*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870811*

Social cohesion, Participation, and Inclusion through Cultural Engagement

# D3.4 Final semantic annotator

| Deliverable information | |
|---|---|
| WP | WP3 |
| Document dissemination level | Public |
| Deliverable type | Final |
| Lead beneficiary | MAIZE |
| Contributors | MAIZE |
| Date | 28/04/2022 |
| Document status | **Final** |
| Document version | 1.0 |

INTENTIONALLY BLANK PAGE

# Project information

**Project start date:** 1st of May 2020

**Project Duration:** 36 months

**Project website:** https://spice-h2020.eu

## Project contacts

**Project Coordinator**
**Silvio Peroni**
ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
Department of Classical Philology and Italian Studies – FICLIT
E-mail: silvio.peroni@unibo.it

**Scientific Coordinator**
**Aldo Gangemi**
Institute for Cognitive Sciences and Technologies of the Italian National Research Council
E-mail: aldo.gangemi@cnr.it

**Project Manager**
**Adriana Dascultu**
ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
Executive Support Services
E-mail: adriana.dascultu@unibo.it

## SPICE consortium

| No. | Short name | Institution name | Country |
|---|---|---|---|
| 1 | UNIBO | ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA | Italy |
| 2 | AALTO | AALTO KORKEAKOULUSAATIO SR | Finland |
| 3 | DMH | DESIGNMUSEON SAATIO - STIFTELSEN FOR DESIGNMUSEET SR | Finland |
| 4 | AAU | AALBORG UNIVERSITET | Denmark |
| 5 | OU | THE OPEN UNIVERSITY | United Kingdom |
| 6 | IMMA | IRISH MUSEUM OF MODERN ART COMPANY | Ireland |
| 7 | GVAM | GVAM GUIAS INTERACTIVAS SL | Spain |
| 8 | PG | PADAONE GAMES SL | Spain |
| 9 | UCM | UNIVERSIDAD COMPLUTENSE DE MADRID | Spain |
| 10 | UNITO | UNIVERSITA DEGLI STUDI DI TORINO | Italy |
| 11 | FTM | FONDAZIONE TORINO MUSEI | Italy |
| 12 | MAIZE | MAIZE SRL (previously CELI SRL) | Italy |
| 13 | UH | UNIVERSITY OF HAIFA | Israel |
| 14 | CNR | CONSIGLIO NAZIONALE DELLE RICERCHE | Italy |

# Executive summary

Spice Semantic Annotator (SSA) is an annotation service for the semantic enrichment of textual contents, targeting user generated contents as well as descriptions of museum artifacts. The service is multilingual and supports English, Finnish, Hebrew, Italian and Spanish. It consists of a natural language processing pipeline that performs Sentiment Analysis, Emotion Detection and Entity Linking.

SSA analyses textual contents collected from museum visitors interacting with the activities scripted in the interfaces (WP5) and realized for the different use cases (WP7). The service annotates contents with respect to the ontological models developed in WP6 and generates as output an RDF graph to be stored in the linked data hub developed by WP4. Such analysis puts the visitor at the centre by interpreting and then enhancing his point of view and contributes to:

- the process of defining profiles of each visitor in order to build Community Models (the profiles and models are generated by task 3.1).
- the design of an advanced recommendation engine (task 3.3)

The novel aspect of the Semantic Annotator lies in the multilingual Emotion Detection component for the Art domain that combines state-of-the-art AI models with language specific domain knowledge. The rule-based system relies on language specific knowledge (i.e., sentiment/emotion lexicons associating linguistic expressions to sentiment/emotions) and it doesn't require any representative dataset. AI models instead allow for tailoring the system to the domain, jargon and style of final users; however, they require representative datasets for each language.

The analysis performed by SSA makes it possible to focus on the visitors, their thoughts, cultural and social context, emotional inclinations so to enhance their role in the curatorial process, both as individuals and as part of a community (or more communities). It also allows for retrospective social studies on how the same type of content can produce different emotions and polarities and, also, how the same emotion or object interpretation is instead shared by people belonging to different communities. It follows that this kind of Semantic Annotation represents a fundamental contribution in putting the visitors at the centre of the curatorial process.

# Document History

| Version | Release date | Summary of changes | Author(s) –Institution |
|---|---|---|---|
| V0.1 | 07/03/2022 | First draft released | Alessio Bosca, Andrea Bolioli |
| V0.2 | 14/04/2022 | Firs revision released | Alessio Bosca, Andrea Bolioli |
| V0.3 | 27/04/2022 | Second Revision released | Alessio Bosca, Andrea Bolioli |
| V0.4 | 28/04/2022 | Final Version released | Alessio Bosca, Andrea Bolioli |
| V1.0 | 28/04/2022 | Version submitted to REA | UNIBO |

# Table of Contents

# 1 Introduction

This deliverable document describes the final release of SPICE Semantic Annotator (SSA). This component is a service for the analysis of curatorial products and the identification of relevant textual information (e.g., sentiment/emotions detection, entities recognition). It is used in order to semantically enrich curatorial contents with metadata that links textual fragments to concepts described in a knowledge graph.

WP3 includes 4 closely related components:

1. individual and community models (existing models in figure 1), which are the data structures that contain information about individuals and communities (concepts taken from WP6 ontologies) and stored in the linked-data hub (included in D3.3);
2. user modeller (circled in yellow in figure 1) and community modeller (circled in green in figure 1) which are the reasoning mechanisms that monitor the users continuously, reason about their behaviour and infer their preferences and community relatedness and update the models accordingly (included in D3.3);
3. **SSA textual content analysis** (detailed in this document) and
4. a recommender system (D3.6) that uses the user models and **scripts** (guidelines/instructions for activities, generated by WP6) for guiding the process of content recommendation to users.

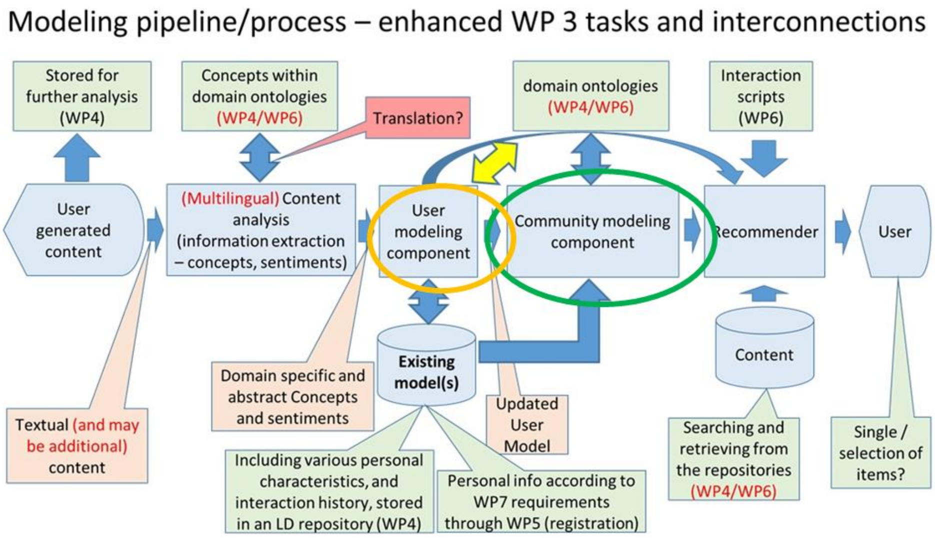


*Figure 1: The user modeller and the community modeller and the internal and external interaction within WP3 and of WP3 with other WPs. The user modeller is circled in yellow; the community modeller is circled in green. The user and community models are stored in the LDH and the modellers continuously reason and update them. The analysed user generated content is used as an input and the user and community models are used by the recommender.*

The overall goal of SSA consists in providing tools and algorithms to enrich user generated textual contents and to provide User and Community modelling components (WP3) with semantic features that leverage the models defined in the SPICE Ontology Network (WP6) and represented as Linked Data. Textual contents originate from museum visitors interacting with the activities scripted in the interfaces (WP5) and realized for the different use cases (WP7). The contents enriched with the extracted semantic features are stored in SPICE Linked Data Hub (WP4).

SSA uses a structured representation for the extracted features (referring to concepts in a knowledge graph) in order to enable abstraction and reasoning over them. More specifically the result of the annotation process consists in the automatic creation of metadata enriching the document (or specific fragments of it) with identifiers of concepts and entities mentioned in the text or relevant to it. Such references link the textual contents to the formal description of concepts/entities in a knowledge graph, and allow for further reasoning over the latter. In the context of the SPICE project, reasoning over such semantic annotation allows for abstracting and generalizing inputs coming from museum visitors, finding commonalities between them and ultimately supporting the activity of users and communities modelling and the design of an advanced recommendation engine. More details can be found in **D3.1.1: *Prototype user and community modelling***.

SSA therefore is a key component within the SPICE infrastructure, since it provides a connection between the contents, coming from the User interfaces, developed in WP5 (more details on User Interfaces can be found in **D5.1.1: *Preliminary interfaces for interpretation***), and SPICE knowledge graph, designed in WP6 (more details on the knowledge graph can be found in **D6.3.1: *Initial ontology network specification***). The semantic annotations produced by this component are stored in the linked data hub developed in WP4 (more details on the Linked Data Hub infrastructure can be found in **D4.1 *Linked Data server technology: requirements and initial prototype***).

In WP4 SSA API was used as part of the dashboard for citizen curation activities analytics, further detailed in D4.1, and this may work as a *de facto* evaluation in a real-world application.

The semantic annotation of curatorial products (i.e. contents generated by the users visiting the museums and engaged in reflection/interpretations activities) is triggered by the interfaces used in the different museum use cases and its result are then used by **User/Community modelling tools** (3.1) and the **Recommender system** (3.3).

The Natural Language Processing (NLP) analysis pipeline of SSA includes the following components:

- Sentiment Analysis,
- Emotion Detection
- Entity Linking;

It is multilingual and supports all the languages used in the museum use cases: English, Finnish, Hebrew, Italian and Spanish. Semantic annotation is performed on the native language, while the annotations representing the structured features extracted from text are expressed in English.

The components providing Sentiment Analysis and Emotion Detection were specifically designed and developed for the project, while the components for the basic language analysis (e.g., lemmatization, PoS tagging) and entity linking were implemented reusing available Open-Source resources and models:

- Language analysis is performed using Stanza[1], a Python natural language analysis package developed by the Stanford NLP Group, exploiting neural networks models built on top of the Pytorch ecosystem.
- Entity linking is performed using ML models from DBpedia Spotlight[2], a solution for linking unstructured information sources to the Linked Open Data cloud through DBpedia.

More details on these resources and how they are used in the analysis pipeline can be found in Deliverable document D3.2 (describing the first implementation of SSA).

The main innovation point of this component consists in the multilingual Emotion Detection for the Arts domain, exploiting state-of-the-art AI algorithms. The application of this integrated analysis to the Arts domain, interpreting and analysing visitors' individual points of view and feelings, represents a step forward in the state of the art of semantic annotation. The conjunction of Language Analysis, Sentiment Analysis, Emotion Detection and Entity Linking allows to obtain a complex analysis of users/visitors' curatorial products provided with complete metadata; these structured metadata features are used to build detailed, structured profiles of each visitor (User Modelling 3.1) and, by having these, to create Community Models (3.1) and to support the action of the Recommendation System (3.3).

