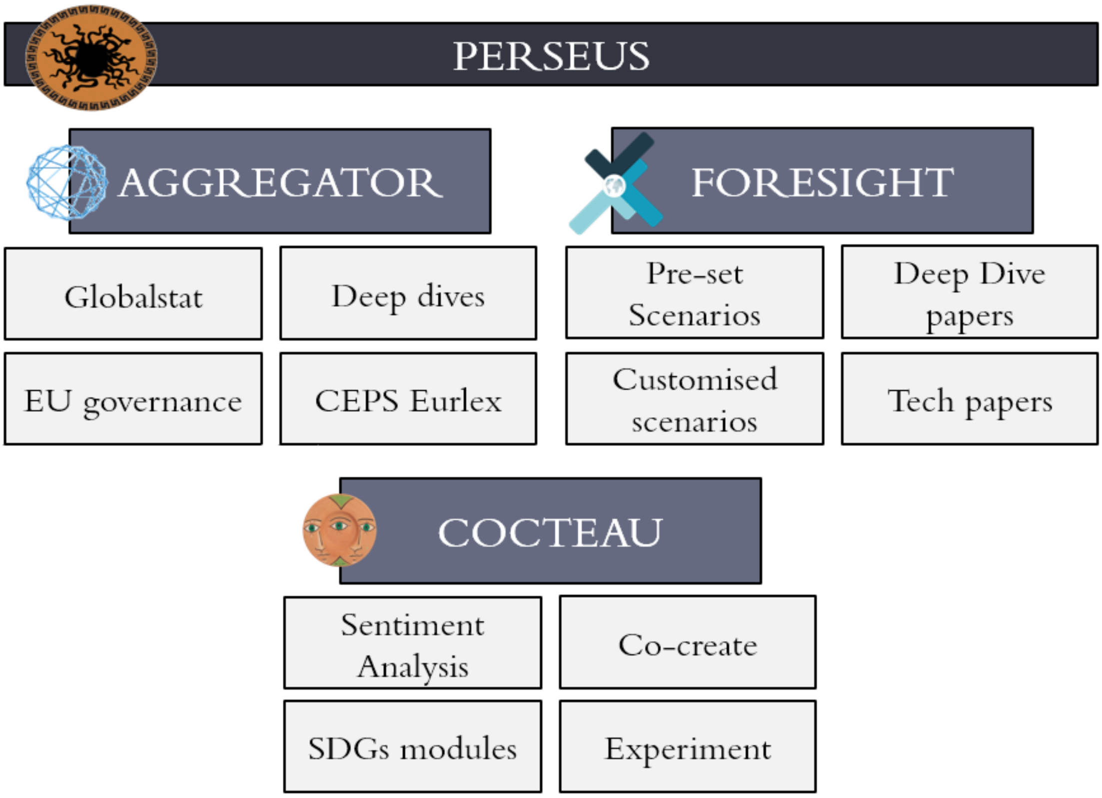

Figure 1 – PERSEUS main components

For each of the envisioned components, different people are involved. The three main actors, identified as potential users and their main tasks, are:

- Policymakers, who can use COCTEAU to define challenges about possible future scenarios, based on specific topics.
- Citizens, who are the end-users of the platform. They interact with each other, generating content on the platform on which analyses will be carried out. They are the main actors engaged in the co-creation activities organized on the COCTEAU platform (more details are provided in deliverable 6.7).
- Researchers, who query the data stored in AGGREGATOR and their purpose is to define new and interesting research questions about EU and global governance.

It is important to notice how AI-enabled Sentiment Analysis will be mostly applied to data from the COCTEAU platform since it is the main source of user-generated content.

3. Background: From Data to Sentiment Analysis

This chapter aims to provide an all-round general explanation of Sentiment Analysis, starting from an introduction to data collection and data analysis.

