EmotionML – an upcoming standard for representing emotions and related states

Marc Schröder[1], Paolo Baggia[2], Felix Burkhardt[3], Catherine Pelachaud[4], Christian Peter[5], and Enrico Zovato[2]

[1] DFKI GmbH, Saarbrücken, Germany
[2] Loquendo S.p.A., Torino, Italy
[3] Deutsche Telekom AG, Berlin, Germany
[4] Telecom ParisTech, Paris, France
[5] Fraunhofer IGD, Rostock, Germany
http://www.w3.org/TR/emotionml/

Abstract. The present paper describes the specification of Emotion Markup Language (EmotionML) 1.0, which is undergoing standardisation at the World Wide Web Consortium (W3C). The language aims to strike a balance between practical applicability and scientific well-foundedness. We briefly review the history of the process leading to the standardisation of EmotionML. We describe the syntax of EmotionML as well as the vocabularies that are made available to describe emotions in terms of categories, dimensions, appraisals and/or action tendencies. The paper concludes with a number of relevant aspects of emotion that are not covered by the current specification.

1 Introduction

Computerised systems, to the extent that they can recognise, simulate or otherwise process emotion-related information, need a representation format. If several components are to work collaboratively on the information, the format must be well-defined. In order to reach the best possible interoperability, a standard representation format should be used. This paper describes a long-running collaborative effort on defining and standardising an Emotion Markup Language.

The word "emotion" is used here in a very broad sense, covering both intense and weak states, short and long term, with and without event focus. This meaning is intended to reflect the understanding of the term "emotion" by the general public rather than any specific scientific theory.

The work started informally in 2006 as an "Emotion Annotation and Representation Language (EARL)" [23]. EARL tried to cover a lot in a short time, spanning the range from scientific descriptions of emotions, via use cases and requirements of technological applications, to the definition of an XML syntax.

Work then moved to the World Wide Web Consortium (W3C) in the form of two Incubator groups: first, the Emotion Incubator Group worked on use cases and requirements [18,17]; next, the Emotion Markup Language Incubator Group prioritised the requirements [1] and proposed elements of a syntax to address them [19]. Three main use cases were identified: (1) manual annotation of data; (2) automatic recognition of emotion-related states from user behavior; and (3) generation of emotion-related system behavior.

This exploratory work was formalised in the "Recommendation Track" at W3C in 2009. A First Public Working Draft (FPWD) of EmotionML 1.0 was published in 2009, followed by a second Working Draft in 2010 [20]. The specification process consisted mainly in resolving the open issues in the second Incubator report, in making the syntactic choices compatible with other works in W3C and in the multimodal interaction working group, and in ensuring that the syntax was sufficiently simple to be usable in real-world settings.

A W3C workshop on EmotionML was organised in October 2010 (`http://www.w3.org/2010/10/emotionml/cfp.html`) to invite feedback on the draft specification from scientific experts as well as from potential users. The workshop provided highly relevant feedback and clarification, and played an important role in the definition of the full specification published as a Last Call Working Draft (LCWD) in spring 2011 [21]. The definition of a number of vocabularies for EmotionML was published as a separate W3C Working Draft [22].

2 Previous work

The representation of emotions and related states has been part of several activities.

In the area of labelling schemes, maybe the most thorough attempt to propose an encompassing labelling scheme for emotion-related phenomena has been the work on the HUMAINE database [6]. The relevant concepts were identified, and made available as a set of configuration files for the video annotation tool Anvil [13]. A formal representation format was not proposed in this work.

Markup languages including emotion-related information were defined mainly in the context of research systems generating emotion-related behaviour of ECAs. The expressive richness is usually limited to a small set of emotion categories, possibly an intensity dimension, and in some cases a three-dimensional continuous representation of activation-evaluation-power space (see [24] for a review). For example, the Affective Presentation Markup Language APML [5] provides an attribute "affect" to encode an emotion category for an utterance (a "performative") or for a part of it:

```
<performative affect="afraid">
  Do I have to go to the dentist?
</performative>
```

An interesting contribution to the domain of computerised processing and representation of emotion-related concepts is A Layered Model of Affect, ALMA [10]. Following the OCC model [15], ALMA uses *appraisal* mechanisms to trigger emotions from events, objects and actions in the world. Emotions have an intensity varying over time. Each individual emotion influences mood as a longer-term affective state. ALMA uses an XML-based markup language named AffectML in two places: to represent the antecedents to emotion, i.e. the appraisals leading to emotions, or to represent the impact that emotions and moods have on a virtual agent's behaviour.

