# StimSeqOnt: An ontology for formal description of multimedia stimuli sequences

Marko Horvat*
* University of Zagreb, Faculty of Electrical Engineering and Computing
Department of Applied Computing
Zagreb, Croatia
marko.horvat3@fer.hr

***Abstract* - Sequences of multimedia documents are successfully used in laboratory settings and in practice to deliberately elicit specific emotional reactions. To ensure a successful experiment the emotion provoking stimuli must be selected carefully and have a specific order in which they are presented to the participants. Temporal aspect – duration of individual stimuli within sequences, duration of whole sequences and pauses between stimuli and sequences – must also be chosen with great care. Construction of effective sequences is a delicate and time consuming activity which requires significant group manual effort from domain experts. To facilitate this task we propose a new ontology called StimSeqOnt for formal description of stimuli sequences. The ontology is written in OWL DL language and provides formal and sufficiently expressive representation of affective concepts, high-level semantics, stimuli documents, multimedia formats and repositories used. In StimSeqOnt all relevant metadata about stimuli sequences may be stored as formal concepts. If available, elicited physiological data of previously exposed participants are available for comparison thereby enabling prediction of emotional responses. The StimSeqOnt is designed in compliance with ontology guidelines to facilitate sharing and reuse of expert knowledge.**

***Keywords - ontology; multimedia; stimulus; emotion; affective computing; psychophysiology***

## I. Introduction

Cognitive science is a compound term for several fields of study which are concerned with understanding the human mind and its processes [1]. As an interdisciplinary field cognitive science brings together philosophy, anthropology, psychology, linguistics, neuroscience, and artificial intelligence [2]. This research is concerned with attaining a comprehensive understanding of intelligence and behaviour, which include mental faculties such as language, perception, memory, attention, reasoning, and emotion [3]. Apart from artificial intelligence, within computer science another field of study is especially close to the goals of cognitive science, and that is affective computing.

Affective computing is a complex interdisciplinary field in computer science. It was defined as a distinct study back in 1997 by Rosalind Picard [4]. Primary goals of researchers in this field are understanding, description, transfer and generation of emotions in computer systems. As such it is closely related to and draws from several subareas in computer sciences like machine learning, signal processing, speech processing, information retrieval, computer vision, knowledge representation and automated reasoning, to name just the most important ones [5].

Elicitation and measurement of emotional reactions is an important area of scientific interest which is aligned both with cognitive science and affective computing. Apart from gaining a deeper insight into how emotions are created, modulated, how they influence perception, attention, thinking, memory or behaviour, such scientific investigations are typically undertaken to study psychological and neurological mechanisms explaining a wide range of behavioural and cognitive phenomena.

It has been experimentally demonstrated that visual stimuli produce measurable physiological responses in skin conductance, startle reflex, breathing and heart rate [6][7][8]. In this regard, other physiological responses have also been intensively explored, as well as multimedia formats for provoking reactions. Per example [9][10][11]. The relationship between multimedia content, emotion, physiology and neuroanatomical correlates is a very complex and intensively researched subject [12].

In any case, for the emotion elicitation using multimedia exposure to be successful, the sequence of the presented content has to have optimal order, duration, pauses, semantic meaning, as well as specific visual and auditory characteristics. The documents in stimuli sequences must be prepared well before an emotion experiment starts and annotated according to semantic and affective models, in order to accurately predict their impact on participants. Because of their use, such multimedia documents are called stimuli. The construction of stimuli sequences, for any particular task such as diagnostics or treatment of mental disorders, is always a complicated task that demands a great deal of expert work [13].

This paper proposes an ontology-based approach for a formal and comprehensive description of affective multimedia sequences for provoking emotions. Such ontology helps to reuse and share expert knowledge, thus contributing to standardization and interoperability in the field of emotion elicitation.

The remainder of the paper is organized as follows; related work is described in the next section. The model of the developed ontology is introduced in the Section 3. This includes core concepts and their relationships. The following section describes in detail the involved ontological concepts, class associations and datatypes. The intended usage of the ontology in a real emotion elicitation experiment is presented in Section 5. Finally, Section 6 discusses various benefits of the proposed ontology and provides insight into future work plans.

## II. Related Work

In recent years, researchers in affective computing, psychology and neuroscience have shown an ever-increasing interest in computer tools for detection and identification of emotions. A large amount of literature exists about each of these topics. Perhaps the best recent insights into this topic come from [14][15].

So far attempts to computationally model construction of emotion have been mostly applied to the appraisal process and cognitive models, including their relationship with stimulation of emotional reactions and the wider phenomena of emotion, i.e. covering all aspects of emotional and affective mental functioning. Per example [16][17][18]. Here the appraisal is defined as the continuous, recursive subjective evaluation of events for their pertinence, as well as the coping potential of the individual. The outcome of the appraisals from these different criteria is predicted to directly drive response patterning of physiological reactions, motor expression, and action preparation [19].

