Application of EmotionML

Felix Burkhardt[1], Christian Becker-Asano[2], Edmon Begoli[3], Roddy Cowie[4], Gerhard Fobe[5], Patrick Gebhard[6], Abe Kazemzadeh[7], Ingmar Steiner[6,8], Tim Llewellyn[9]

[1]Deutsche Telekom Laboratories, Berlin, Germany, [2]Albert-Ludwigs-Universität, Freiburg, Germany, [3]University of Tennessee, Knoxville, USA, [4]Queen's University Belfast, UK, [5]Technische Universität Chemnitz, Germany, [6]DFKI, Saarbrücken, Germany, [7]University of Southern California, USA, [8]Saarland University, Saarbrücken, Germany, [9]nViso, Lausanne, Switzerland

Felix.Burkhardt@telekom.de

Abstract

We present EmotionML, a new W3C recommendation to represent emotion related states in data processing systems, by first introducing the language and then discussing a series of concrete implementations that utilize EmotionML.

Keywords: emotionml, applications, sentiment

1. Introduction

We present EmotionML[1], a new W3C recommendation to represent emotion related states in data processing systems as well as a series of concrete implementations that utilize EmotionML.

EmotionML was developed by a subgroup of the W3C MMI (Multimodal Interaction) Working Group chaired by Deborah Dahl in a first version from approximately 2005 until 2013, most of this time the development was lead by Marc Schröder.

In the scientific literature on emotion research, there is no single agreed description of emotions, not even a clear consensus on the use of terms like affect, emotion or other related phenomena. For a markup language representing emotional phenomena it therefore appears important to allow the representation of their most relevant aspects in the wider sense. Given the lack of agreement in the literature on the most relevant aspects of emotion, it is inevitable to provide a relatively rich set of descriptive mechanisms.

The working group iteratively extracted requirements on the markup language from a number of 39 collected use cases[2]. Based on these requirements, a syntax for EmotionML has been produced.

It is possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts. Emotions can be represented in terms of four types of descriptions taken from the scientific literature: categories, dimensions, appraisals, and action tendencies, with a single <emotion> element containing one or more of such descriptors.

The first part of the paper deals with a short summary of EmotionML by describing selected aspects and the procedure and thinking behind its development. The second half introduces a number of applications that integrated EmotionML and were submitted as implementation reports during the W3C recommendation track process.

2. Overview of EmotionML

Based on the requirements, a syntax for EmotionML (Schröder et al., 2012) has been produced in a sequence of steps.

The following snippet exemplifies the principles of the EmotionML syntax (Burkhardt et al., 2013).

```
<sentence id="sent1">
  Do I have to go to the dentist?
</sentence>
<emotion xmlns="http://www.w3.org/
  2009/10/emotionml" category-set=
  "http://.../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 realized via a reference using a URI, and the reference has an explicit role.

2.1. 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.

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

[1]http://www.w3.org/TR/emotionml/

[2]http://www.w3.org/2005/Incubator/emotion/XGR-emotion/#AppendixUseCases

meta data 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.

2.2. Representations of emotion

Emotions can be represented in terms of four types of descriptions taken from the scientific literature (Schröder et al., 2011): <category>, <dimension>, <appraisal>, and <action-tendency>. An <emotion> element can contain one or 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"/>
```

2.3. Mechanism for referring to an emotion vocabulary

Since there is no single agreed-upon vocabulary for each of the four types of emotion descriptions, 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. 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.
Some vocabularies are suggested by the W3C (Schröder et al., 2012) and to make EmotionML documents interoperable users are encouraged to use them.

2.4. 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 recognized from, or to be generated in, face and voice:

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

For arbitrary additional meta data, EmotionML provides an <info> element which can contain arbitrary XML structures. The <info> element can occur as a child of <emotion> to provide local meta data, 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 meta data, 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.

2.5. 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 behavior 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.

2.6. 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 to point to a time span within the media. Second, time may be represented on an absolute or relative scale. Absolute time is represented in milliseconds since 1 January 1970, using the attributes `start`, `end` and `duration`. 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.

2.7. Representing continuous values and dynamic changes

A 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.

3. Selected Applications

This section discusses several implementations that integrated EmotionML. Common to them is that they were submitted as an implementation report to the W3C during the recommendation track process[3]. The implementations concern very different aspects of emotion related machine processing, which reflects the diversity of the field. We categorize them in four areas: research related, core libraries, frameworks, and commercial applications.

3.1. Research related

These applications deal primarily with research questions on the nature of emotion related states.

3.1.1. EMO20Q

Emotion twenty questions (EMO20Q) is an experimental dialog game that is used to study how people describe emotions with language. By gamifying the question-asking discourse and collecting large amounts of data, EMO20Q aims to to define emotion words through crowd-sourcing (Kazemzadeh et al., 2011). Storing the belief state in EmotionML makes it possible to persist the agent's belief state in cases where the dialog is implemented in a transactional setting, such as HTTP where the agent's context must be reloaded for each request.

