

# EmoGraphArt: A Knowledge Graph for Representing and Explaining Emotions in Abstract Paintings

Rafael Berlanga<sup>1</sup>, Dolores-María Llidó<sup>1</sup> and María-José Aramburu<sup>2,†</sup>

<sup>1</sup>Dpto. Lenguajes y Sistemas Informáticos, Universitat Jaume 1 de Castellón, Spain

<sup>2</sup>Dpto. Ciencias de la Computación, Universitat Jaume 1 de Castellón, Spain

## Abstract

In this paper, we propose a new framework for representing and handling rationales about emotions evoked in abstract images. This representation is based on a logical model for explaining emotions derived from visual stimuli, and a knowledge graph aimed at connecting elements from the stimuli, the appraisal of the annotator and the final evoked emotion. With this approach, we put together in the same space all the components required for explaining emotions in this field. The adopted model relies on the appraisal theory in Psychology, mainly Lazarus' theory, which states that emotions derive from a rationalization process after the stimuli. The proposed knowledge graph provides us with the different perspectives and abstraction levels to analyze the connections between the main components involved in the rationalization of an emotion, namely: the person, the stimuli, and the emotion itself. In this paper, we will mainly focus on the opposites of emotions, that is, how emotions are contrasted in the context of rationalizing the stimuli coming from abstract images. Then, we compare the found opposites with other emotion models that define different axes for complex emotions. The experiments were carried out with the recently published collection FeelingBlue, which associates abstract paintings to basic emotions by means of rationales. We demonstrate that it is possible to extract useful patterns from the rationales by querying the generated knowledge graph.

## Keywords

Affective XAI, Emotion Representation, Knowledge Graphs

## 1. Introduction

Explaining the emotions evoked by abstract images is a challenging task that opens new opportunities for studying key aspects in creativity and education in arts. Gathering rationales (explanations) together with the images (stimuli) allows us to better analyze how emotions are built from stimuli through an appraisal process. In this way, we can achieve a better understanding about how emotional expressions are conducted by artists.

Recent research in affective explainable AI (XAI) has been mainly focused on labeling large datasets of images with small captions and the evoked emotion. However, these datasets lack enough information to capture the relationship between the stimuli and the emotion. Moreover, these models are usually limited to Ekman's model of basic emotions [1]. Even so, no consensus is achieved about the set of labels the annotator must use to assign emotions to images.

Rationales have been recently used to give explanations about decisions for helping in the classification of texts. Several approaches extract the text segments that best align to the rationale and adjust the classification algorithm to improve its performance. This technique has been successfully applied to the sentiment analysis task.

Recently, rationales are used for annotating abstract paintings [2]. In this case, annotators freely express in natural language the emotion an image evokes with respect to a reference basic emotion. Unlike traditional labeling of images, open rationales can provide more details about the evoked emotion as well as its connection to the mentioned stimuli from the image. However, dealing with rationales

---

MAI\_XAI

\*<https://krono.act.uji.es/>

† These authors contributed equally.

✉ [berlanga@uji.es](mailto:berlanga@uji.es) (R. Berlanga); [llido@uji.es](mailto:llido@uji.es) (D. Llidó); [aramburu@uji.es](mailto:aramburu@uji.es) (M. Aramburu)

🆔 0000-0002-0877-7063 (R. Berlanga); 0000-0002-9253-5007 (D. Llidó); 0000-0002-0877-7063 (M. Aramburu)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

is far more complex, requiring an extraction and normalization process similar to those applied in ontology engineering.

In this paper, we propose a knowledge graph to formalize the relationships between stimuli and emotions. We formalize the affective space that results from the appraisal process, connecting the stimuli to the basic emotional reactions or feelings. Afterwards, we extract from rationales all the elements that can be mapped to this space. As far as we know, this is a first attempt to formalize an affective space based on the appraisal theory for abstract art.

