Multimodal Multimodel Emotion Analysis as Linked Data

J. Fernando Sánchez-Rada, Carlos A. Iglesias
GSI Universidad Politécnica de Madrid
Hesam Sagha, Björn Schuller
University of Passau
Ian Wood, Paul Buitelaar
National University of Ireland Galway

Abstract—The lack of a standard emotion representation model hinders emotion analysis due to the incompatibility of annotation formats and models from different sources, tools and annotation services. This is also a limiting factor for multimodal analysis, since recognition services from different modalities (audio, video, text) tend to have different representation models (e. g., continuous vs. discrete emotions).

This work presents a multi-disciplinary effort to alleviate this problem by formalizing conversion between emotion models. The specific contributions are: i) a semantic representation of emotion conversion; ii) an API proposal for services that perform automatic conversion; iii) a reference implementation of such a service; and iv) validation of the proposal through use cases that integrate different emotion models and service providers.

1. Introduction

Emotions permeate every aspect of our lives, from our facial expressions to our comments on social media. However, there is no consensus on the representation of human emotions. So far there has been little attention to unifying measurements, categories, and emotion codes. This is partly due to the field of 'affective computing' being relatively young. As a result, there is a plethora of rivaling emotion representation models with varying degrees of popularity, from categorical models such as Ekman's to Scherer's process model [1].

The lack of a standard is a hindrance when working with different sources, such as datasets annotated by different experts, due to additional effort that has to be spent in understanding the definitions of emotion in every source. It also limits the amount of annotated data for training. In some cases, for the sake of interoperability, a single representation model is chosen on a per-project basis. Then the use of other models is restricted -limiting the resources and quality-, or an ad-hoc conversion mechanism is used, which is costly and inaccurate. However, this compromise is not possible in all cases.

Initiatives such as Emotion Markup Language (EmotionML) [2] and the Onyx Emotion Ontology [3] account for the heterogeneity of models and provide vocabularies or meta-models that enable interoperability. When using these meta-models, annotations do not refer to ambiguous terms (such as *anger*) but to specific definitions (e. g., *Ekman's definition of anger*). This has two consequences: the choice of models is explicit in the annotation itself, and different models may be used in the same data set. As a result, annotations unambiguously refer to the model being used.

However, to the best of our knowledge, none of the meta-models addresses the combination of annotations using different models in a meaningful way. Hence, these annotations are still independent. This work aims to remedy this by formalizing the conversion between emotion models.

The rest of the paper is structured as follows. Section 2 introduces enabling technologies and related research in multimodal and multimodel representation; Section 3 details the proposal for semantic representation of emotion conversion; Section 4 presents our evaluation by means of a reference implementation and a use case; lastly, Section 5 summarizes our conclusions and future work.

2. Background

This section focuses on two aspects: the definition and quantification of emotions (emotion models) and how this information is encoded (representation formats). Previous works have discussed the difference between emotions and related terms (e. g., 'feelings', 'affects', 'sentiment') in detail [4], [5], [6], [7].

2.1. Models for emotions and emotion analysis

There are several models for emotions, ranging from the most simplistic and ancient that come from Chinese philosophers to the most modern theories that refine and expand older models [8], [9]. The literature on the topic is vast, and it is out of the scope of this paper to reproduce it. For the purpose of this paper, it is important to know that emotion models vary in the characteristics of the emotion they represent, and the way in which these characteristics are represented. The main two groups would be: discrete and dimensional models. In discrete models, emotions belong to one of a predefined set of categories, which varies from

model to model. In dimensional models, an emotion is represented by the value in different axes or dimensions. A third category, mixed models, merges both views. The recent work by Cambria et al. [10] contains a comprehensive state of the art on the topic, as well as an introduction to a novel model, *the Hourglass of emotions*, inspired by Plutchik's studies [11]. Plutchik's model is a model of categories that has been extensively used [12], [13] in the area of emotion analysis and affective computing, relating all the different emotions to each other in what is called the wheel of emotions.

A more recent development in emotion representation designed as a principled annotation scheme is the Geneva Emotion Wheel [14]. This scheme combines 20 emotion labels arranged as a circle in a 2-dimensional Valence/Power space with four levels of intensity represented by distance from the centre.