3 Syntax

The following snippet exemplifies the principles of the EmotionML syntax [21].

```
<sentence id="sent1">
  Do I have to go to the dentist?
</sentence>
<emotion xmlns="http://www.w3.org/2009/10/emotionml"
    category-set="http://www.w3.org/TR/emotion-voc/xml#everyday-categories">
  <category name="afraid" value="0.4"/>
  <reference role="expressedBy" uri="#sent1"/>
</emotion>
```

The following properties can be observed.

- The emotion annotation is self-contained within an '<emotion>' element;
- all emotion elements belong to a specific namespace;
- it is explicit in the example that emotion is represented in terms of categories;
- it is explicit from which category set the category label is chosen;
- the link to the annotated material is realised via a reference using a URI, and the reference has an explicit role.

We will now discuss the properties of the EmotionML syntax in some more detail.

3.1 Design principles: self-contained emotion annotation

EmotionML is conceived as a plug-in language, with the aim to be usable in many different contexts. Therefore, proper encapsulation is essential. All information concerning an individual emotion annotation is contained within a single '<emotion>' element. All emotion markup belongs to a unique XML namespace. EmotionML differs from many other markup languages in the sense that it does not *enclose* the annotated material. In order to link the emotion markup with the annotated material, either the reference mechanism in EmotionML or another mechanism external to EmotionML can be used.

Structurally, EmotionML uses element and attribute names to indicate the type of information being represented; attribute values provide the actual information. The use of attribute values (e.g., '<category name="joy"/>') was preferred over enclosed text (e.g., '<category>joy</category>') so that adding EmotionML to an XML node does not change that node's text content.

A top-level element '<emotionml>' enables the creation of stand-alone EmotionML documents, essentially grouping a number of emotion annotations together, but also providing document-level mechanisms for annotating global metadata and for defining emotion vocabularies (see below). It is thus possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts.

3.2 Representations of emotion

Emotions can be represented in terms of four types of descriptions taken from the scientific literature [24]: '`<category>`', '`<dimension>`', '`<appraisal>`' and '`<action-tendency>`'. An '`<emotion>`' element can contain one ore more of these descriptors; each descriptor must have a '`name`' attribute and can have a '`value`' attribute indicating the intensity of the respective descriptor. For '`<dimension>`', the '`value`' attribute is mandatory, since a dimensional emotion description is always a position on one or more scales; for the other descriptions, it is possible to omit the '`value`' to only make a binary statement about the presence of a given category, appraisal or action tendency.

The following example illustrates a number of possible uses of the core emotion representations.

```
<category name="affectionate"/>
<dimension name="valence" value="0.9"/>
<appraisal name="agent-self"/>
<action-tendency name="approach"/>
```

3.3 Mechanism for referring to an emotion vocabulary

Since there is no single agreed vocabulary for each of the four types of emotion descriptions (see Section 4), EmotionML provides a mandatory mechanism for identifying the vocabulary used in a given '`<emotion>`'. The mechanism consists in attributes of '`<emotion>`' named '`category-set`', '`dimension-set`' etc., indicating which vocabulary of descriptors for annotating categories, dimensions, appraisals and action tendencies are used in that emotion annotation. These attributes contain a URI pointing to an XML representation of a vocabulary definition (see Section 4). In order to verify that an emotion annotation is valid, an EmotionML processor must retrieve the vocabulary definition and check that every '`name`' of a corresponding descriptor is part of that vocabulary.