## 2. An Explainable Model of Emotions

Several models have been proposed in the literature to represent emotions in metric and qualitative spaces. The most simple one is that of Ekman [1] which consists of the five basic reactions that people usually express in their faces, namely: anger, joy, sadness, fear, and disgust. Maps of arousal-valence values are very popular in the community, as they provide a means to assign EEG (Electroencephalography) signals to points in this map [3]. These maps inspired other forms of organizing emotions according to other axes different from valence and arousal. One of the most elaborated ones is that of the HourGlass (HG) of emotions presented in [4]. Unfortunately, none of these models include elements of explainability, that is, they do not explain the origin and the process that produce the particular emotions. An exception is the OCC model (proposed by Ortony, Clore, and Collins) [5], which organizes the emotions according to different appraisals, providing in this way an abstract explanation of each emotion. However, this model is limited by the small number of complex emotions it is able to represent.

### 2.1. Space of emotions

In our approach, we combine some interesting concepts of the HG and the OCC models. We adopt from the HG model the axes for representing the reactions or feeling, where composition of reactions is possible to represent more complex feelings. From the OCC model we adopt the way a reaction is explained in terms of the appraisal of an event or situation. On the other hand, the HG model provides us the axes with which emotions can be organized. More specifically, we adapt the the revised HG model [4], where 26 emotion categories are arranged into five axes, namely: Temper, Attention, Sensitivity, Introspection, and a neutral axis named Control with two categories *anticipation* and *surprise*.

Table 1 shows the main emotion categories associated with the adapted axes. In uppercase, we mark the Eckman's core emotions. Notice that there are no basic emotions associated to  $E^0$ ,  $A^+$ ,  $T^+$  and  $S^+$  in the Eckman's core.

For the sake of simplicity, we will not include any expression of intensity in the model, as they could be easily included by modifying the polarities with a number in a continuous range like  $[-1, +1]$ . For example, *annoyance* could be denoted with  $T^{-0.3}$ , *anger* with  $T^{-0.5}$  and *rage* with  $T^{-0.9}$ . Estimating intensities are part of a calculus that is out of the scope of this work [4].

We express the composition of emotions by using the operator  $\oplus$ . For example, love is considered in the HG model as a mixture of acceptance and joy ( $I^+ \oplus A^+$ ). Composition can combine different polarities from different dimensions, and the final polarity of the reaction will depend on the intensity of each of the components [4]. Although in some examples the HG model presents contradicted polarities under the same dimension, it is very unlikely to happen and should be revised.

From a logical point of view, composition  $\oplus$  must be considered as an "and" operator. Thus, reactions can be organized into a hierarchy of reactions. However, as previously mentioned, inference of intensities is out of the logic framework and must be defined through a calculus. For this reason, we will omit intensities from now on.

### 2.2. Explaining emotional reactions

According to the Appraisal Theory [6], the evaluation of a situation or an event's consequences produces an emotional reaction. After this reaction, one has to cope with it according to their abilities to handle,

**Table 1**

Emotion space.

Temper	$T^-$	annoyance - ANGER - rage	$T^+$	serenity - calmness - bliss
Attention	$A^-$	dislike - DISGUST - loathing	$A^+$	acceptance - pleasantness - delight
Sensitivity	$S^-$	anxiety - FEAR - terror	$S^+$	responsiveness - eagerness - enthusiasm
Introspection	$I^-$	pensiveness - SAD - grief	$I^+$	contentment - JOY - ecstasy
Control	$E^0$	anticipation (expected)	$U^0$	surprise (unexpected)

accept or change the situation. Appraisal is a complex process that evaluates relevance, resources, and options in the context of the goals of who judges the future expectancy as favorable or unfavorable. Reasoning and understanding the emotional reaction becomes important for the future appraisals as well (learning).

In our model, we express the main aspects in the appraisal process of an event or a situation similarly to the OCC model. We denote any event/situation as  $EV$ , which has the structure shown in Figure 1. Finally, to put together feelings and explanations, we define causality rules. Thus, a (named) emotion is defined as a reaction caused by the appraisal of an event or situation. We use the causal operator  $\triangleleft$  to relate the reaction with the explanation as follows:

$$Emotion \equiv Reaction \triangleleft EV^{polarity}[targets, properties, types]$$

Now, some complex emotions can be expressed with this language. For example “jealousy” can be represented by the following expression:

$$Jealousy \equiv T^- \oplus A^- \triangleleft EV^-[(intimate, other), (presumed), (social)]$$

This expression means that one feels rejection and anger towards both an intimate and another person because of a negative appraisal due to a presumed relationship between them.

We must point out that, in many cases, the reason for an emotion will be just an appraisal with the same polarity as the reaction.