Other models cover affects in general, which include emotions as part of them. One of them is the work done by Strapparava and Valitutti in WordNet-Affect [15]. It comprises more than 300 affects linked by concept-superconcept relationships, many of which are considered emotions. What makes this categorization interesting is that it effectively provides a taxonomy of emotions. It provides information about relationships between emotions and makes it possible to make choices on the level of granularity of the emotion model.

Despite all efforts, there is no universally accepted model for emotions [7], [16]. This complicates the task of representing emotions. In a discussion regarding EmotionML, Schröder et al. pose that *given the fact that even emotion theorists have very diverse definitions of what an emotion is, and that very different representations have been proposed in different research strand, any attempt to propose a standard way of representing emotions for technological contexts seems doomed to fail* [17]. Instead they claim that the markup should offer users a choice of representation, including the option to specify the affective state that is being labeled, different emotional dimensions and appraisal scales.

EmotionML [2] is one of the most notable general-purpose emotion annotation and representation languages. It was born from the efforts made for Emotion Annotation and Representation Language (EARL) [16], [18]. EARL originally included 48 emotions divided into 10 different categories. EmotionML offers twelve vocabularies for categories, appraisals, dimensions and action tendencies. A vocabulary is a set of possible values for any given attribute of the emotion. A complete description of those vocabularies and its computer-readable form is available in [19].

2.2. Multimodal Linked Data approaches

Recent work has expanded traditional annotation, such as that of EmotionML, by adding semantics and following Linked Data principles [20]. This shift has several important implications. First and foremost, it fosters the integration of different data sources. Whereas traditional annotations are usually tied to a document, this new type of annotation is meant to be queried, consumed and integrated with other sources. As a consequence, it also entails the formal definition of vocabularies and ontologies, which serve as a common representation for all sources.

The most common linked data model for emotion representation is a combination of several existing vocabularies: Onyx [3], a vocabulary to annotate and describe emotions which provides interoperability with EmotionML [21]; and NLP Interchange Format (NIF) 2.0 [22], which defines a semantic format and API for improving interoperability among natural language processing services.

Another important contribution laid the foundation to multimodal annotation [23] by adding a multimedia extension to the NIF model. The bulk of this extension (MESA) is the addition of a URI scheme for multimedia contexts, which complement the original NIF string context. This scheme follows the media fragments recommendation [24] to provide URIs for multimedia segments.

3. Proposal

Our proposal for representation of multimodal multimodel emotion analysis consists of two parts. The main one is the definition of a semantic vocabulary for annotation and conversion of annotations in different models and modalities. The second part is an API for emotion analysis services and tools that leverages the vocabulary.

We selected six basic aspects that a potential vocabulary needs to cover in order to be complete: 1) the definition of emotion annotations; 2) the definition of each emotion model (e. g., Ekman's categories); 3) multimodality; 4) the definition of the process of annotation with emotion (e. g., manual or automatic annotation) and the link between this process and the emotion annotations it generates (provenance); 5) the definition of conversion process between annotation models in a way that is compatible with (4); and 6) integration with RDF and linked data.

We reviewed existing publicly available vocabularies, looking for these criteria (Table 1). None of the vocabularies reviewed included the concept of emotion conversion. However, all the candidates can be extended, either via an XML schema or a semantic extension. Consequently, conversion could be integrated as a small extension of already existing vocabularies. To cover the rest of the criteria, there were two clear alternatives. The first one was to combine several XML schemas: EmotionML (emotions and models), EMMA [25] (multimodality) and Provenance Ontology (PROV-O) [26](provenance). The main advantage of this option is adopting EmotionML, which is well known and already integrated in several tools. Additionally, EMMA is both a W3C recommendation and the encouraged way to integrate multimodality into EmotionML annotations. Unfortunately, although EmotionML and EMMA are semantic in nature, this approach does not meet the linked data requirements, with the exception of Prov-O. However, a subset of EmotionML has already been included as Onyx sub-vocabularies, and Onyx has been successfully used as an

Table 1. COMPARISON OF DIFFERENT VOCABULARIES AND THE PROPOSAL IN THIS PAPER.

	emotions	multimodel	multimodal	provenance	conversion	semantic
Emotion-ML	✓	✓				
EMMA			✓			
PROV-O				✓		✓
Onyx	✓	✓		✓		✓
Onyx+MESA	✓	✓	✓	✓		✓
Emotion-ML + EMMA + PROV-O	✓	✓	✓	✓		
Proposed vocabulary	✓	✓	✓	✓	✓	✓

Figure 1. Extension of the Onyx ontology for emotion conversion. The extension itself shown in bold and without a prefix.

alternative to EmotionML in several projects where linked data was a strong constraint. This reason led us to the second alternative: to extend the Onyx and MESA vocabularies to include the notion of emotion conversion. We will briefly cover how NIF, Onyx, MESA and Prov-O can be used together to cover our first criteria before introducing the extension.