For example, the following annotation uses Mehrabian's PAD model [14] for representing a position in three-dimensional space.

```
<emotion dimension-set="http://www.w3.org/TR/emotion-voc/xml#pad-dimensions">
    <dimension name="arousal" value="0.3"/>   <!-- lower-than-average arousal -->
    <dimension name="pleasure" value="0.9"/>  <!-- very high positive valence -->
    <dimension name="dominance" value="0.8"/> <!-- relatively high potency    -->
</emotion>
```

3.4 Meta-information

Several types of meta-information can be represented in EmotionML.

First, each emotion descriptor (such as '<category>') can have a 'confidence' attribute to indicate the expected reliability of this piece of the annotation. This can reflect the confidence of a human annotator or the probability computed by a machine classifier. If several descriptors are used jointly within an '<emotion>', each descriptor has its own 'confidence' attribute. For example, it is possible to have high confidence in, say, the arousal dimension but be uncertain about the pleasure dimension:

```
<emotion dimension-set="http://www.w3.org/TR/emotion-voc/xml#pad-dimensions">
    <dimension name="arousal" value="0.7" confidence="0.9"/>
    <dimension name="pleasure" value="0.6" confidence="0.3"/>
</emotion>
```

Each '<emotion>' can have an 'expressed-through' attribute providing a list of modalities through which the emotion is expressed. Given the open-ended application domains for EmotionML, it is naturally difficult to provide a complete list of relevant modalities. The solution provided in EmotionML is to propose a list of human-centric modalities, such as 'gaze', 'face', 'voice', etc., and to allow arbitrary additional values. The following example represents a case where an emotion is recognised from, or to be generated in, face and voice:

```
<emotion category-set="http://www.w3.org/TR/emotion-voc/xml#everyday-categories"
         expressed-through="face voice">
    <category name="satisfaction"/>
</emotion>
```

For arbitrary additional metadata, EmotionML provides an '<info>' element which can contain arbitrary XML structures. The '<info>' element can occur as a child of '<emotion>' to provide local metadata, i.e. additional information about the specific emotion annotation; it can also occur in standalone EmotionML documents as a child of the root node '<emotionml>' to provide global metadata, i.e. information that is constant for all emotion annotations in the document. This can include information about sensor settings, annotator identities, situational context etc.

3.5 References to the "rest of the world"

Emotion annotation is always about something. There is a subject "experiencing" (or simulating) the emotion. This can be a human, a virtual agent, a robot, etc. There is observable behaviour expressing the emotion, such as facial expressions, gestures, or vocal effects. With suitable measurement tools, this can also include physiological changes such as sweating or a change in heart rate or blood pressure. Emotions are often caused or triggered by an identifiable entity, such as a person, an object, an event, etc. More precisely, the appraisals leading to the emotion are triggered by that entity. And finally, emotions, or more precisely the emotion-related action tendencies, may be directed towards an entity, such as a person or an object.

EmotionML considers all of these external entities to be out of scope of the language itself; however, it provides a generic mechanism for referring to such entities. Each '<emotion>' can use one or more '<reference>' elements to point to arbitrary URIs. A '<reference>' has a 'role' attribute, which can have one of the following four values: 'expressedBy' (default), 'experiencedBy', 'triggeredBy', and 'targetedAt'. Using this mechanism, it is possible to point to arbitrary entities filling the above-mentioned four roles; all that is required is that these entities be identified by a URI.

3.6 Time

Time is relevant to EmotionML in the sense that it is necessary to represent the time during which an emotion annotation is applicable. In this sense, temporal specification complements the above-mentioned reference mechanism.

Representing time is an astonishingly complex issue. A number of different mechanisms are required to cover the range of possible use cases. First, it may be necessary to link to a time span in media, such as video or audio recordings. For this purpose, the '<reference role="expressedBy">' mechanism can use a so-called Media Fragment URI [25] to point to a time span within the media. Second, time may be represented on an absolute or relative scale. EmotionML follows EMMA [12] in representing time in these cases. Absolute time is represented in milliseconds since 1 January 1970, using the attributes 'start' and 'end'. A combination of the 'start' and 'duration' attributes can also be used to represent time intervals. Absolute times are useful for applications such as affective diaries, which record emotions throughout the day, and whose purpose it is to link back emotions to the situations in which they were encountered. Other applications require relative time, for example time since the start of a session. Here, the mechanism borrowed from EMMA is the combination of 'time-ref-uri' and 'offset-to-start'. The former provides a reference to the entity defining the meaning of time 0; the latter is time, in milliseconds, since that moment.