In the abstract art domain,  $EV$  will represent any visual stimulus whose appraisal evokes a particular reaction in the observer. This will considerably reduce the complexity of the explanations as they usually do not involve complex social relationships. For this domain, we will introduce specific targets/properties for the situations, which are based on visual stimuli like colors, shapes, etc. Notice that targets can be persons, objects or actions.

**Figure 1:** Structure of the event/situation explaining an emotion.

$EV$	<b>Polarity of the appraisal:</b> +/−/0
	<b>Target of the appraisal:</b> self / intimate / other / object / etc.
	<b>Properties:</b> controlled - uncontrolled intended - unintended presumed - proven confirmed / unconfirmed / disconfirmed familiar - unfamiliar
	<b>Type the event:</b> internal, sensory, social, etc.

### 2.3. Logical representation

The introduced model for defining emotions can be easily expressed within a logical framework like description logics (DL). Thus, reactions and events/situations are represented with different classes. We must define all the necessary nominals and properties to represent the different elements of the reactions and the events. Disjoint axioms are then introduced over all the contradictory nominals, like basic emotions within the same axis and different polarities. Also, contradicting properties of events/situations are expressed with disjoint axioms. Finally, the casualty operator  $\triangleleft$  must be interpreted as an “and” operator between reaction and appraisal instead of an implication. This is because the appraisal of a same event can lead to different reactions and therefore different emotions.

## 3. Representing emotion rationales in a KG

Knowledge graphs (KGs) are powerful frameworks for representing connected knowledge and data, providing different degrees of quality by means of ontological axioms [7]. In our proposal, KGs are used to represent the emotional rationales along with the relevant entities involved in them. The KG is built according to the following principles:

- It must contain all the concepts involved in the logical model described in Section 2.
- It must map the different elements mentioned in the rationales to concepts in the KG.
- It should link through properties the elements that take part in the generation of a rationale, mainly: stimuli, reactions and appraisals.
- It should include external concepts to enrich the hierarchies of concepts and allow the abstraction of rationales in order to find useful patterns.

As a result, we have defined two ontologies to support our KG, namely: *emoOnto* and *lemaOnto*. The ontology *emoOnto* contains all concepts related to the emotions organized as proposed in the previous section. The ontology *lemaOnto* comprises the lemmas and compound words hierarchically organized into abstract semantic categories such as objects, actions, parts of body, colors, and so on.

### 3.1. Extracting concepts from texts and corpora

Following the methodology and the same tools proposed by the authors of the FeelingBlue dataset, we apply lemmatization to extract all the concepts from the rationales. These lemmas are then mapped to different semantic resources like WordNet and WikiData to enrich them with useful categories like emotions, colors, shapes, etc. These categories are specially useful to find abstract patterns in the rationales.

All the metadata concerning the stimuli (images) and the rationales (basic reactions) are also included in the KG. For example, we include in the KG all the labels assigned by experts in the Wikiart collection and the labels assigned to the FeelingBlue collection. A normalization process was necessary to harmonize all these labels according to the proposed XAI emotion model.

For multimodal XAI, we have also included the palette of colors extracted from the images. This information is useful to compare the mentioned colors in the rationale with the true colors seen in the paintings. Incoherent rationales are then revised to check whether they are due to some error in the normalization process of colors.

### 3.2. FeelingBlue dataset

FeelingBlue [2] has been constructed from a subset of the WikiArt dataset [8]. Wikiart contains images from The Visual Art Encyclopedia annotated with 20 emotions. This collection was devised to analyze human annotations with and without the title of the images.

FeelingBlue only covers the abstract images of the WikiArt collection with the goal of analyzing emotions from pure visual stimuli like color, shapes and textures. Emotions in FeelingBlue have been

reduced to the five emotions nearly resembling the Ekman’s core ones. FeelingBlue contains a total of 912 images, which has been structured into 1,000 annotation tasks. Each task consists of 4 images and a reference emotion (out of five). Each task is then performed by several human annotators. Each annotator writes an explanation for the image with the emotion intensity (MAX) and another one for the image with the minimum emotion intensity (MIN). As a result, we have multiple annotations for the same image, with expressions of MAX and MIN over the reference emotions. These annotations were included as concepts in our KG.