First of all, emotion models are represented with Onyx *EmotionModel*. All public vocabularies in the EmotionML vocabularies specification have their counterpart in the Onyx EmotionML vocabularies extension. If none of those models cover a specific case, it is also possible to define custom models and to publish them as linked data. The emotion analysis task is encoded by Onyx's *EmotionAnalysis*, a subclass of Prov-O *Activity*. As such, it can provide provenance information. It should also specify the specific model it uses for annotation. Each piece of text to be analyzed is represented as a NIF context. One of the main advantages of NIF is that it defines URI schemes for contexts which only depend on the content itself and its source. Using these unique identifiers, it is possible to aggregate annotations added by independent analysis to the same source. In order to preserve this property, provenance is stored at the annotation level. Hence, contexts are annotated with one or several Onyx *EmotionSet* entities. These *EmotionSet* entities contain all the emotion information, as well as a link to the analysis activity that generated the annotation.

To achieve multimodality in a NIF-compatible manner, MESA includes a specific NIF URI scheme to annotate string contexts within multimedia, based on the media fragments recommendation. An example is shown in Listing 1.

Listing 1. ANNOTATING STRINGS IN MULTIMEDIA.

```
<http://video.com/example#t=0,11>
    a mesa:MediaFragmentsString ;
```

To cover all the criteria, the existing models need to be extended to include the concept of emotion conversion. Figure 1 illustrates how this extension integrates with the existing entities in Onyx and Prov-O. First of all, the extension provides a new class, *Conversion*, which subclasses
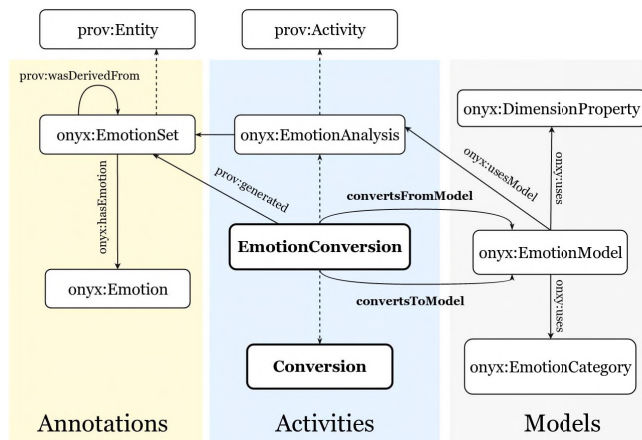
Activity. By making this class independent of emotions, it can be reused for other types of affect conversion. For instance, it could be used to represent the conversion of star-based opinion models to polarity based (thumbs up/down) models. The vocabulary also provides a more specific *EmotionConversion* activity, which subclasses both *Conversion* and *EmotionAnalysis*. The main specific properties of this class are *convertsFrom* and *convertsTo*, both of which point to an *EmotionModel* instance.

Using this extension, we can encode the conversion of the previous example from using Ekman's categorical model to PAD dimensions. An excerpt of this representation is shown in Listing 2.

Listing 2. ONYX EXTENSION TO COVER CONVERSION.

```
:Big6_to_PAD rdf:type onyx:EmotionConversion ;
        onyx:convertsFrom emoml:big6 ;
        onyx:convertsTo emoml:pad .

<http://microblog.com/User1/Post1#char=0,11>
 onyx:hasEmotionSet :EmotionSet1 ,
             :EmotionSet2 .

:EmotionSet2
 rdf:type onyx:EmotionSet;
 onyx:hasEmotion [
    rdf:type onyx:Emotion ;
    emoml:pad_potency 0.8 ;
    emoml:pad_arousal 0.5 ;
    emoml:pad_dominance 0.3 ;
 ] ;
 prov:wasGeneratedBy :Big6_to_PAD .
```

The second part of the proposal is the web API that allows emotion analysis services to integrate emotion conversion. In particular, a service has to: advertise what models they use to annotate; advertise the conversions available; allow users to request a specific model; in case of not being able to convert from the original model to the one requested by the user, raise an error. The complete API for services is included in Table 2.