[1] https://stanfordnlp.github.io/stanza/
[2] https://www.dbpedia-spotlight.org/

The in-depth and multilevel analysis of the curatorial products really makes it possible to focus on the visitors, their thoughts, cultural and social context, emotional inclinations, so to enhance their role in the curatorial process, both as individuals and as part of a community (or more communities); it also allows for retrospective social studies on how the same type of content can produce different emotions and polarities and, also, how the same emotion or object interpretation is instead shared by people belonging to different groups. It follows that this kind of Semantic Annotation represents a fundamental step in putting the visitor at the centre of the curatorial process.

Task T3.2 main objectives for the second year of the project include:

- the **Art & Emotions Experiment** was designed and implemented in order to gather test data in addition and in advance to the data collected in the museum use cases. Such dataset was annotated in order to train and test the Deep Leaning component for Emotion and Sentiment detection.
- a **Deep Learning component for Emotion and Sentiment classification**. Annotating with respect to emotions and sentiment the dataset collected in the Art & Emotions experiment along with data from museum use cases and using it to train and test the Neural Network.
- a rule-based component for **entity detection**, in order to handle entities that are **relevant** to the use cases but **not present in DBPedia** (e.g., artwork titles, artist names)
- Integration of SSA service with the **LDH platform** (WP4) and populating specific datasets for the different WP7 use cases with the analysis results in JSON-LD format.

The rest of the document is structured as follows: Section 2 presents the Art & Emotions Experiment and the dataset collected, then Section 3 presents SSA general architecture, the integration of the new components as well as the results of a performance evaluation of the Emotion and Sentiment detection components. Finally, Section 4 presents the API details as well as the output format of the annotations and a few usage examples.

# 2 *Art & Emotions Experiment*

This section describes the Art & Emotions Experiment realized in collaboration with GAM museum in order to obtain user generated data in all the languages of the SPICE project use cases and use it for training the Deep Learning component for Emotion and Sentiment detection. The collected dataset constitutes as well a source of data for investigating the relationship between art and emotions and can contribute to research in the arts domain. For this experiment we created and shared a form using Google Form that can be accessed online at:

https://docs.google.com/forms/d/e/1FAIpQLScC0vh33NfSKWVPrQjBBzQAlnjcyTKDZ0Wc8Ui1geUQrpN5jQ/viewform

The experiment is based on the GAM collection and consists of 12 artworks, chosen from a group of artworks previously provided by the museum. Each artwork is presented in a different section of the form; for each of the artworks, the user is asked to answer two open questions:

1. *"What do you see in this picture? Write what strikes you most in this image"*
2. *"How does this artwork make you feel? Write your feelings, emotions, thoughts"*

The user is then asked to select one or more emojis; a list of some main emojis is provided as choices and/or there is the possibility to click on "other" and enter other emojis from Emojipedia through a link provided in the question. For each of the artworks, the user can decide whether to skip to the next artwork, if he does not like the one in front of him or go back to the previous artworks and modify the answers.

The question about emotions is left open so as not to force the person to choose emotions from a list of tags which are the tags of a model (e.g., Plutchik), but leaving him free to express the different shades of emotions that can be felt. Before getting to the heart of the experiment, with the artworks sections, the user is asked to leave some personal information (anonymously), to help us getting an idea of the type of users who participated in the experiment.

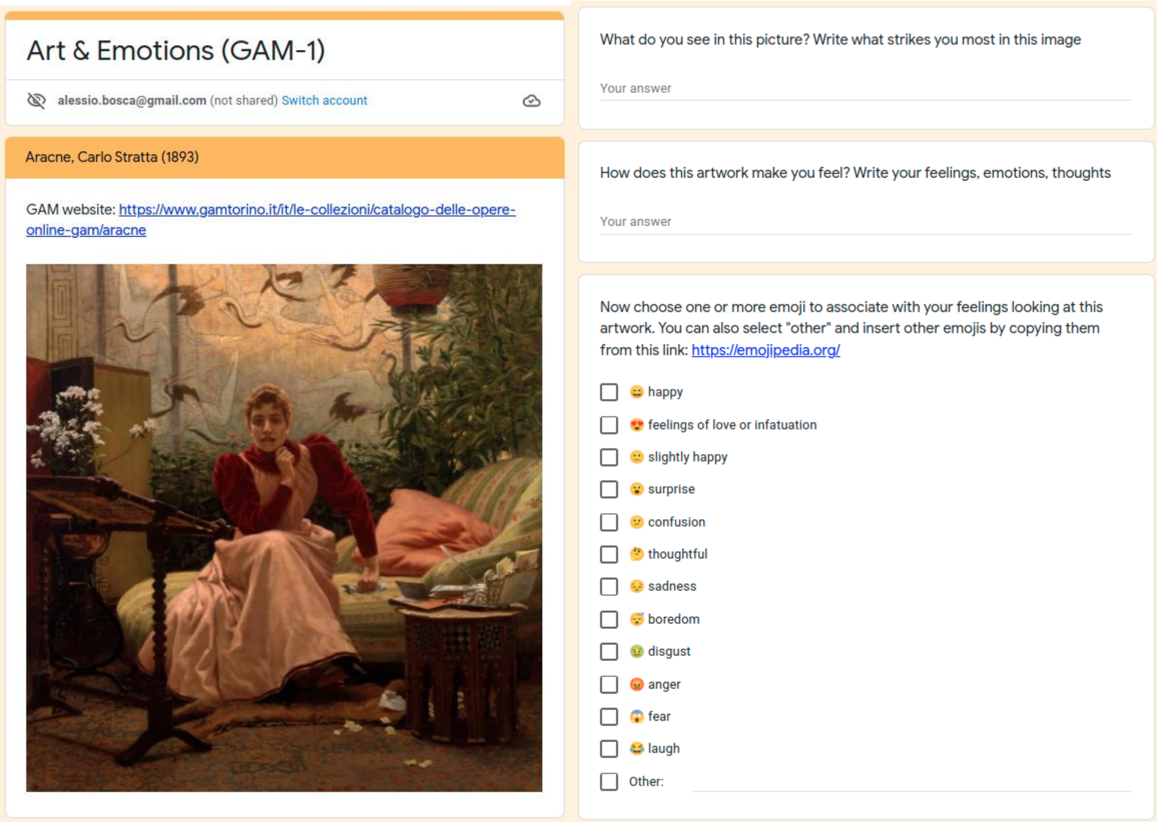


*Figure 1. The final Art & Emotions Experiment Form*

The (optional) questions related to personal information are:

1. Age (open)
2. Gender (male, female, other)
3. How would you define your relationship with art? (Multiple choices allowed)
   - My job is related to the art world
   - I am passionate about the art
   - I am a little interested in art
   - I am not interested in art
4. Do you like going to museums or art exhibitions?
   - I like to visit museums frequently
   - I go occasionally to museums or art exhibitions
   - I rarely visit museums or art exhibitions

The experiment was proposed in all the five languages of SPICE, thanks to the help of the project partners in translating the contents. A different form has been created for each language, maintaining the exact same structure.

It was initially decided to ask three questions:

1. *What do you see in this picture? Write what strikes you most in this image.*
2. What does this work of art make you think of? Write down the thoughts and memories the picture evokes
3. *How does this painting make you feel? Write the feelings and emotions that the picture evokes in you*

After a first round of dissemination (between May and July 2021), once the answers were obtained and read, we decided to reduce the questions to two, combining the question about thoughts and the one about emotions into one, resulting in the question: "*How does it make you feel this work? Write your feelings, emotions, thoughts* ", to avoid redundancy in the answers and lighten the load of requests for the user.

## 3.1 Art & Emotions Experiment Dataset

The dataset collected through the experiment includes:

- 422 answers in English
- 137 answers in Finnish
- 251 answers in Hebrew
- 238 answers in Italian
- 148 answers in Spanish

An example of the collected data is presented in Table 1; all the examples presented in the table refer to the same artwork: "Aracne" from Carlo Stratta - GAM Collection (see Figure 1).

| **What do you see in this picture? Write what strikes you most in this image** | **How does this artwork make you feel? Write your feelings, emotions, thoughts** | **Choose one or more emoji to associate with your feelings looking at this artwork.** |
|---|---|---|
| I see the girl. And her eyes, her powerful glimpse | Curious: I want to know more about the woman | 😍 |
| A woman in a richly decorated room (with items coming from various countries) who has just destroyed a letter. | Determination | 😮 🤔 |
| Birds in the background | Inquisitive | 🙂 |
| Her position, she seems thinking intensely to something and her look is mysterious | Absorbed, observed, intense, slightly scared | 😮 🤔 😱 |
| torn letter, face expression, flying birds in the background, the specific feel of the light ... | like a pause, inner anguish, after something has happened... the biting lip, supressed emotion | 😮 😕 |

*Table 1. Art & Emotions Dataset example*

132 different (anonymous) users provided some personal information along with the open questions. The following charts present details on personal data values distribution.