Previously it has been established that ontologies should be utilized to describe higher-semantics and affect in multimedia stimuli [20]. The benefits of formalizing knowledge descriptions are significant and outweigh foreseeable shortfalls [21].

Several miscellaneous ontologies have already been developed especially for the description of emotion. Representative examples of such ontological models are the Emotion Ontology (EMO) [22] and other diverse models for describing and reasoning on emotion context in order to improve emotion detection based on bodily expression [23], describing emotions and their detection [24], and tag-based music recommendation [25].

Finally, the most comprehensive effort into the systematization of emotional models for the usage in computer data processing systems represents the EmotionML standard recommendation developed under the umbrella of W3C Consortium [26]. Although not an ontology, EmotionML is written in XML and serves as an annotation language for multimedia. It currently offers the most sophisticated emotion glossary which includes 5 category vocabularies, 4 dimension vocabularies, 3 appraisal vocabularies and 1 action tendency vocabulary.

## III. The StimSeqOnt Model

The StimSeqOnt ("Stimuli Sequence Ontology") is designed to − as stated previously − provide a formal and comprehensive, yet simple and manageable, description of sequences containing emotion-provoking multimedia documents content. Such sequences are designed to intentionally provoke targeted emotional reactions in exposed subjects. The ontology model is shown in Figure 1.

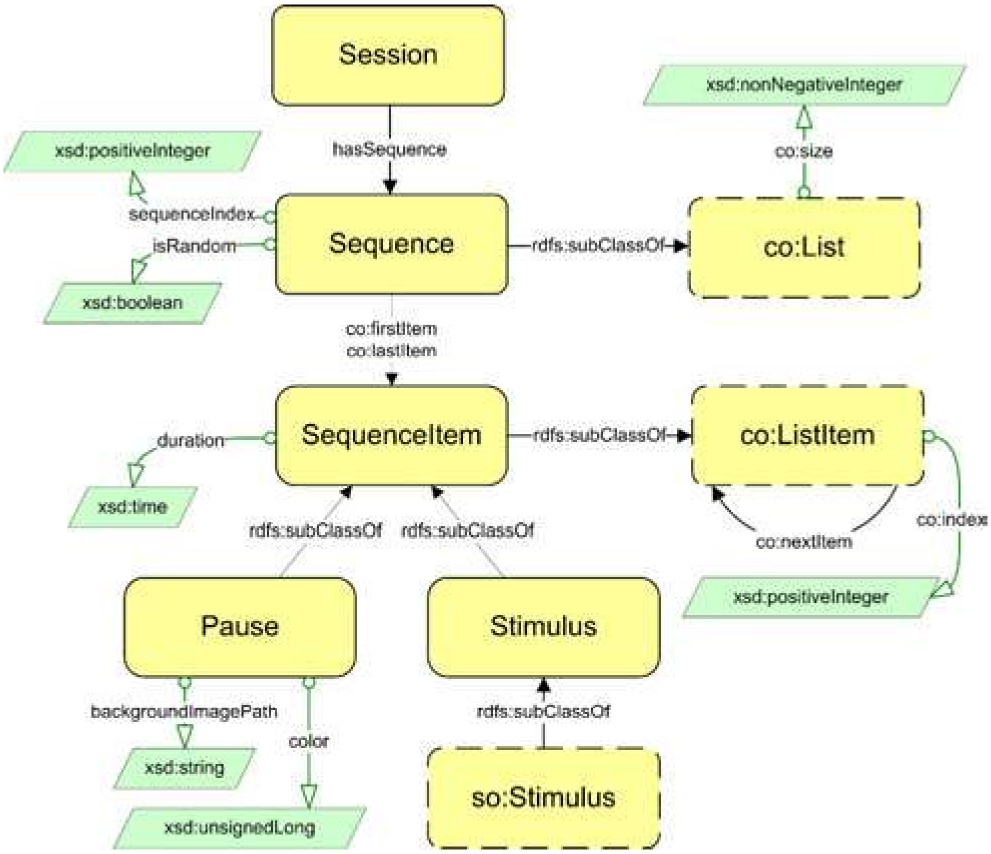

Figure 1. The concepts and relations in the StimSeqOnt model. Data properties (owl:DataProperty) are indicated with green lines while class inheritance (rdfs:subClassOf) and object properties (owl:ObjectProperty) are denoted with black lines. External ontology concepts have dashed borders.

The StimSeqOnt enables a common understanding about the perceived meaning of multimedia stimuli which can be easily shared between different researchers and computer systems. The StimSeqOnt ontology is designed to facilitate querying and retrieval of audio-visual stimuli and stimuli sequences that contain them. The ontology is written in OWL DL and, consequently, is decidable which provides for its practical use. When used with appropriate software tools it may be used to alleviate the construction of stimuli sequences.