The proposed model and API have been integrated in

Senpy, a semantic framework to build sentiment and emotion analysis services [27]. This allows Senpy emotion analysis services to offer automatic emotion model conversion. The framework currently includes a very generic implementation of a centroid-based conversion, inspired by Kim et al [28]. In this centroid-based conversion, each category or label is mapped to a centroid, a point in an N-dimensional space (e.g. VAD). This provides bi-directional conversion. Given an emotion with one or more categories and an optional intensity for each category, the algorithm returns a new emotion whose VAD values are the weighted average of the values of the centroids corresponding to the categories. Given dimensional value, the conversion consists in calculating the distance of the value to each centroid, and either returning a new Emotion with a single category (the closest centroid), or one emotion per centroid and an intensity proportional to the normalized distance between the centroid and the value. This algorithm can be applied with different sets of centroids. We provide centroids for a conversion from Ekman categories to VAD values. To calculate them, we averaged the VAD values of the words in ANEW [29] that were also present in WordNet-Affect [15] under a label that can be mapped to one of Ekman's categories. All the code is Open Source and available on the framework website.

4. Evaluation

This section presents a real scenario where the proposed conversion model and service API have been used. It serves as a starting point to assess the usefulness and adequacy of the proposal. It is also a way to identify possible deficiencies and to encourage further discussion on the topic.

Table 2. THE EXTENDED EMOTION ANALYSIS SERVICE API INCLUDES PARAMETERS TO CONTROL EMOTION CONVERSION.

parameter	description
input(i)	serialized data (i.e. the text or other formats, depends on informat)
informat (f)	format in which the input is provided: turtle, text (default) or json-ld
outformat (o)	format in which the output is serialized: turtle (default), text or json-ld
prefix (p)	prefix used to create and parse URIs
minpolarity (min)	minimum polarity value of the sentiment analysis
maxpolarity (max)	maximum polarity value of the sentiment analysis
language (l)	language of the sentiment or emotion analysis
domain (d)	domain of the sentiment or emotion analysis
algorithm (a)	plugin that should be used for this analysis
emotionModel (emodel, e)	emotion model in which the output is serialized (e.g. WordNet-Affect, PAD, etc.)
conversionType	type of emotion conversion. Currently accepted values: 1) full, results contain both the converted emotions and the original emotions, alongside; 2) nested, converted emotions should appear at the top level, and link to the original ones; 3) filtered, results should only contain the converted emotions.

The use case is as follows. A given video is analyzed by three different services: a video analysis that detects emotions in faces; an emotion analysis on speech; and a text emotion analysis service that annotates the transcription of the speech. Annotations are converted to a continuous space (if necessary) and the results are fused to yield the final outcome. Figure 2 shows an overview of the analysis and their relationships.
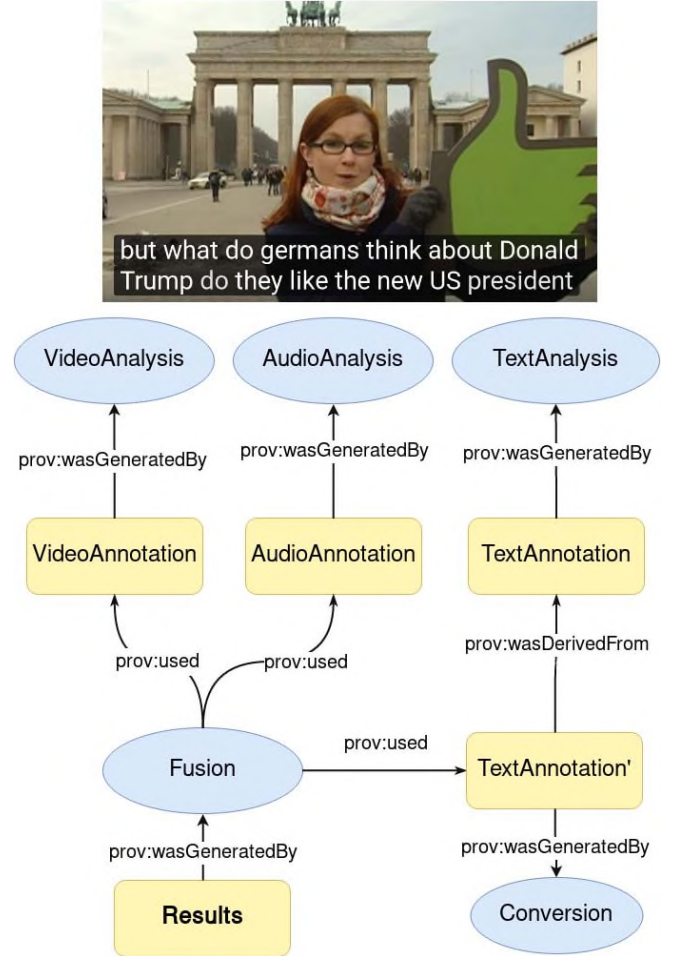