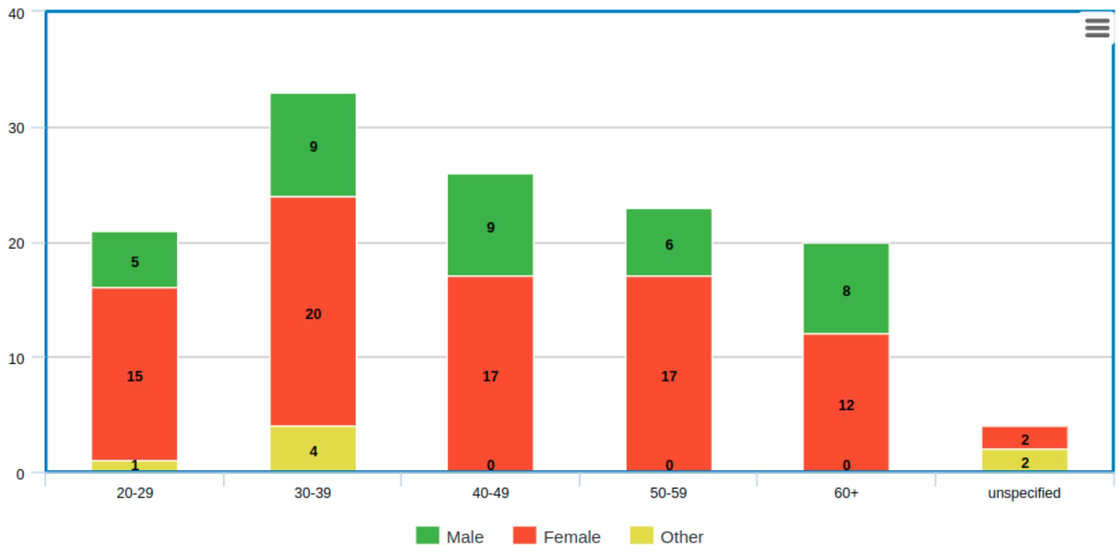


*Figure 2. Gender by Age group distribution*

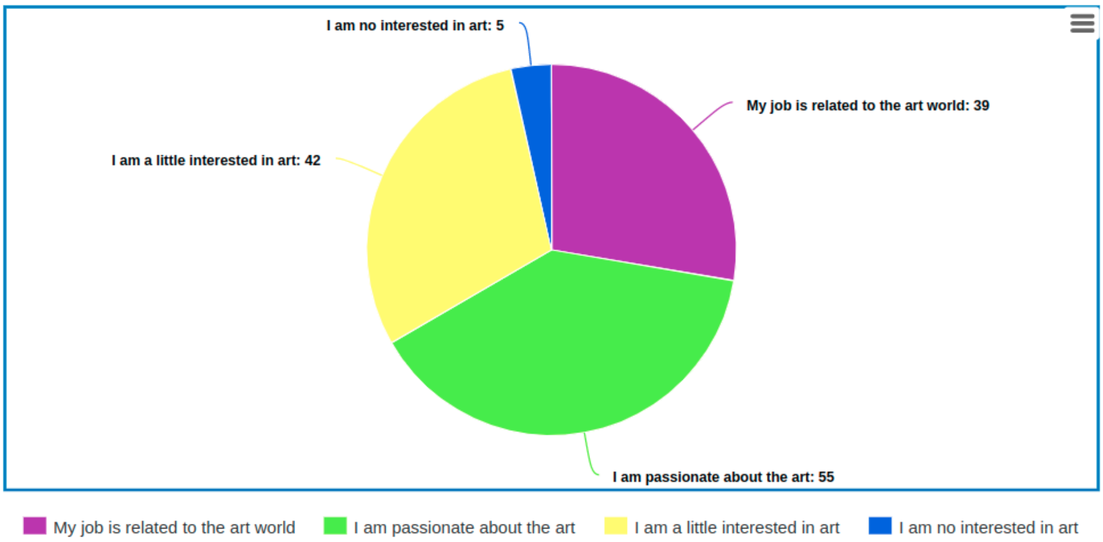


*Figure 3. Relationship with art distribution*

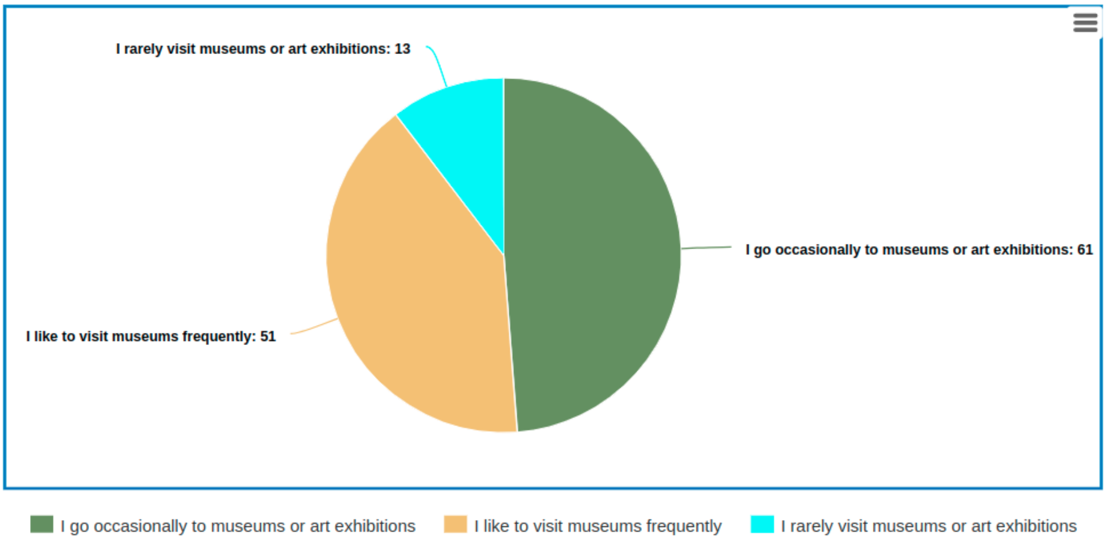


*Figure 4. Museums Visit Frequency distribution*

The dataset collected was manually annotated in order to train and evaluate of the Deep Learning models for Emotion and Sentiment detection; more details on the training process are presented in subsection 3.1.3. The dataset is available in the official project repository and during the third year of the project we are planning on harmonizing the data coming from the different rounds of dissemination and then publishing it in an open data repository, like Zenodo[3].

# 3 SPICE Semantic Annotator Architecture

This section presents the final architecture of the Semantic Annotator and the interaction between the different analysis components. The initial architecture presented in the Deliverable document D3.2 was updated in order to integrate:

- a micro-service for sentiment and emotions detection, trained on textual contents generated by museum visitors using a Deep Learning architecture;
- a microservice for entities detection targeting domain relevant entities not present in DBPedia (and not manged by DBPedia Spotlight models);
- SSA service with the LDH, automatically uploading analysis results

The initial architecture was updated, as well, in order to reduce response times by removing Message Queues (between the Orchestrator and the analysis components) and replacing them with direct API calls (from the Orchestrator to the different components).

The process of semantic annotation is realized by a **Natural Language Processing Pipeline** that includes different analysis modules, each one responsible for annotating the document with respect to a specific aspect: sentiment analysis, emotion detection, entity linking. The overall process is exposed by means of standard **RESTful[4] APIs** and produces a JSON-LD[5] document as output. **JSON-LD** is a JSON-based serialization for Linked Data that can be seamlessly stored in the Linked Data hub of WP4.

[3] https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/Zenodo
[4] https://www.w3.org/2001/sw/wiki/REST
[5] https://www.w3.org/TR/json-ld11/

The architecture (graphically represented in Figure 1) is designed to be modular and configurable in order to allow new analysis components to be included, or to easily replace any of them and experiment with different algorithms and models.

The RESTful service acts as the entry point of the annotation process. It receives as input the textual contents to be analysed along with some metadata (i.e., the contents language and the collection); textual contents and metadata are wrapped into a document object along with an analysis plan (detailing the different modules that should process the contents and in which order) and the document is then submitted to the pipeline. An orchestration component within the pipeline is responsible for forwarding the document to the correct analysis modules; when the analysis plan is completed, the textual content has been **semantically enriched** by the different components and formatted as a JSON-LD document, and it is finally returned to the RESTful service and provided as output.

After SSA service response, a background process **feeds** such **JSON-LD** document to the **Linked Data Hub**; a specific dataset for each collection (representing the museum use cases defined in WP7) is used to store the analysis results as RDF data. Such RDF data can be obtained from LDH (by means of SPARQL queries or requests to LDH APIs) and then used in order to train AI models (as the Group/Community models or the Recommendation System) or in order to perform retrospective social studies on how the same type of content can produce different emotions and polarities and, also, how the same emotion or object interpretation is instead shared by people belonging to different groups.

The whole pipeline is designed following a **Microservice Architecture**[6] approach in order to isolate and decouple the different analysis modules implementing them with different technologies (e.g. Java, Python, R), and exploiting a wide variety of models and solutions available on the open source. The pipeline is deployed as a Microservice Architecture on a Kubernetes[7] cluster with the replication of the analysis components managed by KEDA[8].

- **Kubernetes** is an open-source system for automating deployment, scaling, and management of containerized components (e.g., Docker[9] images).
- **KEDA** instead is a single-purpose and lightweight component that can be added into any Kubernetes cluster and acts as a Kubernetes-based Event Driven Autoscaler; with KEDA it is possible to configure the scaling (up and down) of any container in Kubernetes based on the number of events needing to be processed.

Such architectural solutions were chosen in order to achieve horizontal scalability. In particular, we (1) increase the instances of a given component when the number of documents waiting to be processed exceeds a certain threshold and decrease their number when they go below the threshold. This approach allows us to **ensure service response time** regardless of the system workload and to r**educe economic and energetic costs** by dismissing computational resources when they are not needed.

The whole system is deployed to AWS[10] cloud resources, on servers located in the European region.

[6] Salah, Tasneem, et al. "The evolution of distributed systems towards microservices architecture." *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*. IEEE, 2016.

[7] https://kubernetes.io/

[8] https://keda.sh/

[9] https://www.docker.com/

[10] https://aws.amazon.com/

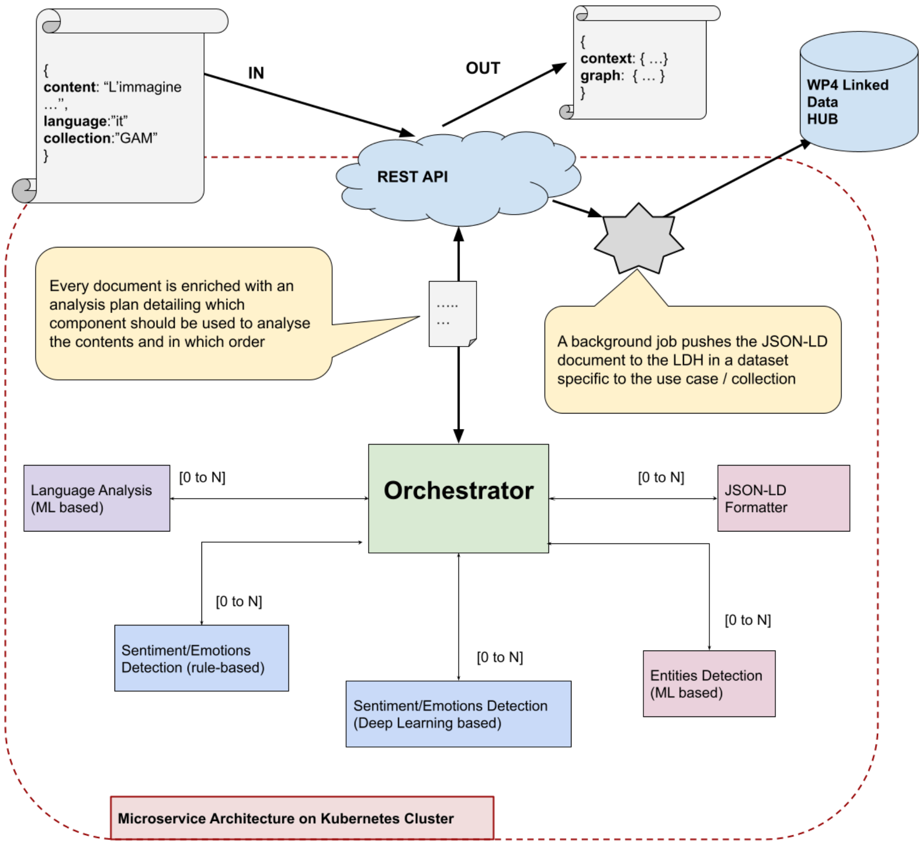


*Figure 5. Semantic Annotator Architecture*

The analysis pipeline represented in Figure 1 enriches the original data with the following components:

- **Language Analysis**, with the goal of performing standard language analysis on the contents (e.g., lemmatization, PoS tagging). Such analysis will be exploited by other components (as the Emotion Detection and Sentiment Analysis).
- **Emotion Detection**, with the goal of detecting textual expressions that can be linked to emotions, referencing emotions from the Plutchik Emotion ontology from WP6. The emotions supported in this component are:
    - Anger, Anticipation, Disapproval, Disgust, Fear, Interest, Joy, Love, Sadness, Serenity, Surprise, Trust
- **Sentiment Analysis**, with the goal of detecting textual expressions that carry a subjective information (e.g., like and dislike statements) along with its polarity: positive, negative or neutral. The conceptual framework used to model sentiment polarity is the MARL[11] ontology; MARL is a standardised data schema designed to annotate and describe subjective opinions expressed on the web or in particular Information Systems.
- **Entity Linking**, with the goal of detecting textual expressions that can be linked to relevant concepts and named entities in order to obtain a representation of the semantics of the contents through the detection of named entities and their types. Since the topics of user generated contents (as well as the subjects of museums use cases) cannot be restricted to a specific domain we decided to use DBPedia[12] as the target conceptual framework. In order to handle entities that are relevant to the use case but not part of DBPedia (i.e., artworks or artists names, collections items) such entities are

[11] http://www.gsi.dit.upm.es:9080/ontologies/marl/
[12] https://www.dbpedia.org/

configured specifically for each use case (e.g., by extracting them from the LDH by means of use case specific SPARQL queries) and added to the component knowledge base.