Nowadays, with the advent of the internet, a large amount of data became available worldwide, especially through social media, online search engines (like Trivago, Booking, etc.), and many other online sources. Data is not only available through online channels, but also corporations understood how relevant it is to collect data about their activities and customers (e.g. user feedback on a product or an overnight stay). Private sectors, like banking, have recently shown growth thanks to the increasing application of data analysis techniques.

3.1. Data Collection

It is known that companies gather their data through internal mechanisms, but which are the most famous online data channels through which data is widely available?

One of the, if not the, best public channel to collect User Generated Content (USG) is Twitter, a microblogging and news service through which people share their thoughts and take part in discussions through mentions to other users (using the @ symbol) and discussion themes (using the # symbol). The latter are also known as "Hashtags".

Accessing data is not always an easy task. Sometimes web APIs and libraries are available to extract and manage data from websites. In few instances (e.g. Twitter) support and guidelines are also provided. Whenever this kind of support is not provided, it is necessary to manually extract the required information. Therefore, some techniques have been developed over the years. The most common one is named "Scraping". A so-called "Scraper" navigates the chosen website simulating human activities throughout the web pages, analysing their HTML structure, and extracting the required pieces of information.

It is important to notice that accurate data management and transformations are necessary due to the inner heterogeneity of data.

3.2. Data Analysis

Wikipedia defines Data Analysis as "the process of inspecting, cleansing, transforming and modelling data to discover useful information, informing conclusions, and supporting decision-making". Data Analysis is a broad term that includes many different kinds of analyses with different data requirements, features, and objectives.

Data Analysis has indeed been classified into four different (ordered) groups:

- Descriptive Analysis aims to answer "what happened" by summarizing data, usually in the form of dashboards.

- Diagnostic Analysis aims to answer “why it happened”, creating connections between the available data and identifying patterns of behaviour.
- Predictive Analysis aims to answer “what is likely to happen”, making predictions about future outcomes through statistical modelling.
- Prescriptive Analysis aims to answer “what do I need to do” through new technologies, like Artificial Intelligence, which consume a huge amount of data to learn and provide enhanced decision-making insights.

It is important to mention how each one of these groups utilizes the outcomes of the previous ones to improve and perform its analyses.

Nowadays a plethora of data analysis tools are available in well-known programming languages like Python, MATLAB, Java, and R. These programs aim to support the user through libraries and pre-defined functionalities, contributing to ease the user experience.

Each “Analysis Group” includes many different types of analyses. Among all of them, this document will focus on Sentiment Analysis, explaining most of its characterizing aspects, alongside some significant use cases.

3.3. Sentiment Analysis

Figure 2 – Examples of words used in rule-based Sentiment Analysis systems

Sentiment Analysis is the interpretation and classification of emotions within textual data, like documents, newspapers, articles, comments, etc. using text analysis techniques to identify the polarity (positive, negative, neutral) of the considered text.

The figure below represents the triangular model, which provides a graphical representation for Sentiment Analysis, highlighting all the different involved aspects.

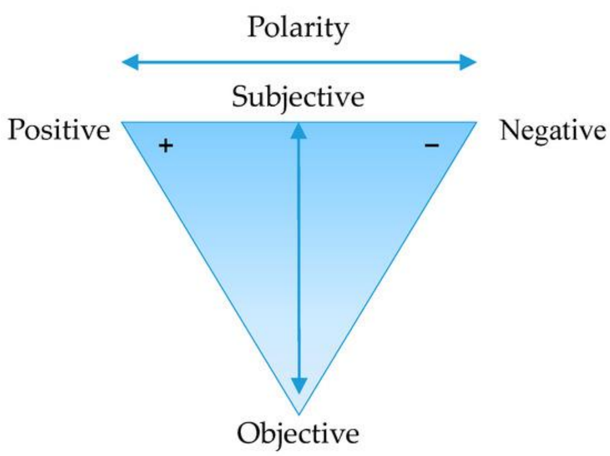

Figure 3 – Sentiment Analysis graphical representation

There are a wide number of cases for which Sentiment Analysis could provide useful insights, like understanding the level of satisfaction of a customer or the overall feel of hotel reviews.