Figure 2. Generation of results combining emotion analysis in three modalities. Ellipses are provenance activities, rectangles are provenance entities.

The video emotion analyzer consists of a face detector (based on on a discriminatively trained deformable part model [30]), face tracking, and facial expression detector (based on convolutional neural networks), which recognises emotions in the continuous arousal/valence space. The speech emotion analyzer is based on openSMILE acoustic feature extractor [31], and Bag-of-Audio-Words [32], a similar concept to the Bag-of-Words of text analysis. Emotions are predicted in a continuous arousal/valence space. The text emotion analysis is performed through the *emotion-wnaffect* senpy plugin, which uses a lexicon-based approach based on WordNet-Affect [15]. It maps every affect label in the WordNet-Affect taxonomy to five of Ekman's categories:

anger, fear, disgust, joy and sadness.

The first two services, audio and video, use the PAD (Pleasure, Arousal, Dominance) model, whereas the analysis on text uses a simpler categorical analysis based on Ekman's model. In order to fuse the three annotations, the categories used in the annotation of text are converted to PAD values, using the centroid-based conversion activity which we developed as a senpy plugin. Among other things, the definition of the activity includes the algorithm being used (*senpy.plugins.conversion.centroids* at version 0.1), the values for each of the centroids, and the corresponding emotion for each of the centroids. An excerpt the definition of the conversion from Ekman dimensions to VAD dimensions is included as Listing 3.

Listing 3. DEFINITION OF THE ACTIVITY THAT CONVERTS ANNOTATIONS FROM EKMAN'S CATEGORIES TO PAD VALUES.

```
<http://servicehost/api/plugins/Ekman2PAD_0.1> a :
    emotionConversionPlugin ;
  onyx:centroids [
      onyx:hasEmotionCategory emoml:big6anger ;
      emoml:arousal 6.95e+00 ;
      emoml:dominance 5.1e+00 ;
      emoml:valence 2.7e+00 ] ;
    [
    onyx:hasEmotionCategory emoml:big6disgust ;
     emoml:arousal 5.3e+00 ;
     emoml:dominance 8.05e+00 ;
     emoml:valence 2.7e+00 ] ;
    [
    onyx:hasEmotionCategory emoml:big6fear ;
     emoml:arousal 6.5e+00 ;
     emoml:dominance 3.6e+00 ;
     emoml:valence 3.2e+00 ] ;
    [
    onyx:hasEmotionCategory emoml:big6happiness ;
     emoml:arousal 7.22e+00 ;
     emoml:dominance 6.28e+00 ;
     emoml:valence 8.6e+00 ] ;
    [
    onyx:hasEmotionCategory emoml:big6sadness ;
     emoml:arousal 5.21e+00 ;
     emoml:dominance 2.82e+00 ;
     emoml:valence 2.21e+00 ] ] ;
  onyx:convertsFrom emoml:big6 ;
  onyx:convertsTo emoml:pad ;
  senpy:description "Plugin to convert emotion sets from
      Ekman to VAD" ;
  senpy:module "senpy.plugins.conversion.emotion.
      centroids" ;
  senpy:name "Ekman2PAD" ;
  senpy:version 1e-01 .
```

Once all the dimensions are mapped into the PAD model, the *fusion service* combines the results of the different modalities and compute the final results. We have chosen the weighted average classifier fusion technique for this task, since within this schema (i) different analyzers are considered independent of each other (comparing with feature fusion techniques), and (ii) each modality may contribute differently for each emotion dimension (e. g., it is well-known that speech has higher impact on arousal detection, while facial movements have higher impact on valence detection). Weights can be trained offline, by heuristics, or the uniform weights can can be used if no information is provided.