∉ **JSON-LD** formatter, with the goal of formatting the NLP pipeline results as RDF, using JSON-LD a JSON-based serialization for Linked Data, with the explicit semantic representation of contents referencing the SPICE ontology network defined in WP6.

The following sub-sections detail on the components added or updated with respect to the first version of Spice Semantic Annotator (described in the Deliverable document D3.2) specifically:

- Sentiment / Emotion Detection components (Deep Learning and Rule Based models)
- Entity Detection
- Integration with SPICE Linked Data Hub

## 3.1 Sentiment / Emotion detection

During the second year of the project:

- the **Rule Based component** for Sentiment and Emotion detection, built on the multilingual lexicon resource created in the first year of the project (described in D3.2) has been **revised and updated** from the first-year baseline to a final version;
- an **AI model** for Emotion and Sentiment detection has been **trained** leveraging the data collected in the Art & Emotions Online Experiment (see Section 2) combined with the preliminary SPICE use cases datasets.

The following subsections present a brief review on the related works for Emotions detection, then describes the activities related to these components followed by an experimental evaluation of the two components.

### 3.1.1 Related Works

Since recent years, a major constraint to Emotion Detection from written text has been the difficulty of extracting emotional signals from small collections of labelled data (Alswaidan and Menai, 2020; Acheampong et al.,2020).

Traditional approaches to Emotion and Sentiment Detection include **Lexicon based approaches** that use one or more lexical resources, like a lexicon or an ontology, linking words to emotions and sentiment values (Mohammad, Saif, 2013 or ) Some approaches exploits these resources by means of **pattern based linguistic rules** (Strapparava, Carlo, 2008 or Shaheen, Shadi, 2014) while others works use Latent Semantic Analysis, a statistical approach for analysing the relationships between a set of documents and the terms mentioned in these documents (Gill, Alastair et al., 2008). Other approaches for Emotion Detection from textual contents involve Machine Learning (ML) techniques. Machine learning is a scientific discipline that deals with the construction and study of algorithms that can learn from data. In particular, supervised learning approaches rely on a labelled training data, algorithms analyse the training data and infer a function, used for mapping new examples. (Balabantaray 2012) presents an Emotion classifier based on multi-class SVM kernels that targets the basic emotions identified by Ekman; in this work the authors automatically collect a dataset from Twitter by filtering tweets using a lexicon resource (containing a list of terms related to each of the basic emotions) and then manually labelled them. The SVM model trained on such data achieved an accuracy level of 73%.

In recent years, the introduction of new language representation models has come to play a central role in machine learning approaches to natural language processing tasks (Devlin et al., 2019). A **Language Model** (LM) assigns probabilities to a sequence of words and is a crucial component in NLP applications such as machine-translation and information extraction (e.g., a translation system might generate multiple translations of the same target sentence and the LM scores all the sentences to pick the one that is most likely). In the last years, the **Deep Learning** era has brought new neural LMs (as Bert[13] or GPT-3[14]) that have outperformed the traditional statistical ones in almost all NLP tasks. **Deep Learning Neural Language Models** are pretrained on very large corpora of textual data (typically extracted from the web) on unsupervised tasks as predicting the next word in a text or filling the blank; they have big learning capacity (e.g., hundreds of millions of parameters) and use novel training algorithms (attention networks).

There are several benefits in using a pretrained model, but the most important one is the possibility of fine-tuning it on a specific task with a (relatively) small amount of domain-related data. Such models are **capable of generalizing/abstracting** the meaning of terms or their usage patterns (thanks to the LM learned in the pretraining) and thus require a **smaller amount of annotated data** in supervised tasks (like classification, entity extraction, etc). Thanks to the capability of abstracting and generalizing contents, this type of LM is suitable for dealing with contents coming from **users** with **different language skills** (e.g., native speakers, non-native speakers, kids, tourists, etc) and is an effective solution for **harmonizing linguistic differences** between the users' groups.

Another important benefit comes from **Multilingual Neural Language Models** in which the tokens from different languages share the same embedding space (i.e., numeric representation of tokens used in the NN) thus the experience (annotated data) learned in one language will be exploited as well in the other languages, leveraging the transfer learning capabilities of the model. This is an important benefit in a multilingual context like SPICE where the number of annotated resources can be different between the project languages.

Recently, larger datasets like EmoNet or GoEmotions have allowed to train neural models that outperform their traditional counterparts based on lexicons (Abdul-Mageed and Ungar, 2017). After the release of BERT (Devlin et al., 2019), an explosion of novel work has focused on fine-tuning transformer models to learn from scarce emotion data. In this direction, GoEmotions introduced a fine-tuned BERT multi-label classifier baseline with **46%** macro-F1 across **28 possible labels** (Demszky et al., 2020).

Despite the benefits coming from approaches based Deep Learning Neural Language Models solutions based on lexical resources are still relevant especially when the data available for fine-tuning are even more scarce, the results of lexicon-based methods may still outperform the more modern deep neural network-based methods (Catelli, Rosario et al., 2022). Another advantage of Lexicon based approaches consists in presenting results that are **easier to interpret and explain** (as they identify textual snippets related to the detected emotions) while Neural Network models have the drawback of not being human-interpretable, raising various problems related to model's explainability. Very few works so far have been proposed to build models that explain their decision-making process (Zucco, Chiara, et al., 2018).

Given these considerations, some works propose **hybrid approaches** that combines the two approaches (Acheampong, Francisca Adoma, et al. 2020 or Pamungkas, Endang Wahyu et al. 2019). In SPICE Semantic Annotator pipeline both approaches have been implemented with specific components.

### 3.1.2 Rule-Based Component for Sentiment / Emotion detection

During the second year of the project the multilingual lexicon for Sentiment and Emotion detection has been revised and updated from the first-year baseline to a final version.

The revision process consisted in confronting Art & Emotions Dataset annotated data with the results from the analysis performed by the Sentiment and Emotions rule-based components. Error analysis for the examples not correctly categorized was performed, identifying and fixing bugs in the code or errors in the

[13] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
[14] Brown, Tom, et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

lexicon and rules (e.g., a lexicon entry misspelled or ambiguous, an error in the rule for handling negations...). A final version of the component was released at the end of this process

### 3.1.3 Multilingual Deep Learning Component for Sentiment / Emotion detection

During the second year of the project, an AI model for Emotion / Sentiment detection in the Arts domain based on a pre-trained Deep Learning LM has been trained leveraging the **data collected in the Art & Emotions Online Experiment** (see Section 2) combined with the **preliminary SPICE use cases datasets** (e.g., the transcription of IMMA and Hecht visitors' scripted interviews) along with some data from the GoEmotions[15] public dataset. We included GoEmotions data (that is outside the Arts domain) for training the models in order to handle emotions that were under-represented in the Art & Emotions and SPICE use cases datasets (as Love or Disgust); in the course of the third year of the project, as more data will become available from Museum use cases, we will update the model by retraining it only on the project data.

Our goal in this task consist in deploying a solution **in line with the current state of the art adapted to the Arts domain** by means of the training/test dataset.

| Collection | Language | Documents Number |
|---|---|---|
| Online Art Emotions | FI | 141 |
| Online Art Emotions | EN | 422 |
| Imma Viewpoints | EN | 999 |
| GoEmotions sample | EN | 200 |
| Online Art Emotions | ES | 148 |
| Online Art Emotions | HE | 252 |
| Hect Experiment | HE | 30 |
| Online Art Emotions | IT | 237 |

*Table 2. Annotated Dataset Composition*

Around two thousand documents were manually analyzed and annotated. Only a subset of the whole collection of documents were actually enriched with manual annotations since not all the documents contain reference to emotions or sentiment. On the other hand, some documents could contain references to more than one emotion as the task of Emotion detection is inherently a **multiclass multilabel problem** (each document can be annotated with 0 or more emotions).

Figure 7 presents the distribution of the number of documents manually associated to the different emotions and sentiment classes. The distribution is not uniform as some emotions were less frequent in the dataset (as Love or Disgust) and in order to cope with such underrepresented classes we included some documents from GoEmotions dataset. Such annotated data was used in order to train the neural networks models used to categorize documents with respect to emotions and sentiment.

[15] Demszky, Dorottya, et al. "GoEmotions: A dataset of fine-grained emotions." *arXiv preprint arXiv:2005.00547* (2020)


*Figure 7. Annotated Classes Frequencies*

The iterative process of model creation can be divided in four phases:

- **Annotation**: textual data is manually labelled with respect to the possible emotions and sentiment values; it is randomly split (keeping classes proportions consistent) in 80% and 20% between training data and evaluation data (plus 100 examples for a test dataset)
- **Training**: training labelled data is used for fine-tuning the pretrained model and train a multiclass classifier,
- **Evaluation**: the trained model is validated against the evaluation labelled data and performance measurements (precision, recall, F1 score) are produced for each class,
- **Error Analysis**: reviewing the performances of the classes and accordingly select raw data for a new annotation phase (e.g., increasing the examples for low-performing classes or for better specifying the semantics of two overlapping classes).

These four phases were repeated a few times in order to improve overall model performance.

The iterative model creation process was performed through Sophia Analytics[16] platform, a commercial solution by MAIZE for Text Analytics and Data Mining. The model created with Sophia Analytics was finally exposed as a microservice in the SSA architecture (see Figure 5 on the analysis pipeline architecture).

[16] https://www.celi.it/en/products/sophia-semantic-engine/


*Figure 6. Model Creation Process*

The pretrained LM we adopted is ***bert-multilingual-base-cased***[17] from **TensorFlow**[18] repository. This Bert model is a 12 stacked encoder with a hidden size of 768 nodes.

The Bert model (pre-trained on a large amount of data from over 100 languages) has been fine-tuned using SPICE annotated data in order to create two models:

- Sentiment detection: a multi-class classifier (one class for each document at most) with respect to positive, negative and mixed classes.
- Emotions detection: multi-label classifier (one or more classes for each document) with respect to Anger, Anticipation, Disapproval, Disgust, Fear, Interest, Joy, Love, Sadness, Serenity, Surprise, Trust classes

The hyperparameters used by Sophia Analytics for training the models adopts the default values suggested in TensorFlow guidelines for Bert fine tuning[19] and consist in:

- Batch size: 16
- Learning Rate: 2e-5
- Learning Rate Warm-Up: 0.1
- Optimizer: Adam optimizer with weight decay
- Max Sequence Length: 128
- Training epochs: 5

The training process implemented by Sophia Analytics splits the annotated dataset in 80% for training and 20% for validation. At the end of each epoch the model is evaluated on the evaluation set and at the end of the 5 training epochs the best model is selected for the predictions. The following figure presents the overall

[17] https://tfhub.dev/google/bert_multi_cased_L-12_H-768_A-12/1
[18] https://www.tensorflow.org/
[19] https://www.tensorflow.org/text/tutorials/fine_tune_bert

F1 score trend (average over all classes) for the Emotion detection model along the iterative model creation process.