3.7 Representing continuous values and dynamic changes

As mentioned above, the emotion descriptors '<category>', '<dimension>', etc. can have a 'value' attribute to indicate the position on a scale corresponding to the respective descriptor. In the case of a dimension, the value indicates the position on that dimension, which is mandatory information for dimensions; in the case of categories, appraisals and action tendencies, the value can be optionally used to indicate the extent to which the respective item is present.

In all cases, the 'value' attribute contains a floating-point number between 0 and 1. The two end points of that scale represent the most extreme possible values, for example the lowest and highest possible positions on a dimension, or the complete absence of an emotion category vs. the most intense possible state of that category.

The 'value' attribute thus provides a fine-grained control of the position on a scale, which is constant throughout the temporal scope of the individual

'<emotion>' annotation. It is also possible to represent changes over time of these scale values, using the '<trace>' element which can be a child of any '<category>', '<dimension>', '<appraisal>' or '<action-tendency>' element. This makes it possible to encode trace-type annotations of emotions as produced, e.g., by FeelTrace [4].

4 Vocabularies for EmotionML

As described above, EmotionML takes into account a number of key concepts from scientific emotion research [24]. Four types of descriptions are available: categories, dimensions, appraisals, and action tendencies. Depending on the tradition of emotion research and on the use case, it may be appropriate to use any single one of these representations; alternatively, it may also make sense to use *combinations* of descriptions to characterise more fully the various aspects of an emotional state that are observed: how an appraisal of triggers caused the emotion; how it can be characterised using a global description in terms of a category and/or a set of dimensions; and the potential actions the individual may be executing as a result. Insofar, EmotionML is a powerful representational device.

This description glosses over one important detail, however. Whereas emotion researchers may agree to some extent on the types of facets that play a role in the emotion process (such as appraisals, feeling, expression, etc.), there is no general consensus on the descriptive vocabularies that should be used. Which set of emotion *categories* is considered appropriate varies dramatically between the different traditions, and even within a tradition such as the Darwinian tradition of emotion research, there is no agreed set of emotion categories that should be considered as the most important ones (see e.g. [2]). Similarly, dimensional accounts of emotion do not agree on either the number or the names that should be given to the different dimensions.

For this reason, any attempt to enforce a closed set of descriptors for emotions would invariably draw heavy criticism from a range of research fields. Given that there is no consensus in the community, it is impossible to produce a consensus annotation in a standard markup language. The obvious alternative is to leave the choice of descriptors up to the users; however, this would dramatically limit interoperability.

The solution pursued in EmotionML is of a third kind. The notion of an 'emotion vocabulary' is introduced: any specific emotion annotation must be specific about the vocabulary that is being used in that annotation. This makes it possible to define in a clear way the terms that make sense in a given research tradition. Components can interoperate if the EmotionML markup that they produce and consume uses one or more emotion vocabularies that all components involved are able to handle.

The specification includes a mechanism for defining emotion vocabularies. It consists of a '<vocabulary>' element containing a number of '<item>' elements. A vocabulary has a 'type' attribute, indicating whether it is a vocabulary for representing categories, dimensions, appraisals or action tendencies. A vocabulary item has a 'name' attribute. Both the entire vocabulary and each individual item can have an '<info>' child to provide arbitrary metadata.

A W3C Working Draft [22] complements the specification to provide EmotionML with a set of emotion vocabularies taken from the scientific literature. When the user considers them suitable, these vocabularies rather than arbitrary other vocabularies should be used in order to promote interoperability. Whenever users have a need for a different vocabulary, however, they can simply define their own custom vocabulary and use it in the same way as the vocabularies listed in the document. This makes it possible to add any vocabularies from scientific research that are missing from the pre-defined set, as well as application-specific vocabularies.

