What is most important for an Emotion Markup Language?

Marc Schröder[1], Ian Wilson[2], William Jarrold[3], Dylan Evans[4], Catherine Pelachaud[5], Enrico Zovato[6], and Kostas Karpouzis[7]

[1] DFKI GmbH, Saarbrücken, Germany
[2] Emotion AI, Cambridge, UK
[3] SRI International, Menlo Park, CA, USA
[4] Cork Constraint Computation Centre, University College Cork, Ireland
[5] University of Paris 8, INRIA, France
[6] Loquendo, S.p.A., Turin, Italy
[7] Image, Video and Multimedia Systems Lab, Nat. Tech. Univ. Athens, Greece
http://www.w3.org/2005/Incubator/emotion

Abstract. The paper describes the consolidated set of mandatory and optional requirements for an Emotion Markup Language as identified by the W3C Emotion Markup Language Incubator group. It also exposes a number of relevant questions that arise from theoretical considerations as well as from the intended use of the markup language.

1 Introduction

Despite the growing relevance of affective computing in technological environments, there is still no common interchange language for applications that deal with emotion and affect. The W3C Emotion Incubator Group was established at the beginning of 2006 with the aim to fill this gap. The group, composed of W3C members and of invited experts from industry and academic institutions, aims to design a general-purpose emotion markup language that can be effectively used in different contexts and applications. Due to the large variety of emotion representations used both in scientific theory and in applications, the language must be compatible with a broad range of emotion theories, and must be customisable to some extent within a standard framework. Furthermore, the language must interoperate with other markups, particularly with those that are standards in multimedia applications.

The first year of the group's work was focused on defining the requirements of the new language [1, 2] on the basis of a bottom-up collection of use cases. Current work is moving towards a concrete specification for the language. Given the long initial list of requirements, it was necessary to distinguish between mandatory ("must have") and optional ("should have") requirements. 14 experts from industry and academia evaluated each of the 22 initial requirements according to its importance [3]. Agreement was high in most cases, leading to a natural separation into two groups. Borderline cases, including *action tendencies* and *regulation*, were discussed in the group before making a choice.

The present paper presents the outcome of the prioritisation on requirements, and reports on discussions of both theoretical and usage considerations.

2 "Must have"s for an EmotionML

This section describes the markup elements that are considered mandatory by the Emotion Markup Language Incubator Group. They comprise a sophisticated description of the emotion or related phenomenon itself; linking mechanisms to other items in the world; meta-information about the emotion annotation; and a generic notion of global metadata.

2.1 Emotion Core

The description of the emotion or related state itself naturally receives the most prominent place in an emotion markup language. Most of the items on the original requirements list [2] are considered mandatory, allowing the user to represent most of the multiple facets of an emotion [4], but also less intense affective states [5].

– **Core 1. Type of emotion-related phenomenon**

First, the language must allow the user to be explicit about the type of phenomenon that is being represented: an *emotion* in the strong sense, i.e. a momentary, intense episode triggered by a concrete event, or rather a mood, an attitude, an interpersonal stance, etc.

Which taxonomy to use for distinguishing types of emotion-related phenomena is an open research question, and as for other elements of the emotion markup language, any standard will only be able to propose a "default" answer. Users must be able to replace the suggested taxonomy with one that fits their own needs. One possible starting point for proposing a default set, from the literature on emotion theory, is provided by Scherer [5].

– **Core 2. Emotion categories**
– **Core 3. Emotion dimensions**
– **Core 4. Appraisals related to the emotion**
– **Core 5. Action tendencies**

Emotions can be described in a number of different ways, using categories, dimensions, appraisals, or action tendencies. Details of these have been described elsewhere (e.g., [4, 2]), so we only point out some issues relevant for defining a markup language.

Emotion categories can be chosen from a set of discrete labels. Dimensions, appraisals and action tendencies seem to be best represented as "scale" values. This raises interesting questions, see Section 4.1.

Each of the four types of representation needs a vocabulary of names for the categories, dimensions etc.; again, the aim will be to propose a meaningful "default", and allow users to use a different set if they have specific needs.