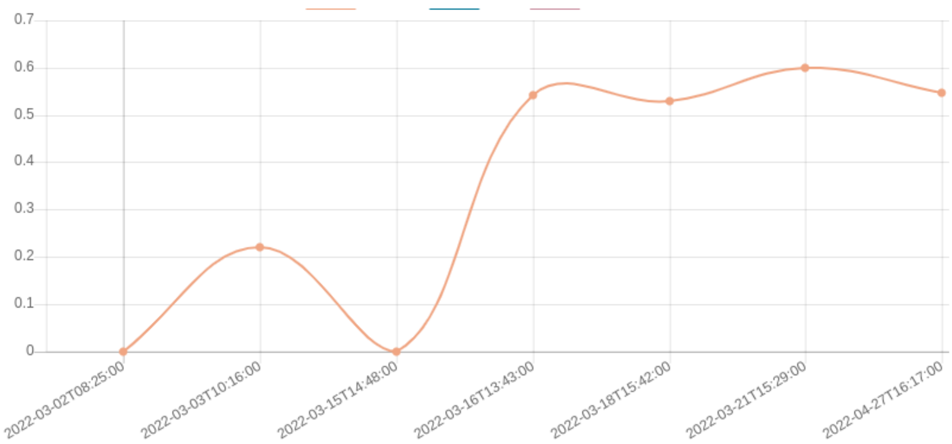


*Figure 7. F1 score trend during Emotion detection Model Creation Process*

### 3.1.4 Combining Rule Based approach with Deep Learning models

The emotions and sentiment annotations coming from the Rule Based and the Deep Learning components within SSA NLP pipeline are joined together, therefore SSA analysis results contain the union of the annotations from the two components.

### 3.1.5 Evaluation

The evaluation was performed on a test dataset of 100 documents extracted from the manually annotated data and not included in the model creation process (see Section 3.1.3). Since the model is selected by evaluating its F1-score against the validation set, best practices suggest to use a whole separate dataset to provide an unbiased evaluation of a final model. The size of such test dataset is small, but the starting pool of documents included only about 2000 documents and further shrinking of the training/validation dataset would be detrimental for the models' creation.

Since we expect an improvement of the ML models performances with the addition of training / evaluation data coming from the use cases, we plan to produce (before the end of the project, as an addition to the current document) a final evaluation test of SSA components in order to integrate and update the experimental evaluation presented in the current Deliverable document. For that round of evaluations, we intend to increase the sizes of all the 3 datasets (training, validation and test).

The dataset was used to evaluate both the Rule Based component and the Deep Learning models. It is included in the TECHNICAL ANNEX B of this document.

#### 3.1.5.1 Rule Based Component Evaluation

An evaluation of the baseline and the final systems can be found in the following tables, detailing respectively sentiment and emotion detection. Each table reports the number of True Positives (i.e., correct predictions), False Positives (i.e., wrong predictions), False Negatives (i.e., missed predictions), Precision, Recall and F1-

Score for each class. A weighted average for the different classes is computed in order to have a single KPI for the sentiment and emotion detection components.

The **baseline** system consists of the Rule Based component released after the first year of the project while the **final** system represents the component after the improvements/updates during the second year (see subsection 3.1.2)

| Class | True Positives | False Positives | False Negatives | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| positive | 25 | 1 | 7 | 0.96 | 0.78 | 0.86 |
| negative | 18 | 7 | 13 | 0.72 | 0.58 | 0.64 |
| mixed | 0 | 11 | 8 | 0.0 | 0.0 | 0.0 |
| Weighted Average of Precision, Recall and F1-score over | | | | **0.68** | **0.59** | **0.63** |

*Table 3. Sentiment Evaluation Metrics – BASELINE lexicon*

| Class | True Positives | False Positives | False Negatives | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| positive | 29 | 9 | 3 | 0.76 | 0.9 | 0.82 |
| negative | 28 | 7 | 3 | 0.8 | 0.9 | 0.85 |
| mixed | 7 | 1 | 1 | 0.87 | 0.87 | 0.87 |
| Weighted Average of Precision, Recall and F1-score over | | | | **0.79** | **0.89** | **0.83** |

*Table 4. Sentiment Evaluation Metrics – FINAL lexicon*

The final version of the Sentiment detection (rule-based) module shows an increase in its KPIs with respect to the baseline:

- Precision from **0.68** to **0.79**;
- Recall from **0.59** to **0.86**.
- F1-score from **0.63** to **0.823**

The main reason for the performance increase in the Sentiment detection component originated from an update of the baseline version, fixing a bug in the sentiment rules handling the negation expressions (that are responsible for inverting the polarity of a sentiment expression: good -> positive; not good -> negative).

| Class | True Positives | False Positives | False Negatives | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| Anger | 2 | 11 | 0 | 0.15 | 1.0 | 0.26 |
| Anticipation | 8 | 8 | 14 | 0.56 | 0.39 | 0.46 |
| Disapproval | 0 | 4 | 3 | 0.0 | 0.0 | 0.0 |
| Disgust | 2 | 6 | 2 | 0.25 | 0.5 | 0.33 |
| Fear | 10 | 6 | 7 | 0.62 | 0.59 | 0.6 |
| Interest | 3 | 9 | 7 | 0.25 | 0.3 | 0.27 |
| Joy | 12 | 14 | 5 | 0.46 | 0.7 | 0.55 |
| Love | 2 | 3 | 3 | 0.4 | 0.4 | 0.4 |
| Sadness | 9 | 11 | 6 | 0.45 | 0.6 | 0.51 |
| Serenity | 8 | 2 | 10 | 0.8 | 0.44 | 0.57 |
| Surprise | 3 | 6 | 2 | 0.33 | 0.6 | 0.37 |
| Trust | 2 | 4 | 4 | 0.33 | 0.33 | 0.33 |
| Weighted Average of Precision, Recall and F1-score | | | | **0.38** | **0.49** | **0.43** |

*Table 5. Emotion Detection Evaluation Metrics – BASELINE lexicon*

| Class | True Positives | False Positive | False Negatives | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| Anger | 2 | 5 | 0 | 0.29 | 1.0 | 0.45 |
| Anticipation | 12 | 6 | 12 | 0.66 | 0.52 | 0.58 |
| Disapproval | 2 | 8 | 1 | 0.2 | 0.66 | 0.31 |
| Disgust | 4 | 2 | 0 | 0.66 | 1.0 | 0.79 |
| Fear | 15 | 8 | 2 | 0.65 | 0.88 | 0.75 |
| Interest | 9 | 11 | 1 | 0.45 | 0.81 | 0.58 |
| Joy | 13 | 4 | 4 | 0.76 | 0.81 | 0.78 |
| Love | 5 | 2 | 0 | 0.71 | 1.0 | 0.83 |
| Sadness | 13 | 10 | 2 | 0.56 | 0.86 | 0.68 |
| Serenity | 16 | 2 | 2 | 0.88 | 0.88 | 0.88 |
| Surprise | 3 | 2 | 2 | 0.6 | 0.6 | 0.6 |
| Trust | 3 | 1 | 3 | 0.75 | 0.5 | 0.6 |
| Weighted Average of Precision, Recall and F1-score | | | | **0.65** | **0.77** | **0.71** |

*Table 6. Emotion Detection Evaluation Metrics – FINAL lexicon*

The final version of the Emotion detection (rule-based) module shows an increase in its KPIs with respect to the baseline:

- Precision from **0.38** to **0.65**;
- Recall from **0.49** to **0.77;**
- Precision from **0.43** to **0.71**

The main reason for the performance increase in the Emotion detection component originated from an update/revision of the lexicon.

### 3.1.5.2 Deep Learning models Evaluation

An evaluation of the Deep Learning models can be found in the following tables, detailing respectively sentiment and emotion detection. Each table reports the number of True Positives (i.e., correct predictions), False Positives (i.e., wrong predictions), False Negatives (i.e., missed predictions), Precision, Recall, F1-Score and the number of examples in the dataset for each class. A weighted average for the different classes is computed in order to have a single KPI for the sentiment and emotion detection components.

| Class | True Positives | False Positives | False Negatives | Precision | Recall | F1-score | N. of examples in the dataset |
|---|---|---|---|---|---|---|---|
| positive | 29 | 1 | 3 | 0.95 | 0.91 | 0.92 | *309* |
| negative | 28 | 3 | 3 | 0.9 | 0.9 | 0.9 | *286* |
| mixed | 1 | 7 | 7 | 0.12 | 0.12 | 0.12 | *108* |
| Weighted Average of Precision, Recall and F1-score | | | | **0.66** | **0.64** | **0.65** | |

*Table 7. Sentiment Evaluation Metrics*

| Class | True Positives | False Positives | False Negatives | Precision | Recall | F1-score | N. of examples in the dataset |
|---|---|---|---|---|---|---|---|
| Anger | 2 | 3 | 0 | 0.4 | 1.0 | 0.57 | *59* |
| Anticipation | 8 | 0 | 14 | 1.0 | 0.36 | 0.52 | *123* |
| Disapproval | 1 | 2 | 2 | 0.33 | 0.33 | 0.33 | *63* |
| Disgust | 2 | 0 | 2 | 1.0 | 0.5 | 0.67 | *61* |
| Fear | 11 | 6 | 6 | 0.65 | 0.65 | 0.65 | *108* |
| Interest | 4 | 4 | 6 | 0.5 | 0.4 | 0.44 | *152* |
| Joy | 12 | 5 | 5 | 0.7 | 0.7 | 0.7 | *128* |
| Love | 1 | 1 | 4 | 0.5 | 0.2 | 0.29 | *59* |
| Sadness | 11 | 4 | 4 | 0.73 | 0.71 | 0.72 | *161* |
| Serenity | 10 | 4 | 8 | 0.71 | 0.55 | 0.62 | *142* |
| Surprise | 2 | 1 | 3 | 0.67 | 0.4 | 0.5 | *87* |
| Trust | 5 | 1 | 1 | 0.83 | 0.82 | 0.82 | *79* |
| Weighted Average of Precision, Recall and F1-score | | | | **0.65** | **0.56** | **0.6** | |

*Table 8. Emotion Detection Evaluation Metrics*

Although the overall performances of these models are in line with similar experiments in literature (i.e., GoEmotions paper from Demszky, Dorottya, et al. or Few shot knowledge transfer from Olah, Justin, et al.) they present significative performances variations across emotions (e.g., Disapproval or Love). This is probably caused by the differences in the number of documents associated to those classes. However, since Neural Network models are a black box, explaining the reason behind their predictions is not possible.

These shortcomings can probably be improved by adding new examples in the dataset and retraining the models. During the third year of SPICE project, we expect an improvement of the AI models performances (with the addition of training and validation data from the use cases) therefore we plan to produce, before the end of the project, an integration to the current document, with a final evaluation test of SSA Deep Learning models.

## 3.2 Entities Detection

The Entities Detection module used by the semantic annotator is based on the open-source models from DBpedia Spotlight[20], a well-known Open-Source library for automatically annotating mentions of DBPedia entities within a textual document - an online demo to try the models is available at https://demo.dbpedia-spotlight.org/. Pretrained models for several languages are available on the project page, including:

- English, Finnish, Italian and Spanish.