3.4. Types of Sentiment Analysis

Sentiment Analysis focuses on polarity (positive, negative, neutral), but it may also focus on sentiment and feelings (happy, sad, etc.) or intentions (interested, not interested, etc.).

Indeed, depending on the desired outcome, different types of Sentiment Analysis have been developed:

- Fine-grained Sentiment Analysis is applied when the polarity is the main aspect to address. It expands the polarity adding further levels (e.g. very positive, very negative, etc.) to improve the precision of the outcome of the analysis.
- Emotion Detection focuses on detecting emotions (like happiness, sadness, disappointment, etc.) through lexicons[1] or machine learning algorithms. This kind of analysis is particularly complex since different people express emotions in different ways.
- Aspect-based Sentiment Analysis focuses on identifying the aspect, element, or feature which has been mentioned in a negative, positive, or neutral way.

Sometimes datasets contain texts in different languages due to the source of data from which it was built. This problem increases the complexity of the analyses, requiring further computations, like translations or language detection. Therefore, the system through which our analyses are performed should be improved to manage multiple languages at once.

[1] Elements made of words and the corresponding emotion they convey

3.5. Sentiment Analysis in practice

Sentiment Analysis uses various Natural Language Processing (NLP) methods and algorithms, on which different kinds of systems are built:

- Rule-based systems perform Sentiment Analysis based on a set of manually crafted rules. It exploits simple NLP techniques like stemming[2], tokenization[3], part-of-speech tagging[4], parsing[5], and lexicons. This is the simplest kind of system and doesn't take into account how words are combined in a sequence. This problem could be fixed by including further rules at the cost of increasing the overall complexity, which may influence results.
- Machine Learning-based systems rely on machine learning techniques to learn from data. A Sentiment Analysis task is usually modelled as a classification problem, whereby a classifier is fed a text and returns a category (e.g. positive, negative, or neutral). Pre-trained models for Sentiment Analysis tasks are widely available through libraries (e.g. Python libraries) and some of them could also be fine-tuned to achieve even better results. Alternatively, a custom classifier can be trained from scratch following machine learning principles.
- Hybrid systems combine both rule-based and automatic approaches, often achieving more accurate results.

3.6. Sentiment Analysis challenges

Even though modern Sentiment Analysis techniques can achieve good accuracy there are a few cases in which identifying the sentiment or emotions becomes even more complex.

Most of these cases occur when dealing with linguistic features and online language "slangs" (e.g. shortened terms, emoji, etc.), like:

- Subjectivity and Tone. Detecting subjective or objective sentences is relevant since objective texts do not contain any sentiment, therefore being able of analysing the tone of the sentence is core.
- Context and Polarity. Due to the presence of context, the polarity may change, and analysing sentiment without context is pretty difficult. Addressing this problem requires either a pre-processing or a post-processing phase, which is not that easy to manage and implement.

---

[2] Process of reducing inflection in words to their root forms, even if the stem itself is not a valid word in the language.
[3] Process of reduces the inflected words properly ensuring that the root word belongs to the language.
[4] Process of marking up a word in a corpus as corresponding to a particular part of speech, based on both its definition and its context.
[5] Process of determining the syntactic structure of a text by analysing its constituent words based on the underlying grammar of the language.

- Irony and Sarcasm. In this case, people express their negative sentiments using positive words. Therefore, it can be difficult to retrieve the correct polarity without understanding the context of the situation.
- Comparisons. Sometimes it is really difficult to retrieve the correct polarity without the context when comparing two things.
- Emojis. They are a longer combination of characters of a vertical nature which play an important role in the sentiment of texts, particularly in tweets.
- Defining Neutral. It is particularly relevant when training models, since depending on the meaning of “neutral” the outcome may change, depending on the labels assigned to the input in the training phase.

As mentioned, some of these challenges could be tackled either with pre-processing, post-processing, or other techniques, with the only drawback of dealing with an increase in the overall complexity of the system.