Listing 4 shows an edited fragment of the annotations at the end of the whole process. The actual results were collected in JSON-LD format and stored in an elasticsearch database, and Kibana was used for visualization [1].

Listing 4. RESULTS FROM THE FUSION PHASE

```
ex:video.mp4:time=0,2 a nif:Context ;
  onyx:hasEmotionSet [
    prov:wasGeneratedBy ex:videoAnalysis ;
    onyx:hasEmotion [
     emoml:pad_arousal 0.5 ;
     emoml:pad_valence 0.1
     ] .
    ] ,
    [
    prov:wasGeneratedBy ex:audioAnalysis ;
    onyx:hasEmotion [
     emoml:pad_pleasure -0.2 ;
     emoml:pad_arousal -0.4
     ] .
    ] ;
    _:textResults ;
    [
    prov:wasGeneratedBy <http://servicehost/api/plugins/
        Ekman2PAD_0.1> ;
    prov:wasDerivedFrom _:testResults ;
    onyx:hasEmotion [
     emoml:pad_pleasure 0 ;
     emoml:pad_arousal 0 ;
     emoml:pad_dominance 0
     ] .
    ] ;
    prov:wasGeneratedBy ex:fusion ;
    onyx:hasEmotion [
     emoml:pad_arousal 0.049 ;
     emoml:pad_valence -0.05 ;
     ] .
    ] .

_:textResults prov:wasGeneratedBy ex:textAnalysis ;
    onyx:hasEmotion [] ,

ex:fusion
  a onyx:EmotionConversion ;
  onyx:convertsFrom emoml:pad-dimensions ;
  onyx:convertsTo emoml:pad-dimensions .
```

5. Conclusions

In this paper, we proposed an approach for the integration of emotion analysis in different modalities (multimodal) and using different emotion representation models (multimodel). The proposed linked data vocabulary unifies and extends existing vocabularies to provide a complete coverage of multimodal multimodel emotion annotations, including the unambiguous definition of conversion to different emotion models. The vocabulary is compatible with existing specifications and recommendations, such as EmotionML. Additionally, it integrates with the provenance ontology, which means annotations are modeled as entities whose provenance (origin) can be traced to either an annotation or a conversion activity. These activities can in turn be precisely modeled, including the resources being used, the emotion models adopted, and other entities that were transformed by them. In addition to the model, a reference implementation

---

1. Elasticsearch and Kibana: https://www.elastic.co/

of automatic emotion conversion has been integrated into senpy (a framework for sentiment and emotion analysis).

Lastly, the applicability and completeness of this approach and the reference implementation has been assessed through a use case that integrates multimodel multimodal annotations.

Acknowledgments

References

[1] K. R. Scherer, "Psychological models of emotion," *The neuropsychology of emotion*, vol. 137, no. 3, pp. 137–162, 2000.

[2] P. Baggia, C. Pelachaud, C. Peter, and E. Zovato, "Emotion Markup Language (EmotionML) 1.0 W3C Recommendation," W3C, Tech. Rep., May 2014. [Online]. Available: http://www.w3.org/TR/emotionml/

[3] J. F. Sánchez-Rada and C. A. Iglesias, "Onyx: A Linked Data Approach to Emotion Representation," *Information Processing & Management*, vol. 52, pp. 99–114, January 2016.

[4] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *Intelligent Systems, IEEE*, vol. 28, no. 2, pp. 15–21, March 2013.

[5] M. Munezero, C. Montero, E. Sutinen, and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text," *Affective Computing, IEEE Transactions on*, vol. 5, no. 2, pp. 101–111, April 2014.

[6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[7] S. M. Mohammad, "Sentiment analysis: Detecting valence, emotions, and other affectual states from text," in *Emotion Measurement*, H. Meiselman, Ed. Woodhead Publishing, 2016.

[8] P. Ekman, "Basic emotions," *Handbook of cognition and emotion*, vol. 98, pp. 45–60, 1999.

[9] J. J. Prinz, *Gut reactions: A perceptual theory of emotion*. Oxford University Press, 2004.