For the Hebrew language DBpedia Spotlight models are not available and we used DBpedia based models from Wikifier[21]

[20] https://www.dbpedia-spotlight.org/

[21] https://wikifier.org/

The entity recognition module provides for each detected entity a unique ID and one or more types. For instance, in the sentence:

- *This picture reminds me of the Mona Lisa*

the Entities Detection module identifies "Mona Lisa" as an entity and returns as output its ID (dbr:Mona_Lisa) and 3 types:

- http://dbpedia.org/ontology/Person,
- http://dbpedia.org/ontology/Work,
- http://dbpedia.org/ontology/Artwork

Some entities, however, might not be captured by the models either because they are not part of DBPedia. In order to customize Entity Detection and better adapt it to the different use cases a rule-based component (following the same approach detailed for the Lexicon based Emotion Detection component described in Deliverable document D3.2) was integrated in order to handle the use case specific relevant entities non included in DBPedia.

During a preliminary evaluation of SSA service it emerged the need to integrate the recognition of custom entities in the Entity Detection module; entities that are relevant for the use case but not present in DBPedia. Most of such custom entities are present in the metadata of museum artifacts included in the use cases collections (as artwork titles, collections' items names, artists names, subjects represented in the artwork, etc).

Whenever such information is present in the LDH repository it can be accessed and retrieved by means of specific SPARQL queries. An example query for extracting catalogue items labels (based on IMMA LDH Dataset[22]) can be found in Table 4.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <http://schema.org/>


SELECT ?uri ?label ?type
WHERE {
 graph ?any {
  ?uri a ?type .
  { ?uri schema:caption ?label } UNION { ?uri rdfs:label ?label }
 }
}
```

*Table 4. Custom Entities extraction from LDH - SPARQL query example*

The labels retrieved by means of SPARQL queries are then indexed using Lucene[23], a well-known search engine, and then used as a lookup resource for entity detection. In order to identify the labels (and slight variations of them) within a textual document (e.g., comments, answers to scripted activities) a sliding window of n-grams is extracted from user generated contents and then used as queries in order to perform a composite search over the index (combining a search over the normal forms, the lemmatized/stemmed

[22] https://spice.kmi.open.ac.uk/dataset/details/41
[23] https://lucene.apache.org/

forms and a fuzzy search[24] based on Levenshtein distance) in order to identify NER candidates, following the same approach described in Bosca, Alessio, et al. (2014) for Cross-Language Information Retrieval.

An experimental evaluation of the Entities Detection module was not possible at this stage of the project because the experimental datasets collected so far (as the Art Online Experiment, IMMA Viewpoints, GAM game) contain, at the moment, few mentions to Entities. This might be related to the script used in these experiments, involving direct questions about specific items from a collection (see section 2 on Arts & Emotions experiment).

## 3.3 Integration with LDH

During the second year of the project SSA has been fully integrated with LDH. A background process is in charge to feed the Linked Data Hub with the JSON-LD response document of each textual content analyzed by SSA; thus, enabling retrospective social studies by the curators on how the same type of content can produce different emotions and polarities and, also, how the same emotion or object interpretation is instead shared by people belonging to different groups.

In the LDH, a specific dataset for each museum is used to collect all users' generated contents related to a specific use case. One of the parameters of SSA API consists of a label for the collection of the contents to be analyzed (see section 4 for more details on SSA APIs). If the value of the collection parameter refers to one of the museum use cases, then the JSON-LD document is saved in a use case specific dataset, otherwise a fallback test dataset is used.

The following table details the museum specific collections along with the relative dataset UUID; the fallback test dataset details are also reported at the end of the table.

| **Collection - Museum Use Case** | **Dataset UUID in LDH** |
|---|---|
| IMMA | *b3631f48-2657-4cd3-96fa-4887c6e0c63a* |
| GAM | *810d60a6-c7be-4299-be2e-c86d988f58ad* |
| HECHT | *4125ba0c-adbe-4b0b-a2ff-3a5dde29d088* |
| MNCN | *2ae73c0c-84ad-416c-b17b-23032a75f0ef* |
| DMH | *514c5676-2560-47a9-bab4-76ff42eb0b83* |
| test | *85c109bb-6090-4110-9422-79303183fae5* |

*Table 5. Collection to LDH UUID mapping*

The dataset UUID is needed to programmatically access the dataset either through the LDH APIs or by means of SPARQL queries (filtering the dataset by their annotations values or timestamp). The Data model used to represent information in the JSON-LD document is presented in Section 4.2.

# 4 Spice Semantic Annotator APIs

This section describes SSA API detailing about its input, output and usage. The service is exposed through standard REST API behind a Basic Authentication[25] scheme. The service can be accessed at the URL:

24 https://lucene.apache.org/core/8_0_0/core/org/apache/lucene/search/FuzzyQuery.html

25 https://tools.ietf.org/html/rfc7617

- https://sophia42-demo.aws.celi.it/***<LANGCODE>****/spice/analysis*

<LANGCODE> is a path parameter and it is used to specify the language content, the supported values are: **en, es, fi, it, he**

## 4.1 Service Input

The service can be accessed with:

- POST requests: accepting a json document as input, with the following properties:
    - **content**: *mandatory* - the textual contents to be analyzed
    - **ns_prefix**: *optional* - the prefix used for representing the textual content in the JSON-LD response document, default value is *"spice"*
    - **ns_uri**: *optional* - the URI of the ontology used for representing the textual contents in the JSON-LD document, default value is *"https://w3id.org/spice/resource/"*
    - **collection**: *optional* - a textual label representing the collection/museum/use case, default value is *"spice"*

An example API request to SSA service API, using curl[26]:

```
curl --user USR:PWD[27] -X POST https://sophia42-demo.aws.celi.it/en/spice/analysis  -H 'Content-Type: application/json' -d '{"content":"I love Picasso'\''s Guernica but I am absolutely terrified by the screaming horse!", "collection":"test"}'
```

*Table 6. SSA API request example via CURL*

The same request expressed in python, using the popular requests[28] lib:

```
import requests

def testService(text: str, lang: str) -> object:
  r = requests.post('https://sophia42-demo.aws.celi.it/'+lang+'/spice/analysis',
          json={"content":text, "collection":"test"},
          auth=('USR', 'PWD'))
  print(r.json())

if __name__ == "__main__":
  testService("I love Picasso's Guernica, but I am absolutely terrified by the screaming horse!", 'en')
```

*Table 7. SSA API request example via Python*

Please notice that USR and PWD **MUST** be substituted with a real authentication in order to access the API.

The response document for the previous example request can be found in the technical annex at the end of the current deliverable document.

[26] https://curl.se/
[27] USR and PWD **MUST** be substituted with a real authentication in order to access SSA APIs.
[28] https://docs.python-requests.org/en/latest/

## 4.2 Service Output

The Semantic Annotator exposes the NLP pipeline analysis results as a JSON-LD[29] document. JSON-LD is a method of encoding linked data using JSON. Linked Data is structured data which is interlinked with other data so it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs. More details on the Linked Data Hub designed and deployed by WP4 can be found in **D4.1** ***Linked Data server technology: requirements and initial prototype.***

The JSON-LD document contains two main sections:

- **Context**: detailing the ontologies used to describe data along with their prefix (used for compact notations in the graph section)
- **Graph**: containing a set of RDF triples represented as JSON objects; in our case the textual contents along with some metadata, followed by a set of annotations referencing the textual spans that can be linked to an emotion, a sentiment value or an entity (within DBPedia knowledge graph)

The following picture represents the service output for the input: *"I love Picasso's Guernica, but I am absolutely terrified by the screaming horse!"*


```
{ ⊟
   "@context":{ ⊟
      "spice":"https://w3id.org/spice/resource/",
      "owl":"http://www.w3.org/2002/07/owl#",
      "dbr":"http://dbpedia.org/resource/",
      "earmark":"http://www.essepuntato.it/2008/12/earmark#",
      "xsd":"http://www.w3.org/2001/XMLSchema#",
      "rdfs":"http://www.w3.org/2000/01/rdf-schema#",
      "dcterms":"http://purl.org/dc/terms/",
      "semiotics":"http://ontologydesignpatterns.org/cp/owl/semiotics.owl#",
      "emotion":"https://w3id.org/spice/SON/PlutchikEmotion/",
      "marl":"http://www.gsi.upm.es/ontologies/marl/ns#",
      "sentilo":"http://ontologydesignpatterns.org/ont/sentilo.owl"
   },
   "@graph":[ ⊞ ]
}
```


*Table 8. SSA JSON-LD output - @context section*

[29] https://json-ld.org/