3.1.2. Gtrace

Gtrace (General Trace program) by the Queen's University Belfast is the successor to FEELtrace and the tools used to label the HUMAINE database (Cowie and Douglas-Cowie, 2012). It allows users to play a video of a person and create "traces" which show how the person's emotions appear to be changing over time. It includes over 50 scales, and also allows users to create their own. Alternative ways of using the scales are provided. It runs on current versions of Windows. A manual provides broad background as well as instructions for use. The system currently implements EmotionML by tracing for category and dimensional descriptors.

3.2. Libraries

Some libraries for different programming languages have already been developed by the community. In addition, there is also one for Java from Alexandre Denis at LORIA (Nancy, France)[4] and a library to check on the validity of EmotionML documents by Marc Schröder[5].

3.2.1. C# library

The EmotionML C# library[6] was developed at the University of Chemnitz as part of a project dealing with emoticons like smileys or emojis and the issues of this kind of emotion representation during the interaction in an intercultural text based chat (Fobe, 2012).
With the help of the integrated EmotionML-parser it is possible to create related object instances automatically. Furthermore object instances can be converted to EmotionML as well (in DOM and XML mode). Beside a standalone EmotionML document the plug-in version for the inclusion of emotions in other languages is supported.

[3] http://www.w3.org/2002/mmi/2013/emotionml-ir/

[4] http://code.google.com/p/loria-synalp-emotionml/

[5] https://github.com/marc1s/emotionml-checker-java

[6] https://github.com/gfobe/EmotionML-Lib-CSharp

3.2.2. EMLPy

EMLPy is a generator library for EmotionML documents[7]. It is a Python based library intended as a utility to be invoked from other EmotionML programs. EMLPy generates EmotionML documents by transforming the user specified and populated Python object tree into a XML representation. EMLPy performs EmotionML checks covered in assertions while executing this object to XML transformation. From an API perspective, the user interacts with an object tree hierarchy that maps directly to an EmotionML hierarchy of elements and attributes. EMLPy validates the object tree and its properties against the EmotionML schema and specification rules.

3.3. Frameworks

The following examples illustrate the use as part of a larger framework used in different contexts.

3.3.1. ALMA

ALMA EmotionML is an extension extension to the ALMA computational model of affect. ALMA allows the the real-time simulation of three basic types of affective features that humans can experience: (1) emotions reflect discrete short-term affect that decays after a short period of time; (2) moods reflect continuous medium-term affect, which is generally not related to a concrete event, action, or object; and (3) personality reflects discrete individual differences in mental characteristics and affective dispositions. The simulation is based on situational appraisal of the current situation according to the cognitive model of emotions created by Ortony, Clore, and Collins (OCC) (Ortony et al., 1988). ALMA combines this with the Big Five model of personality (McCrae and John, 1992) and a simulation of mood based on the Pleasure, Arousal, and Dominance (PAD) model (Mehrabian, 1996).

The ALMA EmotionML[8] implementation supports most of the EmotionML standard: (1) appraisal representation, (2) discrete and continuous emotion and mood representation, and (3) PAD and OCC emotional vocabularies. All computational output, e.g. intensities of current active emotions, or the current mood are described in an EmotionML representation. The EmotionML extension allows a fine-grained control of affect related body behavior of virtual characters, like emotional facial expressions or mood related posture control.

3.3.2. WASABI

WASABI[9] is an architecture for affect simulation for believable interactivity (Becker-Asano., 2008). It was initially developed to enhance the believability of the virtual human MAX at University of Bielefeld. Since then, it was integrated into several virtual and robotic agent systems (Becker-Asano, 2014). It realizes the concurrent simulation of emotion dynamics based on the interaction between emotion and mood and it utilizes the PAD emotional vocabulary. Its specification uses EmotionML extended by several `<info>` elements to define WASABI-specific parameters. Its UDP-based network output can be configured to represent its internal dynamics in terms of `<dimension>` elements in combination with the `<trace>` element. Thereby, it has proven easy to adjust WASABI's configuration to the project's needs and to interface it with other soft- and hardware modules, such as MARY TTS.

3.3.3. MARY TTS

MARY TTS[10] is an open-source, multilingual text-to-speech synthesis platform that includes modules for expressive speech synthesis (Charfuelan and Steiner, 2013). Particularly the support for both categorical and dimensional representations of emotions by EmotionML is important to MARY's expressive speech synthesis. These categories and dimensions are implemented by modifying the predicted pitch contours, pitch level, and speaking rate.