[10] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in *Cognitive Behavioural Systems*. Springer, 2012, pp. 144–157.

[11] R. Plutchik, *Emotion: A psychoevolutionary synthesis*. Harper & Row New York, 1980.

[12] D. Borth, T. Chen, R. Ji, and S.-F. Chang, "Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content," in *Proc. 21st ACM international conference on Multimedia*. New York, NY, USA: ACM, 2013, pp. 459–460.

[13] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis." in *FLAIRS Conference*, 2012, pp. 202–207.

[14] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.

[15] C. Strapparava and A. Valitutti, "Wordnet-affect: An affective extension of wordnet," in *Proc. LREC*, vol. 4, 2004, pp. 1083–1086.

[16] M. Schröder, H. Pirker, and M. Lamolle, "First suggestions for an emotion annotation and representation language," in *Proc. LREC*, vol. 6, 2006, pp. 88–92.

[17] M. Schröder, L. Devillers, K. Karpouzis, J.-C. Martin, C. Pelachaud, C. Peter, H. Pirker, B. Schuller, J. Tao, and I. Wilson, "What should a generic emotion markup language be able to represent?" in *Affective Computing and Intelligent Interaction*. Springer, 2007, pp. 440–451.

[18] H. N. of Excellence, "Humaine emotion annotation and representation language (earl): Proposal." HUMAINE Network of Excellence, Tech. Rep., Jun. 2006. [Online]. Available: http://emotion-research.net/projects/humaine/earl/proposal#Dialects

[19] K. Ashimura, P. Baggia, F. Burkhardt, A. Oltramari, C. Peter, and E. Zovato, "Emotionml vocabularies," W3C, Tech. Rep., May 2012. [Online]. Available: http://www.w3.org/TR/2012/NOTE-emotion-voc-20120510/

[20] C. A. Iglesias, J. F. Sánchez-Rada, G. Vulcu, and P. Buitelaar, "Linked Data Models for Sentiment and Emotion Analysis in Social Networks," in *Sentiment Analysis in Social Networks*. Morgan Kauffman, October 2016, ch. Linked Dat, pp. 46–66.

[21] M. Schröder, P. Baggia, F. Burkhardt, C. Pelachaud, C. Peter, and E. Zovato, "Emotionml – an upcoming standard for representing emotions and related states," in *Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Springer Berlin Heidelberg, 2011, vol. 6974, pp. 316–325.

[22] S. Hellmann, J. Lehmann, S. Auer, and M. Brümmer, "Integrating nlp using linked data," in *The Semantic Web–ISWC 2013*. Springer, 2013, pp. 98–113.

[23] J. F. Sánchez-Rada, C. A. Iglesias, and R. Gil, "A Linked Data Model for Multimodal Sentiment and Emotion Analysis," Beijing, China, July 2015, pp. 11–19.

[24] Media fragments URI 1.0 (basic). 00005.

[25] W. Chou, D. A. Dahl, G. Mccobb, and D. Raggett, "EMMA: Extensible Multi-Modal Annotation Markup language," 2005.

[26] P. Groth and L. Moreau, "Prov-o w3c recommendation," W3C, Tech. Rep., 2013. [Online]. Available: http://www.w3.org/TR/prov-o/

[27] J. F. Sánchez-Rada, C. A. Iglesias, I. Corcuera-Platas, and O. Araque, "Senpy: A pragmatic linked sentiment analysis framework," in *Proc. DSAA Special Track on Emotion and Sentiment in Intelligent Systems and Big Social Data Analysis (SentISData)*, 2016.

[28] S. M. Kim, A. Valitutti, and R. A. Calvo, "Evaluation of unsupervised emotion models to textual affect recognition," in *Proc. NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Association for Computational Linguistics, 2010, pp. 62–70.

[29] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," University of Florida, Tech. Rep., 2010.

[30] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1627–45, sep 2010.

[31] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proc. 21st ACM International Conference on Multimedia*, ser. MM '13. New York, NY, USA: ACM, 2013, pp. 835–838.

[32] M. Schmitt, F. Ringeval, and B. Schuller, "At the border of acoustics and linguistics: Bag-of-audio-words for the recognition of emotions in speech," in *Proc. 17th International Speech Communication Association*. San Francisco, USA: ISCA, 2016, pp. 495–499.