```
{⊟
  "@context":{⊞},
  "@graph":[⊟
    {⊟
      "@id":"spice:sa_1646400673363",
      "@type":"earmark:StringDocuverse",
      "@language":"en",
      "dcterms:source":"spice:test",
      "earmark:hasContent":"I love Picasso's Guernica, but I am absolute
      "sentilo:scoreType":{⊟
        "@id":"marl:Positive",
        "sentilo:avgScore":0.5
      }
    },
    {⊞},
    {⊟
      "@id":"ex:anno_2_emotion_47-56",
      "@type":"earmark:PointerRange",
      "rdfs:label":"terrified",
      "semiotics:denotes":{⊟
        "@id":"ex:anno_2_emotion_47-56_fear",
        "@type":"emotion:Fear"
      },
      "earmark:refersTo":{⊞},
      "earmark:begins":{⊞},
      "earmark:ends":{⊞}
    },
    {⊞},
    {⊞},
    {⊟
      "@id":"ex:anno_5_entity_7-14",
      "@type":"earmark:PointerRange",
      "rdfs:label":"Picasso",
      "semiotics:denotes":{⊟
        "@id":"dbr:Pablo_Picasso",
        "@types":[⊟
          "http://dbpedia.org/ontology/Artist",
          "http://dbpedia.org/ontology/Animal",
          "http://dbpedia.org/ontology/Person",
          "http://dbpedia.org/ontology/Species",
          "http://dbpedia.org/ontology/Writer",
          "http://dbpedia.org/ontology/Eukaryote"
        ]
      },
      "earmark:refersTo":{⊞},
      "earmark:begins":{⊞},
      "earmark:ends":{⊞}
```


*Table 9. SSA JSON-LD output - @graph Section*

The main element of the graph section contains a unique identifier of the textual contents and the content itself. The following *PointerRange* elements specify character offsets (with the properties *earmark:begins* and *earmark:ends*) that identifies an expression within the text, while the property *semiotic:denotes* contains the semantic connotation of the element along with its value and type.

### 4.3 Online Testing

A simple web page allows to test the system without the need of a valid account (username and password) for SSA APIs. The test page is accessible at:

- [https://spice.saas.celi.it/](https://spice.saas.celi.it/)

The tool allows users to specify a language and enter a text, submit it to SSA and presents the result with a simple diagram that highlights textual fragments associated to emotions/sentiment expression or entities. The "**Raw**" tab presents the JSON-LD response document.

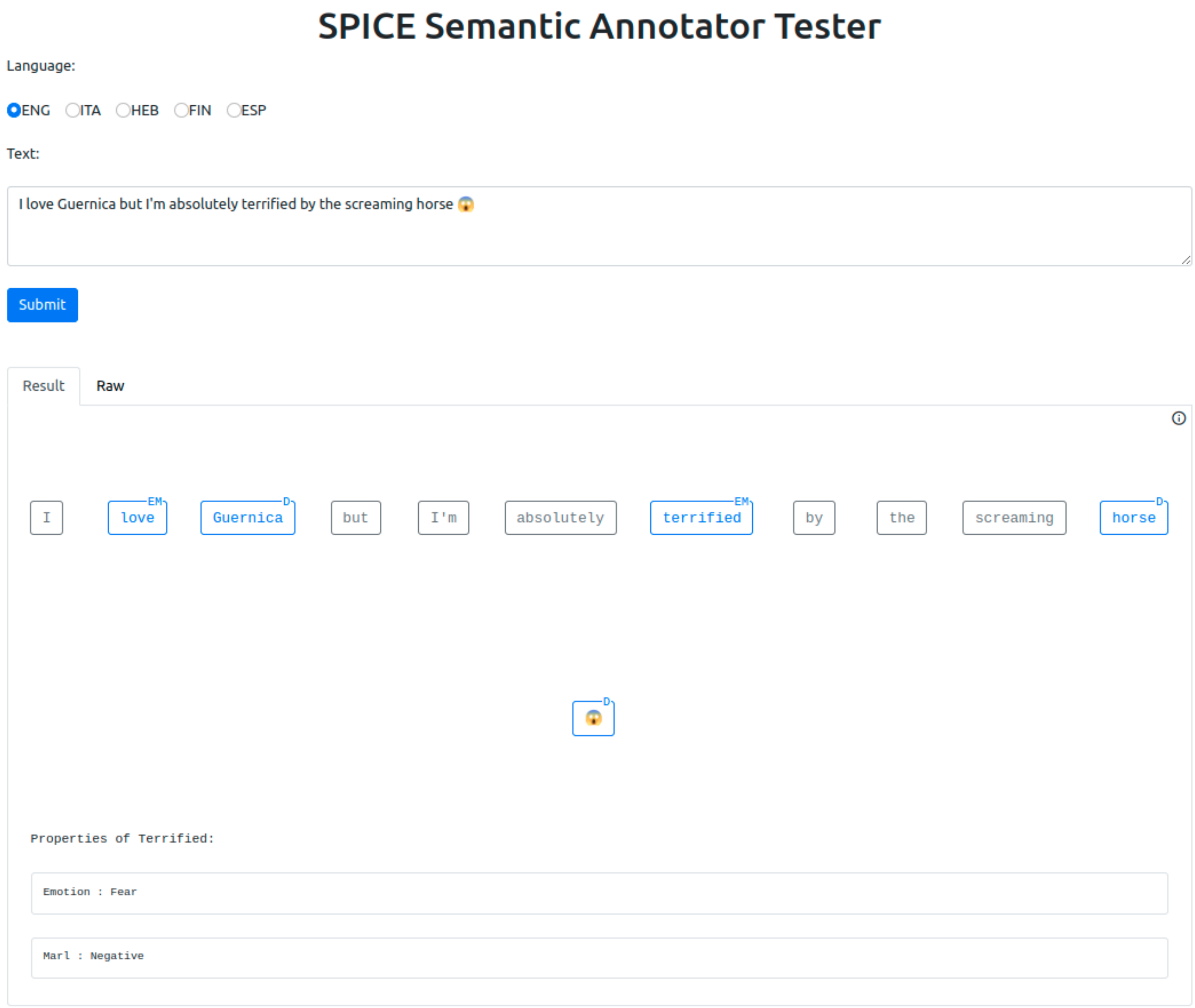


*Figure 7. SSA Test Interface*

# 5 Conclusions and future works

In the last year of the project, we will continue the revision, domain adaptation and refinement of SSA components:

- Sentiment & Emotions lexicons: with cycles of evaluation, error analysis and lexicon update,
- the Deep Learning models: adding data from museums' use cases in models train/test sets

- the Entity Linking module: by integrating new entities whenever required from the use cases.

An additional activity emerged in the context of the IMMA use case consists in developing and integrating in SSA a component for hate speech detection in order to filter out inappropriate contents (generated by museum visitors) and avoid to present them to other users. We plan to train an AI classifier (with the same approach and neural LM used for sentiment/emotion detection) leveraging the labelled dataset for hate speech that are available in the opens source (as the Hate Speech Dataset Catalogue[30] or HaSpeeDe [31]dataset).

Since we expect an improvement of the ML models performances with the addition of training and validation data from the use cases, we plan to produce an addition to the current document with a final evaluation test of SSA Deep Learning models in order to integrate and update the experimental evaluation presented in the current Deliverable document.

[30] https://hatespeechdata.com/
[31] http://www.di.unito.it/~tutreeb/haspeede-evalita18/index.html

# 6 References


- Salah, Tasneem, et al. "The evolution of distributed systems towards microservices architecture." 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST). IEEE, 2016.
- Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.
- Demszky, Dorottya, et al. "GoEmotions: A dataset of fine-grained emotions." arXiv preprint arXiv:2005.00547 (2020)
- Mendes, Pablo N., et al. "DBpedia spotlight: shedding light on the web of documents." Proceedings of the 7th international conference on semantic systems. 2011.
- Noraset, Thanapon, Chandra Bhagavatula, and Doug Downey. "WebSAIL wikifier at ERD 2014." Proceedings of the first international workshop on Entity recognition & disambiguation. 2014.
- Olah, Justin, et al. "Cross Domain Emotion Recognition using Few Shot Knowledge Transfer." arXiv preprint arXiv:2110.05021 (2021).
- Bosca, Alessio, et al. "Using semantic and domain-based information in clir systems." European Semantic Web Conference. Springer, Cham, 2014.
- Alswaidan, Nourah, and Mohamed El Bachir Menai. "A survey of state-of-the-art approaches for emotion recognition in text." Knowledge and Information Systems 62.8 (2020): 2937-2987.
- Abdul-Mageed, Muhammad, and Lyle Ungar. "Emonet: Fine-grained emotion detection with gated recurrent neural networks." Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers). 2017.
- Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- Demszky, Dorottya, et al. "GoEmotions: A dataset of fine-grained emotions." arXiv preprint arXiv:2005.00547 (2020)
- Mohammad, Saif M., and Peter D. Turney. "Nrc emotion lexicon." National Research Council, Canada 2 (2013).
- Strapparava, Carlo, and Rada Mihalcea. "Learning to identify emotions in text." Proceedings of the 2008 ACM symposium on Applied computing. 2008.
- Gill, Alastair J., et al. "Identifying emotional characteristics from short blog texts." 30th Annual Conference of the Cognitive Science Society. Washington, DC: Cognitive Science Society, 2008.
- Balabantaray, Rakesh C., Mudasir Mohammad, and Nibha Sharma. "Multi-class twitter emotion classification: A new approach." International Journal of Applied Information Systems 4.1 (2012): 48-53.
- Shaheen, Shadi, et al. "Emotion recognition from text based on automatically generated rules." 2014 IEEE International Conference on Data Mining Workshop. IEEE, 2014.
- Catelli, Rosario, Serena Pelosi, and Massimo Esposito. "Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian." Electronics 11.3 (2022): 374.
- Acheampong, Francisca Adoma, Chen Wenyu, and Henry Nunoo-Mensah. "Text-based emotion detection: Advances, challenges, and opportunities." Engineering Reports 2.7 (2020): e12189.
- Zucco, Chiara, et al. "Explainable sentiment analysis with applications in medicine." 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018.
- Pamungkas, Endang Wahyu, and Viviana Patti. "Cross-domain and cross-lingual abusive language detection: A hybrid approach with deep learning and a multilingual lexicon." Proceedings of the 57th annual meeting of the association for computational linguistics: Student research workshop. 2019.

# TECHNICAL ANNEX - A

JSON Response document to the example requested presented in Section 2:


```
{
   "@context":{
      "spice":"https://w3id.org/spice/resource/",
      "owl":"http://www.w3.org/2002/07/owl#",
      "dbr":"http://dbpedia.org/resource/",
      "earmark":"http://www.essepuntato.it/2008/12/earmark#",
      "xsd":"http://www.w3.org/2001/XMLSchema#",
      "rdfs":"http://www.w3.org/2000/01/rdf-schema#",
      "dcterms":"http://purl.org/dc/terms/",
      "semiotics":"http://ontologydesignpatterns.org/cp/owl/semiotics.owl#",
      "emotion":"https://w3id.org/spice/SON/PlutchikEmotion/",
      "marl":"http://www.gsi.upm.es/ontologies/marl/ns#",
      "sentilo":"http://ontologydesignpatterns.org/ont/sentilo.owl"
   },
   "@graph":[
      {
         "@id":"spice:sa_1646385943683",
         "@type":"earmark:StringDocuverse",
         "@language":"en",
         "dcterms:source":"spice:test",
         "earmark:hasContent":"I love Picasso's Guernica but I am absolutely terrified by
the screaming horse!",
         "sentilo:scoreType":{
            "@id":"marl:Positive",
            "sentilo:avgScore":0.5
         }
      },
      {
         "@id":"ex:anno_1_emotion_2-6",
         "@type":"earmark:PointerRange",
         "rdfs:label":"love",
         "semiotics:denotes":{
            "@id":"ex:anno_1_emotion_2-6_love",
            "@type":"emotion:Love"
         },
         "earmark:refersTo":{
            "@id":"ex:docuverse"
         },
         "earmark:begins":{
```

```
            "@type":"xsd:nonNegativeInteger",
            "@value":2
         },
         "earmark:ends":{
            "@type":"xsd:nonNegativeInteger",
            "@value":6
         }
      },
      {
         "@id":"ex:anno_2_emotion_46-55",
         "@type":"earmark:PointerRange",
         "rdfs:label":"terrified",
         "semiotics:denotes":{
            "@id":"ex:anno_2_emotion_46-55_fear",
            "@type":"emotion:Fear"
         },
         "earmark:refersTo":{
            "@id":"ex:docuverse"
         },
         "earmark:begins":{
            "@type":"xsd:nonNegativeInteger",
            "@value":46
         },
         "earmark:ends":{
            "@type":"xsd:nonNegativeInteger",
            "@value":55
         }
      },
      {
         "@id":"ex:anno_3_sentiment_2-6",
         "@type":"earmark:PointerRange",
         "rdfs:label":"love",
         "semiotics:denotes":{
            "@id":"marl:Positive"
         },
         "earmark:refersTo":{
            "@id":"ex:docuverse"
         },
         "earmark:begins":{
            "@type":"xsd:nonNegativeInteger",
            "@value":2
         },
         "earmark:ends":{
```

```
            "@type":"xsd:nonNegativeInteger",
            "@value":6
        }
    },
    {
        "@id":"ex:anno_4_sentiment_46-55",
        "@type":"earmark:PointerRange",
        "rdfs:label":"terrified",
        "semiotics:denotes":{
            "@id":"marl:Negative"
        },
        "earmark:refersTo":{
            "@id":"ex:docuverse"
        },
        "earmark:begins":{
            "@type":"xsd:nonNegativeInteger",
            "@value":46
        },
        "earmark:ends":{
            "@type":"xsd:nonNegativeInteger",
            "@value":55
        }
    },
    {
        "@id":"ex:anno_5_entity_73-78",
        "@type":"earmark:PointerRange",
        "rdfs:label":"horse",
        "semiotics:denotes":{
            "@id":"dbr:Horse",
            "@types":[
                "http://dbpedia.org/ontology/PersonFunction",
                "http://dbpedia.org/ontology/Animal"
            ]
        },
        "earmark:refersTo":{
            "@id":"ex:docuverse"
        },
        "earmark:begins":{
            "@type":"xsd:nonNegativeInteger",
            "@value":73
        },
        "earmark:ends":{
            "@type":"xsd:nonNegativeInteger",
```