3.7. Sentiment Analysis use cases

As stated before, Sentiment Analysis has been widely applied in different environments, like the business, political, and social ones, while others are new to these techniques and are beginning to apply them only recently, like the banking one. Indeed, most companies, like Apple, Google, Amazon, TripAdvisor, etc. apply it to improve their services by analysing users’ comments and feedback. This chapter will illustrate how Sentiment Analysis was used to analyse the feelings behind some cases which took place in the last years.

In [4] the authors utilized a lexicon-based approach to analyse the sentiment of a dataset made of news published by the BBC. For each one of the news categories, they classified the feelings focusing on three polarities: positive, neutral, and negative. They found out how the categories Sport and Business had more positive articles, while Entertainment and Tech had more negative articles. From the representation they provided, it is also possible to notice how only a few articles were classified as neutral.

The second case we would like to focus on is Uber. Due to its worldwide diffusion, Uber receives a lot of feedback, comments, and complaints through social media. In 2018, 60’000 data points involving all the public interactions among customers and the society were collected from Facebook, Twitter, and the news. In particular, the key product thematics were Cancel, Payment, Price, Safety, and Service. The gathered data were filtered to remove all the spam and the junk coming from social media and then analysed to understand the overall brand perceptions. The outcomes revealed how comments from Facebook and Twitter were mostly negative for all the previously mentioned topics, while these coming from the news were mostly positive. In the end, this study revealed how overall the perception of the service provided by Uber was mostly negative.

Some studies focused on achieving an overall opinion of service, while others had more specific objectives. The study [1] carried out by Carmen Pérez Cabañero in 2020 aims to answer questions focused on the expertise of the reviewer, Sentiment Analysis of a tweet, and its content. One of these is strictly related to Sentiment Analysis: "Does the expertise of the reviewer influence the sentiment of his/her tweets?". They carried out text mining analysis on over 14'000 tweets collected on Twitter regarding Venice as a tourist destination. They discovered how the variables "Years", representing the years of activity on the platform, is the most influential one when it comes to the sentiment expressed by the posts. The more years the user had been using Twitter, the more negative the messages posted were.

Another study [2] carried out by Abdur Rasool in 2018, focused on finding the public opinion about Adidas and Nike and compare the positive and negative attitudes of common users about each brand. Analysing 100'000 tweets, they discovered Adidas had more positive reviews than Nike and how most of the data contained mostly comparison with other brands, especially among the mentioned ones.

Another interesting usage of Sentiment Analysis comes from research [3] whose objective was to analyse how Donald Trump's behaviour on Twitter changed over time. They focused their analysis on three periods: Campaign, Transition, and Presidency. Overall, they analysed how the number and sentiment of the tweets varied during the aforementioned periods in the different hours of the day. They discovered many interesting facts about the tweeting behaviour of the current President of the USA, in particular, his tweets are neutral overall, with a slight edge toward more positive tweets, but the tweets that received the highest number of retweets were negative!

In [5] the authors focused on Brexit and alongside an in-depth analysis of the whole phenomena, they utilized Sentiment Analysis to identify whether the tweets could be classified as 'pro-Leave' or 'pro-Remain' and their polarity. They discovered how the tweets on the "New referendum request" topic were mainly from the 'pro-Remain' side and they were mainly negative, while tweets on the "Cabinet" topic were mainly coming from the 'pro-Leave' and tweets on the "Irish border issue" were mainly positive. They also extracted the public stance and the sentiment from Brexit-related topics and the politicians' Twitter accounts.

In recent times, due to the outbreak of the virus COVID-19, it has become really important to be able to track the spreading of diseases. At https://covid19obs.fbk.eu/ an Italian group of researchers from Trento, developed a dashboard that performs data analysis on comments from social media and datasets from publicly available repositories. Among all of the performed analyses, Sentiment, Emotional, and Psychological Analysis of the collected comments were performed, representing their evolution over time. From the dashboard, it is possible to notice how the overall sentiment was negative during the first months of the pandemic, while after a few months it became positive or neutral.