To complete the KG, we translate the whole FeelingBlue collection by structuring the explicit MIN-MAX rationales on abstract images. Table 2 shows an example of MIN-MAX explanations and their corresponding expressions in the proposed model of emotions. Basically, we link the images to their annotations as well as to the elements participating in the logical expressions of the stimuli-appraisal-reaction involved by the rationale. The final statistics of the resulting KG for the FeelingBlue dataset is summarized in Table 3. In the table we have not included the inferred statements which are around 22,700. In Figure 2 we show a fragment of the graph for one painting after loading the KG into the GraphDB tool [9].

To allow sentiment analysis, we have used the spaCy<sup>1</sup> Python library. So we can associate a global polarity to each annotation rationale. Finally, for each image we summarize the polarity of the sentiment of all its rationales.

**Table 2**

Rationales and their formalization in the proposed model (we only include the *targets* component of the appraisal)

Annotation	Rationale	Expression
MAX.disgust	The blue part is like a shape of a monster.	$A^- \triangleleft EV^-[(color(blue), shape(monster))]$
MIN.disgust	Some sort of serenity in the picture.	$T^+ \triangleleft EV^+[(object(picture))]$

**Table 3**

Statistics of the EmoGraphArt for FeelingBlue

Number of Statements	506,673
Number of Images	912
Number of Task	1000
Number of Emotions	100
Number of Web Colors	139
Nodes for Emotion Category	5
Number of Colors Category	11
Nodes for stimuli category	42
Number of stimuli	16,354

In Figure 2, we show the annotated emotions in WikiArt (*emotionA*). Only emotions annotated by more than 10% of the annotators were included in the KG. In this case, the image was labelled in WikiArt with the emotions *Love*, *Arrogance*, *Disgust*, *Humility* and *Optimism*. Regarding the FeelingBlue annotations, the image has been explained through 16 tasks related with the emotions *Disgust* and *Anger*.

The *Anger* task has been annotated nine times as *MAX.anger* with text polarity  $-1.0$  and eight times as *MIN.anger* with polarity  $1.04$ . This indicates that the image is controversial as has a similar number of annotations expressing opposite emotions.

We can now analyze the divergences of emotions between WikiArt and FeelingBlue by retrieving the polarities implied by the labeled emotions. We can do this through a SPARQL query on the KG as follows.

In the query 3, we have selected all the data about images that have assigned different emotion