```
        "@value":78
      }
    },
    {
      "@id":"ex:anno_6_entity_7-14",
      "@type":"earmark:PointerRange",
      "rdfs:label":"Picasso",
      "semiotics:denotes":{
        "@id":"dbr:Pablo_Picasso",
        "@types":[
          "http://dbpedia.org/ontology/Artist",
          "http://dbpedia.org/ontology/Animal",
          "http://dbpedia.org/ontology/Person",
          "http://dbpedia.org/ontology/Species",
          "http://dbpedia.org/ontology/Writer",
          "http://dbpedia.org/ontology/Eukaryote"
        ]
      },
      "earmark:refersTo":{
        "@id":"ex:docuverse"
      },
      "earmark:begins":{
        "@type":"xsd:nonNegativeInteger",
        "@value":7
      },
      "earmark:ends":{
        "@type":"xsd:nonNegativeInteger",
        "@value":14
      }
    },
    {
      "@id":"ex:anno_7_entity_17-25",
      "@type":"earmark:PointerRange",
      "rdfs:label":"Guernica",
      "semiotics:denotes":{
        "@id":"dbr:Guernica",
        "@types":[
          "http://dbpedia.org/ontology/City",
          "http://dbpedia.org/ontology/Location",
          "http://dbpedia.org/ontology/Place",
          "http://dbpedia.org/ontology/Settlement",
          "http://dbpedia.org/ontology/PopulatedPlace"
        ]
```

```
        },
        "earmark:refersTo":{
           "@id":"ex:docuverse"
        },
        "earmark:begins":{
           "@type":"xsd:nonNegativeInteger",
           "@value":17
        },
        "earmark:ends":{
           "@type":"xsd:nonNegativeInteger",
           "@value":25
        }
     }
  ]
}
```

# TECHNICAL ANNEX - B

Test Dataset used in the evaluations of Sentiment and Emotion Detection (for Rule Based and Deep Learning modules)

| TEXT | LANG | EMOTION | SENTIMENT |
|---|---|---|---|
| **It has a positive energy to it. Like something good coming up.** | en | Joy | Positive |
| **Nostálgica** | es | Sadness | Negative |
| **El vestido** | es | NONE | NONE |
| **Quietud** | es | Serenity | Positive |
| **Mujer relajada** | es | Serenity | Positive |
| **Ningún sentimiento en concreto. Me gusta la vista del paisaje y casi me sobra la figura humana.** | es | Interest | NONE |
| **Lo que más me llama la atención es la vista esquemática del paisaje y al combinación de verdes fríos.** | es | NONE | NONE |
| **Un poco perturbada e inquieta** | es | Fear, Anticipation, Sadness | Negative |
| **Un retrato y un ojo perturbador** | es | Fear, Anticipation | Negative |
| **Sentimental y melancólico.** | es | Sadness, Love | Negative |
| **Una mujer de clase medi alta.** | es | NONE | NONE |
| **Me agrada la gama del tono, me gustan las formas, pero la expresión de la doña me incomoda.** | es | Anticipation | Negative |
| **Rostro de mujer en tonos rojizos. Formas a veces definidas, bastante geométricas. Me recuerda máscaras de lugares/tiempos no desarrollados. La expresión del sujeto también es una suerte de máscara.** | es | NONE | NONE |
| **Me da una sensación contradictoria ciertos elementos antiguos, descoloridos y sobre la izquierda las flores blancas muy vivas. El gesto de la mujer no encaja eno que se ve a su alrededor** | es | NONE | Mixed |
| **Impresionada y fortaleza** | es | Surprise, Trust | Positive |
| **Cierto desagrado** | es | Disgust | Negative |
| **No me gusta lo que veo y no me atrae.** | es | Disapproval, Disgust | Negative |
| **Relajado y contento.** | es | Serenity, Joy | Positive |
| **Una mujer en una hamaca.** | es | NONE | NONE |
| **Es agradable tomar una siesta entre pinos y helechos, pero la representación no es muy interesante.** | es | NONE | Mixed |
| **Un poco intrigada por conocer su historia/contexto. No me perturba mucho. Es intrigante si pudiera sacar sus armas desgarradoras a voluntad. Es que se quiere presentar menos amenazante, o que es un ser extraño que quizás sea rechazado?** | es | Interest | Mixed |
| **Intensiivinen keskittyminen, inspiraatio, toimintaan virittÃ¤ytyminen, innoitus** | fi | Surprise, Serenity | Mixed |
| **Tulee jotenkin hämmentynyt olo koska kohde katsoo suoraan ulos teoksesta, ja ilme ja asentokin ovat kovin monitulkintaisia. Ihmettelen, mitä on oikein tapahtunut** | fi | Interest | NONE |
| **Pompin abstraktin ja esittävän välillä.** | fi | Surprise | Positive |
| **Tunkkaiselta** | fi | Interest | NONE |
| **Ärsyttää, kamalasti siivottavaa jollekin!** | fi | Disgust, Disapproval | Negative |
| **Teos aiheuttaa vastakkaisia tunteita - siinä on jotain todella mielenkiintoista ja jopa harmonista, mutta samanaikaisesti suuren keltaisen alueen sinapinomainen sävy ja varsinkin vasemmalla keskellä oleva punainen muoto tuntuvat hyvin epämukavilta** | fi | Serenity, Fear, Anticipation | Mixed |
| **Että nainen ei oikeasti ole pulassa** | fi | NONE | NONE |
| **"Merenneito" näyttää hukkuneelta, naisen kuoleman romantisointia / fetisointiako taaskin? Bored now.** | fi | NONE | NONE |
| **Rauhallinen, utelias** | fi | Serenity, Interest | Positive |
| **En pidÃ¤ tÃ¤stÃ¤ kuvasta. VÃ¤rit hermostuttavat.** | fi | Disapproval | Negative |
| **על שברון לב שנשאר סגור כסוד** | he | Sadness | Negative |
| **זה אנטי מה שבדרך כלל תמונה עם דמות עם שמלה גדולה נראית בדרך כלל** | he | NONE | NONE |
| **על דברים עצובים** | he | Sadness | Negative |
| **טרגדיה** | he | Sadness, Fear, Anticipation | Negative |
| **קצת פחד ומעט חמלה** | he | Fear, Love | Mixed |
| **איפוק וחשש- האישה שבתמונה נראית מעט חוששת ומאופקת** | he | Fear, Trust, Interest | Mixed |
| **חוסר נוחות, הבעת פניה ותנוחתה שמביעות טרדה וקוצר רוח גורמות להרגשה לא נוחה** | he | Sadness, Anticipation | Negative |
| **בלבול - חוסר התאמה בין השולחן הערוך יפה לפרחים והלכלוך על הרצפה.** | he | Surprise, Anticipation | Mixed |
| **נראה מפואר וחגיגי** | he | NONE | NONE |
| **אהבה , אושר, הנאה;שמח, חושק** | he | Love, Joy | Positive |

| | | | |
|---|---|---|---|
| **Senso di impotenza e ineluttabilità** | it | Sadness, Fear, Anticipation | Negative |
| **Leggerezza** | it | Serenity | Positive |
| **Leggera sicura** | it | Serenity, Trust | Positive |
| **Speranzoso** | it | Trust, Anticipation | Positive |
| **Rilassata** | it | Serenity | Positive |
| **Inquietudine** | it | Fear, Anticipation | Negative |
| **In grande pace** | it | Serenity | Positive |
| **Mi riporta alla terra di cui ci siamo privati** | it | NONE | Mixed |
| **Quest'opera mi fa pensare ad una persona che ho amato molto** | it | Love | Mixed |
| **un'intrusa nella scena** | it | Anticipation | NONE |
| **Non mi suscita particolari emozioni** | it | NONE | NONE |
| **Distaccato** | it | NONE | NONE |
| **Calma** | it | Serenity | Positive |
| **Bene, invita al riposo** | it | Joy | Positive |
| **pacato, ammirato dalla bellezza** | it | Serenity | Positive |
| **Mi fa pensare ai sogni ad occhi aperti, che sono isole di riposo** | it | NONE | Positive |
| **Preoccupata, le gote rosse riesco solo a collegarle ad uno stato di salute precario, mi fa pensare stia male, e provo pena per lei** | it | Fear, Anticipation | Negative |
| **Dolcezza** | it | Love, Joy | Positive |
| **Felice** | it | Joy | Positive |
| **A disagio** | it | Fear, Anticipation | Negative |
| **Absorbed, observed, intense, slightly scared** | en | Interest, Fear, Anticipation | Mixed |
| **Solitude** | en | Sadness | Negative |
| **melancholic** | en | Sadness, Joy | Negative |
| **loneliness, stuffy air** | en | Sadness, Anticipation | Negative |
| **Tired, impatient** | en | Anticipation | Negative |
| **It's not a happy image. Feel a bit on edge.** | en | Fear, Anticipation | Negative |
| **I like it. But not a strong emotional response.** | en | Interest | Positive |
| **kalm** | en | NONE | NONE |
| **Anxious** | en | Fear, Anticipation | Negative |
| **A little disturbed** | en | Disgust | Negative |
| **Calm, but also positive.** | en | Serenity | Positive |
| **Happy** | en | Joy | Positive |
| **It reminds me sailing with my old friends. Nice memories but also frightening experiences** | en | Joy | Positive |
| **Freedom, breath, fresh air, wind** | en | Serenity, Joy | Positive |
| **empty** | en | Sadness | Negative |
| **Silence, equilibrium, energy (by colors)** | en | Serenity | Positive |
| **Abstract, interesting, calm, uncertain** | en | Serenity | Mixed |
| **Hopeful like a new day** | en | Anticipation, Trust | Positive |
| **Interested, bright, summer. I feel a trajectory of movement in this painting and it feels calm/happy.** | en | Interest, Joy | Positive |
| **Makes me smile** | en | Joy | Positive |
| **Life can be a drudge.** | en | NONE | Mixed |
| **Oppressed. Melancholic.** | en | Fear, Sadness | Negative |
| **Scared** | en | Fear | Negative |
| **Suspended, something must have happened** | en | Anticipation | Negative |
| **the chaos on the table and on the floor is a bit annoying** | en | Anger | Negative |
| **it looks like a couple had a bad fight, or even a disturbing forensic scene - however most likely it was the kids who were responsible for this scene** | en | NONE | NONE |
| **Happy and tired** | en | Joy | Mixed |
| **Feeling unsure, a bit dark, and chilly calm** | en | Serenity, Anticipation | Mixed |
| **Glad; unsure about those flowers;** | en | Joy | Positive |
| **Amused & entranced.** | en | Interest, Joy | Positive |
| **Tired, nostalgic, loving** | en | Sadness, Joy | Negative |
| **It's all over.** | en | Sadness | Negative |
| **Unsettled.** | en | Anger, Fear | Negative |
| **It has a positive energy to it. Like something good coming up.** | en | Joy | Positive |
| **Dark** | en | NONE | NONE |
| **Thinking of the sense of life, anxiety** | en | Anticipation, Fear | Negative |
| **Bored** | en | NONE | Negative |
| **Relax, fulfilment** | en | Serenity | Positive |
| **Curious and awake** | en | Trust, Surprise | Positive |
| **Everyday life** | en | NONE | NONE |