In selecting emotion vocabularies, the group has applied the following criteria. The primary guiding principle has been to select vocabularies that are either commonly used in technological contexts, or represent current emotion models from the scientific literature. A further criterion is related to the difficulty to define mappings between categories, dimensions, appraisals and action tendencies. For this reason, groups of vocabularies were included for which some of these mappings are likely to be definable in the future.

The following vocabularies are defined. For categorical descriptions, the "big six" basic emotion vocabulary by Ekman [7], an everyday emotion vocabulary by Cowie et al. [3], and three sets of categories that lend themselves to mappings to appraisals, dimensions and action tendencies: the OCC categories [15], the categories used by Fontaine et al. [8], and the categories from the work by Frijda [9]. Three dimensional vocabularies are provided, the pleasure-arousal-dominance (PAD) vocabulary by Mehrabian [14], the four-dimensional vocabulary proposed by Fontaine et al. [8], and a vocabulary providing a single 'intensity' dimension for such use cases that want to represent solely the intensity of an emotion without any statement regarding the nature of that emotion. For appraisal, three vocabularies are proposed: the OCC appraisals [15], Scherer's Stimulus Evaluation Checks [16], and the EMA appraisals [11]. Finally, for action tendencies, only a single vocabulary is currently listed, namely that proposed by Frijda [9].

While these vocabularies should provide users with a solid basis, it is likely that additional vocabularies or clarifications about the current vocabularies will be requested. Due to the rather informal nature of a non-Recommendation-track Working Draft, it is rather easy to provide future versions of the document that provide the additional information required.

5 Conclusion and future work

The EmotionML 1.0 specification addresses the majority of the requirements that arise from use cases. In a future call for implementations, the implementability of all features provided by the specification will be verified.

A number of important issues have been noted as important but too difficult to handle in the first version of EmotionML. Among these is a careful solution for representing *regulation* in EmotionML, i.e. the fact that an emotion was suppressed, simulated, masked by another emotion, etc. Another requirement that is not covered in EmotionML 1.0 is the use of ontologies to define the terms in an emotion vocabulary, to relate the terms to one another, and to define mappings between emotion vocabularies where possible. Another difficulty regards the specification of scales. Should it be discrete, continuous, unipolar or bipolar, etc.? Due to the difficulty of finding a consensus in the emotion community on best practice for scales, we have postponed a more detailed definition of scales.

Once that EmotionML 1.0 has reached its full maturity, these directions can be developed in future versions of EmotionML.

Acknowledgements

The preparation of this paper was supported by the W3C and the EU project ALIZ-E (IST-248116). Thanks to all who have provided input to this work, especially Roddy Cowie, Sylwia Hyniewska, and Jean-Claude Martin.

References

1. Burkhardt, F., Schröder, M.: Emotion markup language: Requirements with priorities. W3C incubator group report, World Wide Web Consortium (2008), http://www.w3.org/2005/Incubator/emotion/XGR-requirements-20080513/
2. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Communication 40(1-2), 5–32 (2003)
3. Cowie, R., Douglas-Cowie, E., Appolloni, B., Taylor, J., Romano, A., Fellenz, W.: What a neural net needs to know about emotion words. In: Mastorakis, N. (ed.) Computational Intelligence and Applications, pp. 109–114. World Scientific & Engineering Society Press (1999)
4. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröder, M.: 'FEELTRACE': an instrument for recording perceived emotion in real time. In: Proc. ISCA Workshop on Speech and Emotion. pp. 19–24. Northern Ireland (2000)
5. De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a markup language for believable behavior generation. In: Prendinger, H., Ishizuka, M. (eds.) Life-Like Characters, pp. 65–85. Springer, New York (2004)
6. Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., Martin, J., Devillers, L., Abrilian, S., Batliner, A., Amir, N., Karpouzis, K.: The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In: Proc. Affective Computing and Intelligent Interaction. pp. 488–500. Lisbon, Portugal (2007), http://dx.doi.org/10.1007/978-3-540-74889-2_43
7. Ekman, P.: Universals and cultural differences in facial expressions of emotion. In: Cole, J. (ed.) Nebraska Symposium on Motivation, vol. 19, pp. 207–282. University of Nebraska Press (1972)
8. Fontaine, J.R., Scherer, K.R., Roesch, E.B., Ellsworth, P.C.: The world of emotions is not Two-Dimensional. Psychological Science 18(12), 1050–1057 (2007), doi:10.1111/j.1467-9280.2007.02024.x