4. Sentiment Analysis modules

Nowadays there are a lot of different libraries and tools which provide pre-trained models for Sentiment Analysis. In general, since Sentiment Analysis is a part of NLP (Natural Language Processing), these libraries do not only provide features for Sentiment Analysis. Indeed, they come with a lot of built-in NLP functionalities.

Some of the most known ones are:

- spaCy, a free open-source Python library for Natural Language Processing. It features Named Entity Recognition (NER), Part-Of-Speech (POS) tagging, dependency parsing, word vectors, and more.
- Natural Language ToolKit (a.k.a. NLTK), a suite of Python libraries and programs for symbolic and statistical NLP on data in English.
- TextBlob, a Python library for processing textual data. It provides a simple API for diving into common NLP tasks such as part-of-speech tagging, noun phrase extraction, Sentiment Analysis, classification, translation, and more.
- Stanford Core NLP, which enables users to derive linguistic annotations for text, including token and sentence boundaries, parts of speech, named entities, numeric and time values, dependency and constituency parses, coreference, sentiment, quote attributions, and relations.

A particular case is “Bidirectional Encoder Representations from Transformers” (a.k.a. BERT) from Google. It is not a library. It is a method of pre-training language representations, which aims at training a general-purpose "language understanding" model on a large plain text corpus, and then use that model for NLP tasks, like Sentiment Analysis. BERT outperforms previous methods since it is the first unsupervised, deeply bidirectional system for pre-training NLP. Pre-trained representations can also either be context-free (like word2vec or GloVe) or contextual (like ELMo). The latter category of models, before the introduction of BERT, included only unidirectional or shallowly bidirectional models, meaning that the context of a word was only taken from either one of the sides of a word. Instead, BERT is deeply bidirectional and utilizes both the left and right contexts to contextualize a word, which improves the overall performance of the model. One of the libraries that currently provides an implementation of the BERT architecture is “Transformers” by Hugging Face (https://github.com/huggingface/transformers).

Most of the mentioned libraries directly provide a pre-trained model through which perform Sentiment Analysis, making it easy and quick. Indeed, it is enough to load the library and invoke the specific function, passing the text to analyze. On the other hand, it is possible to either fine-tune these models using the collected data (this is the case with BERT), improving its performance in a specific context, or train a whole new model from scratch. The latter is seldomly recommended, because of the computational resources and the amount of data required to perform the training. For example, spaCy provides a free course named “Advanced NLP with spaCy”, available at https://course.spacy.io/en/, in which they teach you the basics of language processing and models.

4.1. Applications of Sentiment Analysis in PERSEUS

In order to have satisfactory results, the data on top of which Sentiment Analysis is ran should be chosen carefully. Textual data that does not express subjective content should not be considered as an input for the Sentiment Analysis component. An example of this is the EurLex dataset. As such, the data used for these analyses comprises:

- Tweets
- Media texts (news articles)
- User generated content from COCTEAU (comments and keywords)

A more thorough explanation of this is given in deliverable D7.4, “Guidance on how to use AI-enabled Sentiment Analysis for Deep Dives”.

The collected data and outcomes of the analyses will be available in dedicated data visualizations for policymakers to explore, improving their understanding of manifold research questions like the ones related to the Deep Dives.

4.2. Linking Sentiment Analysis and Actorness

A key objective of TRIGGER is to develop new ways to analyze the EU’s actorness. The consortium is therefore exploring different ways for how to combine the method of sentiment analysis with the theory of actorness. Only by creating a link between method and theory, can we improve our understanding of actorness.

The TRIGGER model for actorness defines actorness via seven dimensions: authority, autonomy, coherence, recognition, attractiveness, credibility and opportunity/necessity to act (see D3.1, D3.2 and the deep dives in WP7). The credibility dimension refers to the external perception of the EU as a reliable and trustworthy actor by its international partners. This dimension can be analysed in several different ways both qualitatively and quantitatively – one of them is quantitative text analysis and sentiment analysis.