– **Core 6. Multiple and/or complex emotions**

Multiple emotions may be co-occurring in the same experiencer, e.g. when a person is angry about one thing and sad about another, when the face shows one emotion and the voice another, or in cases of regulation (see Core 9, below).

– **Core 7. Emotion intensity**

The intensity of an emotion is a unipolar scale; the question how to represent scales is not easily answered, though, see Section 4.1 for a short discussion.

– **Core 8. Emotion timing**

The temporal scope of an emotion markup may be defined through a combination of start and end times, or by linking to items located on the time line such as utterances or gestures. The *time course* of an emotion markup may be defined through a sampling mechanism, providing values at fixed intervals.

2.2 Meta-information about emotion annotation

– **Meta 1. Confidence / probability**

Both machine classifiers and human annotators need to indicate the degree of confidence that a certain element of the representation is correct.

– **Meta 2. Modality**

Emotion may be expressed specifically in a certain modality, e.g. face, voice body posture or hand gestures, but also lighting, font shape, etc.

2.3 Links to the "rest of the world"

Emotion markup is always about something. Providing suitable links to external entities is essential for the interpretation of the emotion markup.

– **Links 1. Links to media**
– **Links 2. Position on a time line in externally linked objects**

A generic linking mechanism is envisaged. A link may point to a media object, such as a picture, an audio or video file, or a node in an XML document; this may be complemented with timing information, such as a start time and a duration.

– **Links 3. The semantics of links to the "rest of the world"**

Links must be assigned a meaning. Initially, the following meanings are envisaged: the experiencer (who "has" the emotion); the observable behaviour "expressing" the emotion; the trigger, cause or eliciting event of the emotion; and the object or target of the emotion (i.e., the thing that the emotion is about).

2.4 Global metadata

In order to facilitate communication between a producer and a consumer of emotional data with respect to application-specific information, the emotional markup may need to contain global metadata.

– **Global 0. A generic mechanism to represent global metadata**

In the Must Have section, we have previewed only a relatively unspecific placeholder (hence the identifier "Global 0") for the various more specific but currently optional requirements for global metadata (see below).

3 "Should have"s for an EmotionML

The elements in this section are considered less important and urgent; if their implementation poses non-trivial problems, a first draft of the emotion markup language does not need to implement them. Nevertheless, they are needed for certain use cases, and should be added in future versions of the markup language.

3.1 Emotion Core

– **Core 9. Emotion regulation**

Regulation covers a range of manipulations of an emotion or its expression by the experiencer. In a very basic interpretation, this includes a difference between the internal and the externalised state, i.e. cases of simulation and suppression. However, considerably more complex models of emotion regulation are described in the literature (e.g., [6]), and a suitable degree of abstraction will need to be found; see also the discussion in Section 4.1.

3.2 Meta-information about emotion annotation

– **Meta 3. Acting**

Specific annotations are needed for describing the properties of acted material, such as perceived naturalness, authenticity, quality of acting, etc.

3.3 Global Metadata

– **Global 1. Info on Person(s)**
– **Global 2. Social and communicative environment**
– **Global 3. Purpose of classification**
– **Global 4. Technical environment**

Several specific kinds of global metadata are needed in different use cases in order to properly interpret the emotion markup in context. For some kinds of metadata (e.g., information about persons), it may be possible to reuse existing metadata annotation schemes; others (e.g., social environment) are specifically relevant in the context of emotion markup, and will need to be modelled explicitly. Purpose of classification and technical environment are needed for the emotion recognition use case, where they are needed to interpret the markup.

3.4 Ontologies of emotion descriptions

– **Onto 1. Mappings between different emotion representations**
– **Onto 2. Relationships between concepts in an emotion description**

Different emotion representations, and different concepts within an emotion representation, are not independent of one another; if their relation could be made explicit, that would allow for mappings (such as, locating an emotion category on dimensional scales), or for better interpretations (e.g., by making the similarity or difference between emotion categories explicit).

It remains to be seen whether such relationships are best modelled within the emotion markup language, or whether it is better to model them as a complement to the markup language.

4 Discussion

The targeted specification raises a considerable number of issues, combining questions of theoretical interest with practical considerations. The present section outlines a number of key questions that have been raised in the group.