<sup>1</sup><https://spacy.io>

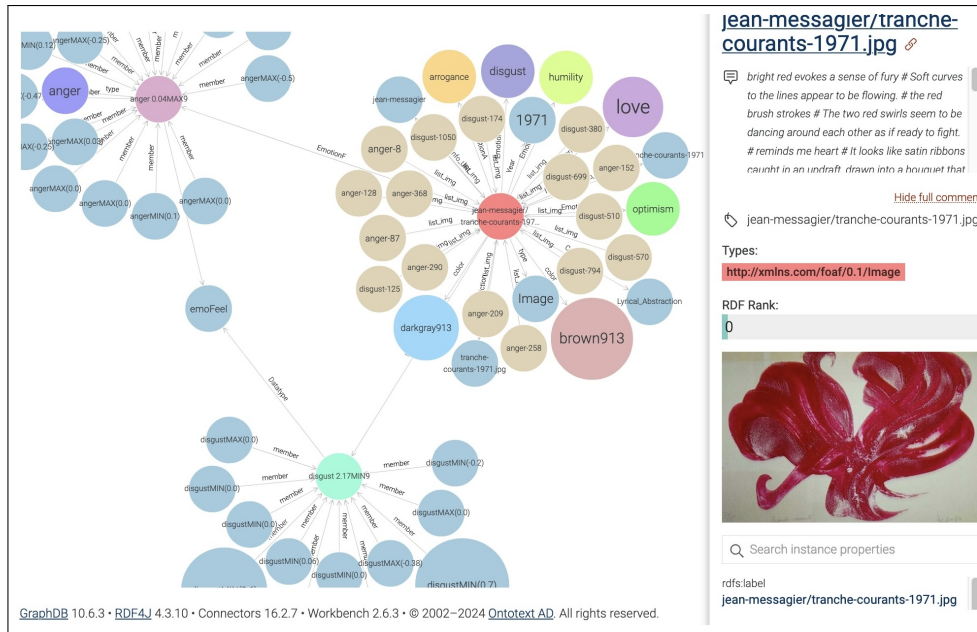


Figure 2: Knowledge graph representing emotions in abstract paintings.

```

select ?term ?emoA ?emoAvalue ?emoPA ?emoF ?emoFvalue ?emoPF
where {
  ?term rdf:type foaf:Image.

  ?term base:EmotionA ?emoAnnoA.
  ?emoAnnoA rdf:value ?emoAvalue.
  ?emoAnnoA rdf:type ?emoA.
  ?emoA base:emoOnto ?emoAnnoAA.
  ?emoAnnoAA base:polarity ?emoPA.

  ?term base:EmotionF ?emoAnnoF.
  ?emoAnnoF rdfs:label ?emoFvalue.
  ?emoAnnoF rdf:type ?emoF.
  ?emoF base:emoOnto ?emoAnnoFF.
  ?emoAnnoFF base:polarity ?emoPF.

  FILTER ( ?emoPF != ?emoPA )
} groupby ?term ?emoA ?emoAvalue ?emoPA ?emoF ?emoFvalue ?emoPF

```

Figure 3: SQL to contrast annotations of emotions and text polarities.

polarities with the two emotions annotators: WikiArt (*emoA*) and FeelingBlue (*emoF*), where *emoPA* and *emoPF* represents their respective polarities and *emoAvalue* represents the polarity value.

Table 4

SPARQL Result for dataImages: img913. Task summaries consists of the task, the number of annotators and the overall polarity.

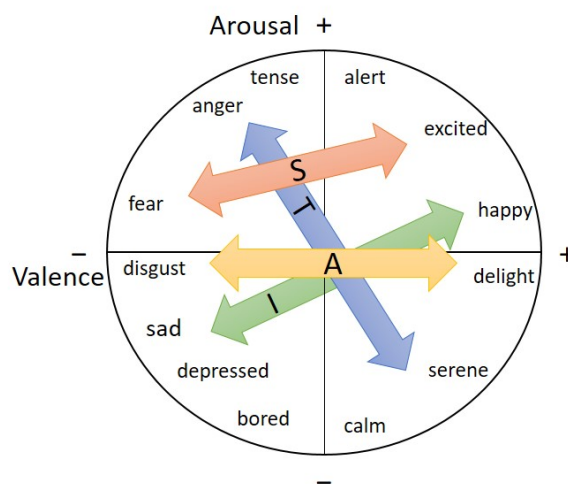
<i>emoA</i>	<i>emoAvalue</i>	<i>emoPA</i>	<i>emoF</i>	Task Summary	<i>emoPF</i>
base:M_love	0.273	emo:Positive	base:F_disgust	MIN.disgust 9 2.17	emo:Negative
base:M_love	0.273	emo:Positive	base:F_anger	MAX.anger 9 0.04	emo:Negative

From the results of the query, we will inspect the case of task "anger.218", where rationales showed some contradictions. Table 4 shows that the image has been associated to the anger emotion with a preference to MAX by 9 annotators. However, the overall rationale polarity is positive. In this case, the text polarity has been wrongly calculated. One of the annotators explains that the image is *MAX.anger* because "bright red evokes a sense of "fury". The positive polarity comes from the word "bright" with polarity +0.7, whereas "fury" has not been regarded as a negative word by the sentiment analysis tool. On the other hand, the image has been also annotated as *MIN.disgust* with overall positive polarity. In this case, the rationale associated to the "minimum disgusting image" task was "bright color so better feel". Notice that both rationales points to the same stimulus (bright color/red) producing opposite appraisals and reactions: "evokes fury" vs. "better feel".

#### 4. Extracting patterns from the knowledge graph

The resulting EmoGraphArt can serve us to test hypotheses about the stimuli and the emotions. For example, an interesting question is how emotions are negated when annotators contrast MIN and MAX annotations. Other relevant questions can be as follows: Are the MIN-MAX oppositions related to the model axes? Is the annotator prone to select an image with a switched polarity in the same axis of the reference emotion?

As described in Section 2, we deal with four axes, namely: Attention (*A*), Introspection (*I*), Temper (*T*), and Sensibility (*S*). We ignore the Control axis because it cannot be placed in this map, and it is poorly represented in the dataset. The selected axes can be visualized in the arousal-valence map as the Figure 4 shows.