We argue that sentiment analysis can provide a rough proxy for the EU’s perceived credibility. When an actor is mentioned in a mostly positive or negative context in, for example, international media articles, we assume that this can be interpreted as a rough proxy for high or low credibility and trust in public perception.[6]

[6] Please note that no text source can be understood as a perfect representation of public perception. Strictly speaking, media articles only represent the publicized perspective of the respective newspapers. In addition, many news articles contain more factual descriptions and

It is important to note that existing lexicon or machine learning based sentiment analysis methods are not specifically designed for measuring credibility, but more generally for measuring positive or negative sentiment. Our research in this domain is therefore an experimental combination of these data science techniques with a social science concept. The following paragraphs showcase our current approach using the example of our news media analysis for the deep dives.

Our news media analysis designed to analyse the EU's perceived credibility follows several steps: (1) data collection; (2) data preparation; (3) data analysis; (4) interpretation and validation.

First, we **collect news articles** from the leading news media in the G7 and BRICS countries. The list of newspapers was manually defined and comprises more than 69 different newspapers. The newspapers were selected based on two primary criteria: their relevance in terms of circulation and their political orientation to ensure a representative view of the respective media landscape. The topic of the articles is defined via a search query in the Google search engine such as "sustainable development" or "data protection". The Google search engine than returns the links to the articles and their title and text can be downloaded directly from the website of the respective newspaper.

Second, we **prepare the raw news media texts** for analysis. The articles are split into separate sentences and only sentences that mention the actor(s) of interest are maintained. The unit of analysis for our sentiment analysis are therefore individual sentences, as opposed to entire articles. This is primarily done in order to facilitate attributing positive or negative sentiment to a specific actor. If sentiment analysis is applied to the entire article, it is unclear to which among the many actors in the article the positive/negative sentiment refers. By applying sentiment analysis only on a sentence in which e.g. "the EU" is mentioned, we can be more confident that the positive/negative sentiment refers to the EU. Please note, however, that it is not possible to discern with certainty, whether positive/negative sentiment refers specifically to the EU or another actor event mentioned in the same sentence. The method only allows us to say that the EU is mentioned in a positive or negative sentence, which we use as a proxy for the sentiment towards the EU. This method is not 100% accurate for each individual sentence, but individual misclassifications can be outweighed in aggregate if enough sentences are analyzed.

Third, after splitting each article into sentences, we **apply sentiment analysis** to all sentences that mention at least one actor from a predefined list of actors. This list of actors contains the EU and all G7 and BRICS countries. We apply sentiment analysis not only to the EU, but also the G7 and BRICS in order to create a reference point for defining 'comparatively positive' or 'comparatively negative' sentiment. If we only analysed the sentiment towards the EU, the algorithms could e.g. return a mean sentiment of 0.5. This number in itself is, however, not very

---

sentences loaded with sentiment need to be identified. In a similar vein, Twitter data only represent the views of twitter users, which overrepresent certain socio-economic groups. [6]

meaningful as we would not have a reference point to which we could compare this number. By computing the sentiment score of the EU and all G7 and BRICS countries, we can understand the sentiment scores for leading powers in general and can then compare whether the EU's score is comparatively low or high. Based on this analysis, we can create a ranking of sentiment scores for leading powers (see Table 1).