9. Frijda, N.H.: The Emotions. Cambridge University Press, Cambridge, UK (1986)
10. Gebhard, P.: ALMA - a layered model of affect. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-05). Utrecht (2005)
11. Gratch, J., Marsella, S.: A domain-independent framework for modeling emotion. Cognitive Systems Research 5(4), 269–306 (2004)
12. Johnston, M., Baggia, P., Burnett, D.C., Carter, J., Dahl, D.A., McCobb, G., Raggett, D.: EMMA: extensible MultiModal annotation markup language (Feb 2009), http://www.w3.org/TR/emma/
13. Kipp, M.: Anvil - a generic annotation tool for multimodal dialogue. In: Proc. Eurospeech. pp. 1367–1370. Aalborg, Denmark (2001)
14. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology 14(4), 261–292 (Dec 1996), http://dx.doi.org/10.1007/BF02686918
15. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotion. Cambridge University Press, Cambridge, UK (1988)
16. Scherer, K.R.: Appraisal theory. In: Dalgleish, T., Power, M.J. (eds.) Handbook of Cognition & Emotion, pp. 637–663. John Wiley, New York (1999)
17. Schröder, M., Devillers, L., Karpouzis, K., Martin, J., Pelachaud, C., Peter, C., Pirker, H., Schuller, B., Tao, J., Wilson, I.: What should a generic emotion markup language be able to represent? In: Proc. 2nd International Conference on Affective Computing and Intelligent Interaction (ACII'2007). Lisbon, Portugal (2007)
18. Schröder, M., Zovato, E., Pirker, H., Peter, C., Burkhardt, F.: W3C emotion incubator group final report. Tech. rep., World Wide Web Consortium (2007), http://www.w3.org/2005/Incubator/emotion/XGR-emotion-20070710
19. Schröder, M., Baggia, P., Burkhardt, F., Martin, J., Pelachaud, C., Peter, C., Schuller, B., Wilson, I., Zovato, E.: Elements of an EmotionML 1.0. W3C final incubator group report, World Wide Web Consortium (2008), http://www.w3.org/2005/Incubator/emotion/XGR-emotionml-20081120/
20. Schröder, M., Baggia, P., Burkhardt, F., Oltramari, A., Pelachaud, C., Peter, C., Zovato, E.: Emotion markup language (EmotionML) 1.0. W3C working draft, World Wide Web Consortium (Jul 2010), http://www.w3.org/TR/2010/WD-emotionml-20100729/
21. Schröder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C., Zovato, E.: Emotion markup language (EmotionML) 1.0. W3C Last Call Working Draft, World Wide Web Consortium (Apr 2011), http://www.w3.org/TR/2011/WD-emotionml-20110407/
22. Schröder, M., Pelachaud, C., Ashimura, K., Baggia, P., Burkhardt, F., Oltramari, A., Peter, C., Zovato, E.: Vocabularies for EmotionML. W3C Working Draft, World Wide Web Consortium (Apr 2011), http://www.w3.org/TR/2011/WD-emotion-voc-20110407/
23. Schröder, M., Pirker, H., Lamolle, M.: First suggestions for an emotion annotation and representation language. In: Proceedings of LREC'06 Workshop on Corpora for Research on Emotion and Affect. pp. 88–92. Genoa, Italy (2006)
24. Schröder, M., Pirker, H., Lamolle, M., Burkhardt, F., Peter, C., Zovato, E.: Representing emotions and related states in technological systems. In: Petta, P., Cowie, R., Pelachaud, C. (eds.) Emotion-Oriented Systems – The Humaine Handbook, pp. 367–386. Springer (2011)
25. Troncy, R., Mannens, E., Pfeiffer, S., van Deursen, D.: Media fragments URI 1.0. W3C last call working draft, World Wide Web Consortium (Jun 2010), http://www.w3.org/TR/1998/REC-xml-19980210