Using this approach, expressive synthesis is most effective when using HMM-based voices, since the statistical parametric synthesis framework allows appropriate prosody to be realized with consistent quality. Expressive unit-selection voices support EmotionML best if they are built from multiple-style speech databases (Steiner et al., 2013), which preserves intonation and voice quality better than when applying signal manipulation to conventional unit-selection output.

3.4. Applications

Lastly, the following lists commercial applications that utilize EmotionML to represent emotion related models.

3.4.1. NViso

NViso uses emotion detection to analyze customer reaction on brands and (web) interfaces (nViso, 2011). It provides a cloud service to measure instantaneous emotional reactions of consumers in online environments and thus provides real-time information for Market Research, Brands, Creative Agencies and R&D Product Development.

The NViso 3D Facial Imaging API is an online service for the recognition of emotions depicted through facial expressions in still images and videos. The focus of the integration of EmotionML into the tool is on using the media type and URI time for video.

3.4.2. Speechalyzer

The Speechalyzer by Deutsche Telekom Laboratories is an open source project[11] for analysis, annotation and transcription of speech files (Burkhardt, 2011). It can be used to rapidly judge large numbers of audio files emotionally, an automatic classification is integrated. The Speechalyzer was part of a project to identify disgruntled customers in an automated voice service portal (Burkhardt et al., 2009) with two use cases in mode: a) transfer angry users to a trained human agent, and b) gain some statistic insight on the number of angry customers at the end of each day. It utilizes EmotionML as an exchange format to import and export emotionally annotated speech data.

---

[7] https://github.com/ebegoli/EMLPy

[8] ALMA is freely available for download: http://www.dfki.de/~gebhard/alma

[9] https://github.com/CBA2011

[10] http://mary.dfki.de/ and https://github.com/marytts

[11] https://github.com/dtag-dbu/speechalyzer

4. Conclusions

We presented EmotionML, a new W3C recommendation to represent emotion related states. The first part of the paper deals with a short summary of EmotionML and the second half introduces a number of applications that integrated EmotionML and were submitted as implementation reports during the W3C recommendation track process. We hope this article encourages the reader to use EmotionML in own projects and give feedback to the W3C to pave the way towards EmotionML version 2.0.

5. References

Becker-Asano., C. (2008). *WASABI: Affect Simulation for Agents with Believable Interactivity*. Ph.D. thesis, University of Bielefeld.

Becker-Asano, C. (2014). WASABI for affect simulation in human-computer interaction. In *Proc. International Workshop on Emotion Representations and Modelling for HCI Systems*.

Burkhardt, F., Polzehl, T., Stegmann, J., Metze, F., and Huber, R. (2009). Detecting real life anger. In *Proc. International Conference on Acoustics, Speech and Signal Processing*.

Burkhardt, F., Schröder, M., Baggia, P., Pelachaud, C., Peter, C., and Zovato, E. (2013). W3C Emotion Markup Language (EmotionML) 1.0 proposed recommendation. http://www.w3.org/TR/emotionml/.

Burkhardt, F. (2011). Speechalyzer: a software tool to process speech data. In *Proc. Elektronische Sprachsignalverarbeitung*.

Charfuelan, M. and Steiner, I. (2013). Expressive speech synthesis in MARY TTS using audiobook data and EmotionML. In *Proc. Interspeech*.

Cowie, Roddy, G. M. and Douglas-Cowie, E. (2012). Tracing emotion: An overview. *International Journal of Synthetic Emotions (IJSE)*, 3(1):1–17.

Fobe, G. (2012). Serialisierung von Emotionen in der textuellen Kommunikation. Master's thesis, Technical University of Chemnitz.

Kazemzadeh, A., Lee, S., Georgiou, P. G., and Narayanan, S. S. (2011). Emotion twenty questions: Toward a crowd-sourced theory of emotions. In *Proc. Affective Computing and Intelligent Interaction (ACII)*.

McCrae, R. and John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292.

nViso. (2011). http://nviso.ch.

Ortony, A., Clore, G. L., and Collins., A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

Schröder, M., Pirker, H., Lamolle, M., Burkhardt, F., Peter, C., and Zovato, E. (2011). Representing emotions and related states in technological systems. In Petta, P., Cowie, R., and Pelachaud, C., editors, *Emotion-Oriented Systems – The Humaine Handbook*, pages 367–386. Springer.

Schröder, M., Pelachaud, C., Ashimura, K., Baggia, P., Burkhardt, F., Oltramari, A., Peter, C., and Zovato, E. (2012). Vocabularies for EmotionML. http://www.w3.org/TR/emotion-voc/.

Steiner, I., Schröder, M., and Klepp, A. (2013). The PAVOQUE corpus as a resource for analysis and synthesis of expressive speech. In *Proc. Phonetik & Phonologie 9*.