Sustainable Development (experimental)			
Rank	Selected entities	Mean_sentiment_score_2010-2020	Media sentiment score[7]
1	Canada*	0.23	5
2	USA*	0.23	5
3	EU*	0.23	5
4	South Africa*	0.21	4.5
5	China*	0.21	4.5
6	Germany*	0.21	4.5
7	Japan*	0.19	3.9
8	UK*	0.16	3.1
9	France*	0.16	3.1
10	Brazil*	0.13	2.3
11	India*	0.12	2.1
12	Italy*	0.11	1.8
13	Russia*	0.08	1

Table 1. Sentiment analysis scores for articles on sustainable development

Please note that this research is currently in the experimentation stage

Fourth, we have to **interpret and validate the output** of the sentiment analysis algorithm. The analysis in table 1, for example, was created with an open-source sentiment lexicon called VADER[8], which was not specifically designed for analyzing news media sentiment and trust/credibility. We therefore have to read a sample of sentences and their respective sentiment score to test whether the sentiment score actually represents the sentiment towards the respective actor and provides a meaningful proxy for trust and credibility. These experiments will continue throughout the TRIGGER project.

The key advantage of this media sentiment analysis pipeline is its reproducibility, scalability and flexibility. Once the code for the pipeline is written, it only takes a few days to analyze several

[7] The sentiment scores can be translated to a scale from 1 to 5 with a 'Min-Max-Scaler'. See details here: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html
[8] https://github.com/cjhutto/vaderSentiment

thousand news articles on any topic, across dozens of news media and spanning over an entire decade. This pipeline is currently being developed for the four deep dive topics (SDGs, data protection, climate change and EU-Africa partnership/development aid) but will be transferrable to other topics for which digital news articles are publicly available. It is important to note, however, that these methods do not offer the same accuracy as qualitative text analyses and they can only be understood as a rough proxy for social science concepts like 'credibility'.

The use value for policy makers or other users lies in this flexibility and scalability. The pipeline can produce a sentiment analysis on many topics and many actors in a comparably short period of time. It can produce indicative answers to questions such as "What is the global media sentiment towards the EU when it comes to climate change?", "How does the sentiment towards the EU compare to the sentiment towards the US or China?", or "How did the sentiment evolve over the past 10 years?". As mentioned above, there are, however, several limitations and hurdles to take.

4.3. Limitations

Due their experimental nature, PERSEUS, AGGREGATOR and COCTEAU are still under active development by different teams within the consortium and we are still collecting evidence for their effectiveness. Some preliminary tests have been carried out internally while more meaningful tests will be carried out during the last year of TRIGGER. These will highlight strengths and weaknesses of the aforementioned platforms, enabling an iterative refinement process. The results will eventually be presented in the last deliverable related to PERSEUS (D8.7).

5. Conclusions

In this document, an overview of PERSEUS and an in-depth explanation of Sentiment Analysis and software libraries that support it have been provided. Sentiment Analysis is a powerful tool that can contribute to decision-making in many different ways, also providing interesting insights. Many tools are currently available online, most of which are free and their effectiveness in PERSEUS has yet to be demonstrated. Therefore, further investigations are needed before being able to tell which library is better to be applied in PERSEUS.

6. Bibliography

[1] Cabañero, Carmen & Bigne, Enrique & Mafe, Carla & Cuenca, Antonio. (2020). Sentiment Analysis of Twitter in Tourism Destinations. 1-3. 10.4995/CARMA2020.2020.11621.

[2] Rasool, Abdul & Tao, Ran & Kamyab, Alimarjan & Naveed, Tayyab. (2019). Twitter Sentiment Analysis: A Case Study for Apparel Brands. Journal of Physics: Conference Series. 1176. 022015. 10.1088/1742-6596/1176/2/022015.

[3] Ouyang, Yu & Waterman, Richard. (2020). Trump Tweets: A Text Sentiment Analysis. 10.1007/978-3-030-44242-2_4.

[4] Taj, Soonh & Meghji, Areej & Shaikh, Baby. (2019). Sentiment Analysis of News Articles: A Lexicon based Approach.

[5] Calisir, Emre & Brambilla, Marco. (2018). Characterizing Long-Running Political Phenomena on Social Media.

[6] Wojcik, Stefan & Hughes, Adam. (2019). Sizing Up Tiwtter Users. Pew Research Centre. Available online: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

trigger

Trends In Global Governance and Europe's Role

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 822735. This document reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.