4.1 Theoretical Issues

The Status of Action Tendencies Action tendencies are listed among the possible descriptions of emotions that are mandatory for the EmotionML. The concept stems from emotion theory – for example, according to Frijda [7, p. 88, Table 2.1], desire is linked to a tendency to approach, fear is linked to a tendency to avoid, and so on. Action tendencies are potentially very relevant for use cases where emotions play a role in driving behaviour, such as in the behaviour planning component of non-player characters in games, or in robot companions.

However, there are some interesting difficulties with this requirement. One of them is the dependency of action tendencies on the effectors available to the system, which have huge variation. As an example, an action tendency of "eat food" would make sense for systems capable of (simulating) eating whereas a more informational "consume energy" might be a more useful and wide ranging description that would equally apply to robots that recharge their batteries.

Another difficulty concerns the distinction between action tendencies and other forms of observable emotion behaviour such as facial expression. It may seem arbitrary to include a specific requirement for action tendencies in the markup language, while all other information about the observable behaviour "expressing" the emotion is relegated to a subcategory of metadata that specifies the semantics of links to the "rest of the world". However, there are sound practical reasons for this decision. For one thing, there are well-established standards for certain visual expressions of emotion that have already been specified outside of the proposed markup language, such as MPEG-4, H-anim, and FACS, while no such 'standard' representation language currently exists for behaviours. Also, there are indefinitely many forms of emotional expression, from facial expressions, physiological parameters in humans and humanoid agents to colours and flashing lights in robots. This is generically unlimited, and so it was decided to leave this information outside the emotion markup language, in order to keep the language reasonably compact.

Complex Emotions and Regulation Complex emotions are listed as a mandatory requirement for the EmotionML. This enables the representation of cases where an event may be evaluated from different perspectives leading to the superposition of two emotions, as well as cases where different emotions are apparent in different modalities.

Regulation, however, is listed as an optional requirement, despite the fact that it is an important source for complex emotions. For example, the display of an emotional state may be impeded due to some socio-cultural rules [8]: the expression of one emotion may be masked by another one, it may be inhibited, minimised or even exaggerated. Alternatively, it is possible to regulate, to some

extent, the emotion itself rather than its expression, through a process of re-appraisal [6]. Representing regulation in a scientifically appropriate way appears to be a non-trivial challenge; for this reason, we avoided making regulation a mandatory requirement for EmotionML, despite its importance for modelling certain types of complex emotions.

Scale Values Several of the elements listed in Sections 2 and 3 seem to be best represented as values along some sort of scale. A straightforward representation of scale values, and an obvious candidate for a "default" representation, would be a continuous unitless scale such as [0,1]. Such a standard value range would have the advantage of easily supporting interoperability between technological components; however, it would limit the use of EmotionML in several ways.

One potential deficiency of [0,1] is that it may be overly restrictive for representing emotion intensity. For example, is there some maximum amount of experienceable joy or despair? Independently of that theoretical question, it may be desirable for some applications to generate exaggerated values for emotions and their expressions, e.g. for cartoon animation.

Furthermore, human annotators often use a set of discrete labels such as a five-point Likert scale ("strongly disagree", "disagree", ..., "strongly agree"). Such scales are at best ordinal, but it would be misleading to map them to a metric interval such as [0,1] – the psychometric difference between "neither agree nor disagree" and "agree" cannot be assumed to be as large as the difference between "agree" and "strongly agree". Any such mapping would necessarily imply a loss of information and of accuracy.

Finally, human judgements may be class-specific (e.g., "For a child, Young Johnny exhibited a low tendency to avoid strangers."), and may provide only partial ordering (human raters may state consistently that $A > B$ and $B > C$ but may not agree that $A > C$).

These considerations show that a seemingly simple aspect of the language, such as the values on a scale, raise complex questions in view of both scientific validity and general usability.

4.2 Usage issues

Target Audience The participants in the group represent a broad cross-section of users, and as such are in themselves a reasonable reflection of the potential user base for an EmotionML. These user groups come under two general categories, industry and academia, and their potential products, services and research can broadly be defined as emotion annotation, emotion recognition and emotion generation, as we have previously defined in our representative use cases [1]. We can further summarise the use cases into two groups, data storage (annotation) and data transmission (recognition and generation).