**Figure 4:** Axes of the HG represented in the arousal-valence map.

We analyze the MIN rationales to estimate the likelihood of each emotion to be contrasted to that of the reference. Table 5 shows the statistics comparing the rationale emotion with the reference emotion for MIN annotations.

We can see that many of the MIN explanations fall into the neutral emotion, which means that the annotator usually does not have any feeling with respect to the expected one. This happens specially for the *MIN.disgust* task ( $A^-$ ), whose opposite is the Pleasant emotion ( $A^+$ ), which is not the most likely mentioned emotion. Notice that for *MIN.disgust* the most likely non-neutral evoked emotions are calm ( $T^+$ ) and joy ( $I^+$ ), both positive as expected, but not in the same axis as disgust. Calm ( $T^+$ ) is a frequent non-neutral emotion for negative emotions, except sadness. This is because anger, disgust, and fear have a high arousal and sadness a low one. Thus, the opposites to these emotions are positive with low arousal like calm. Finally, the contrast sadness-joy is quite clear in the results. This means

that when the reference emotion is either sadness ( $I^-$ ) or happiness ( $I^+$ ), the annotator looks for an image with the opposite feeling in the same axis.

We have also included some expressions like “less happy” and “less sad” present in some rationales to show that for some tasks the user looks for a less intense feeling and not the opposite one.

**Table 5**

Opposites of emotions expressed in the MIN rationales.

Emotion	MIN.anger	MIN.disgust	MIN.fear	MIN.happy	MIN.sad	ALL
neutral	725	2121	1434	1227	609	6116
anger	1	0	2	32	1	36
pleasant	14	107	27	0	9	157
calm	189	343	385	23	66	1006
sadness	7	2	8	558	33	500
disgust	0	0	3	52	2	57
fear	2	3	8	53	1	67
joy	50	220	185	2	149	606
less happy	0	1	4	105	6	116
less sad	0	0	1	1	13	15
conflict	0	0	0	15	0	15

## 5. Conclusion

Representing explanations of abstract art requires the combination of visual stimuli along with appraisals that come from the subjective interpretation of the observer. Due to the complexity of this information, we propose to represent them in a knowledge graph governed by a logical model of emotions able to connect the necessary elements for the explanations. With this KG, we are able to find out useful correlations between stimuli and emotions contextualized to specific images and observers.

In the preliminary experiments, we have shown how this graph can be applied to find contradictions from the different sources of annotations, as well as to understand how emotions are contrasted for different images. This is a first step to study the dynamics of the emotions, that is, how emotions can lead to further emotions through the course of a series of stimuli. Future work will go further in this line as we plan to extract patterns from images to find connections between these stimuli and the emotions that are evoked in the rationales.

## Acknowledgments

This research has been partially funded by the Spanish Ministry of Science under grants PID2021-123152OB-C22 funded by the MCIN/AEI/10.13039/501100011033 and by the European Union and FEDER/ERDF (European Regional Development Funds).

## References

- [1] P. Ekman, W. Friesen, *Unmasking the Face*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
- [2] A. Ananthram, O. Winn, S. Muresan, *Feelingblue: A corpus for understanding the emotional connotation of color in context*, *Transactions of the Association for Computational Linguistics* 11 (2023) 176–190.
- [3] J. Russell, L. Barrett, *Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant*, *Journal of Personality and Social Psychology* 76 (1999) 805–819.
- [4] Y. Susanto, G. Livingstone, B. C. Ng, E. Cambria, *The hourglass model revisited*, *IEEE Intelligent Systems* 35 (2020) 96–102.



- [5] G. Clore, A. Ortony, Psychological construction in the occ model of emotion, *Emotion Review* 5 (2013) 335–343.
- [6] R. Lazarus, Progress on a cognitive-motivational-relational theory of emotion, *American Psychologist* 46 (1991) 819–834.
- [7] A. Hogan, et al., Knowledge graphs, *ACM Computing Surveys* 54 (2021) 1–37.
- [8] S. Mohammad, S. Kiritchenko, Wikiart emotions: An annotated dataset of emotions evoked by art, in: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [9] B. Bishop, S. Bojanov, Implementing owl 2 rl and owl 2 ql rule-sets for owlim, in: *CEUR Workshop Proceedings*, volume 796, 2011.