This is a wide spectrum of potential uses and it is a challenge for the group to define a standard that is flexible enough to be suitable for the requirements of each user group and functional task. While academia and industry can and should share standards, users who require a data storage standard and those who

require a data transmission standard may have quite different requirements. Furthermore, any emerging standard should respect and support existing processes concerning data extraction, storage and manipulation; if it does not, then commercial adoption will be minimal and as a result the standard may be considered irrelevant.

Future Proofing Working from a well defined set of use cases has allowed us to better understand the requirements of potential users and form constraints that define how far we expect the standard to stretch. However, the standard should also be flexible enough to accommodate future uses that we have not currently envisioned.

To this end, one of our guiding principles has been to define only a core standard but allow and encourage the core to be extended by the use of custom vocabularies, notably for the emotion description itself. We aim to produce, within the core specification, a set of default values for categories, dimensions, appraisals, etc., appropriate to the transmission and storage of emotion data that will serve many, perhaps non specialist, users needs. These can, however, be replaced or "overloaded" by the user's own defined set.

This extensibility should give the standard a longer life span and avoid rapid obsolescence.

Interoperability Another centrally important objective of the standard is to enable or facilitate interoperability between systems that process and/or store emotional data. In this respect, EmotionML faces a number of challenges. Its main challenge is that within the area of emotion research there are currently no agreed and accepted standards for concepts such as categories, labels, dimensions and scales. To enable interoperability it will be necessary for us to define a mechanism that can either translate between different formats or enforce a set of core data types within the standard and then allow users to translate from this to their own preferred formats. In this way any transmission of data is guaranteed to be in a standard format and so the recipient need not worry or know about the sender's internally used format.

4.3 Towards a customisable standard?

The above discussion has highlighted the diversity of potential uses of the Emotion Markup Language in different contexts. To accomodate these, the language will have to provide sufficient flexibility: to describe emotions using one or several descriptors (categories, dimensions, appraisals and/or action tendencies); to use a pre-defined or a custom vocabulary for a descriptor (e.g., a custom list of emotion words or of appraisal dimensions); discussion is ongoing whether to provide users with the choice among one of several ways to indicate scale values – continuous bounded or unbounded, pre-defined or custom ordered lists, etc. At the same time, the discussion has pointed out the need to use standard representations for exchanging data between technological components.

The task of the group is to weigh the need for standardisation, required for interoperability, against the need for customisation, required for representing the concepts with which potential users are already working nowadays. The guiding principle used is to provide a choice only where it is truly required; to propose default sets for every choice; and to introduce a mechanism for mapping between representations where possible.

5 Conclusion

The paper has presented the mandatory and the optional elements of an emotion markup language, and has discussed a number of key aspects to take into account when designing the language.

The next step will now be the formulation of a concrete specification draft, and to collect feedback from potential users, in order to ensure that the language can actually be used as intended: as the most carefully designed and most generally usable emotion markup language to date.

Acknowledgements

Preparation of this paper was supported by the W3C and has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 211486 (SEMAINE).

References

1. Schröder, M., Devillers, L., Karpouzis, K., Martin, J.C., 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)
2. Schröder, M., Zovato, E., Pirker, H., Peter, C., Burkhardt, F.: W3C Emotion Incubator Group final report (2007)
3. Burkhardt, F., Schröder, M.: Emotion markup language: Requirements with priorities. W3C Incubator Group report (2008)
4. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Communication Special Issue on Speech and Emotion **40** (2003) 5–32
5. Scherer, K.R.: Psychological models of emotion. In Borod, J.C., ed.: The Neuropsychology of Emotion. Oxford University Press, New York (2000) 137–162
6. Gross, J.J.: Emotion regulation in adulthood: Timing is everything. Current Directions in Psychological Science **10** (2001) 214–219
7. Frijda, N.H.: The Emotions. Cambridge University Press, Cambridge, UK (1986)
8. Ekman, P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Times Books (2003)