In accordance with the most important ontology designing objectives the StimSeqOnt reuses two existing ontologies. They are utilized for formal representation of collections of emotionally annotated multimedia documents: *i*) STIMONT and *ii*) Collections Ontology. Both ontologies have necessary and sufficient expressiveness and decidability and can be used in inferences and SPARQL queries. The STIMONT (prefix "stim") [27] is an upper core ontology designed to provide an integrated and formal description of emotion, high-level semantics, context and physiology content of a multimedia stimulus. The most important feature of the STIMONT is that it provides a formal framework for supporting the explicit, human and machine-processable definition of affective multimedia content. This ontology also facilitates the storage of stimuli in emotionally-annotated databases, stimuli querying and retrieval and construction of stimuli sequences. Its model relies on W3C EmotionML format and extends it with additional emotion vocabularies. High-level semantic expressivity is made possible by reuse of SUMO common sense ontology and SUMO to WordNet mappings. The STIMONT is written in OWL DL.

The Collections Ontology (prefix "co") [28] is an OWL 2 DL ontology developed for creating sets, bags and lists of resources, and for inferring collection properties even in the presence of incomplete information. Although RDF data can be used to define collections and containers to group resources as one entity, this important feature has not been included in OWL and even in OWL 2 DL specifications. The Collections Ontology has been created to address this issue.

The StimSeqOnt is a formalism for the representation of all relevant knowledge about any type of multimedia such as images, sounds, video and text in stimuli sequences. This is achieved by reusing OWL constructs from the STIMONT model that helps to convey high-level semantics interpretation, and even induced emotional states, document metadata and emotion-related physiology.

## IV. StimSeqOnt Concepts

The top concept in the StimSeqOnt ontology (prefix "sso") is sso:Session which is directly subsumed by owl:Thing class. An instance of sso:Session is related to one or more instances of sso:Sequence class with hasSequence object relationship:

$$Session \equiv \exists_{\geq 1} Sequence$$

$$hasSequence.Session \sqsubseteq Sequence$$

Sessions can only to be labeled because other important information is stored in other classes and their mutual relationships. The model was intentionally made simple and compact as to be easier to use by experts and in software tools.

In the StimSeqOnt model one session is divided into one or more sequences which contain particular stimuli. Sequences have their index and can be numbered. Each sequence must contain at least one stimulus, and every session consists of at least one sequence. Item in the sequence can either be a pause or stimulus. Pauses can be introduced between adjacent stimuli.

Stimuli are hierarchically organized into sequences. Sequence is a type of List from the Collections Ontology [28]. The co:List cannot be empty, i.e. data property co:index of the class co:ListItem is a positive integer. Therefore, at least one instance of sso:SequenceItem class must exist in ABox.

Each multimedia stimulus or pause is represented in the ABox as exactly one instance of sso:SequenceItem concept. Concept sso:pause can have a background image or a color, and sso:Stimulus is specified by so:Stimulus from STIMONT which provides for a rich description of emotion, high-level semantics, context and physiology content [27].

The order of stimuli in the sequence is defined as in a linked list. In this scheme the first sequence member in the list is attached to a sso:Sequence individual using co:firstItem object relation. The second member in the list is linked to the first with co:nextItem object relation and so on, until, eventually, the last member is denoted using co:lastItem object relation Thus the sequence can be easily traversed by following all sso:Stimulus individuals in the chained list. Duration of the exposure in seconds is captured by sso:SequenceItem duration property. Also, each list item has its index that uniquely identifies it.

## V. Formal Modelling of Stimuli Sequences Example

As a realistic example how the StimSeqOnt ontology is supposed to be used, it will be demonstrated how to encode emotionally annotated multimedia stimuli sequences from a previously published experiment [29]. This study was performed at University of Zagreb, Faculty of Electrical Engineering and Computing in cooperation with experts from Department of Psychology, Faculty of Humanities and Social Sciences. In total *N*=10 participants where exposed to 20 optimal pictures and 20 sounds manually selected by experts for inclusion in the elicitation sequences. The aim of the experiment was to objectively compare emotion elicitation strength between two sequences consisting only of static images and only of videos, respectively. Video clips were constructed using congruent affective pictures and sounds. The required files were selected from the International Affective Picture System (IAPS) [30] and the International Affective Digital Sounds System (IADS) [31] affective multimedia databases which were designed specifically for this type of experiments [32]. A congruent picture from IAPS and a sound from IADS were combined to make one video-clip. Identical pictures were used in both sequences.

A sample of ten IAPS pictures used in the experiment are displayed in Figure 3.

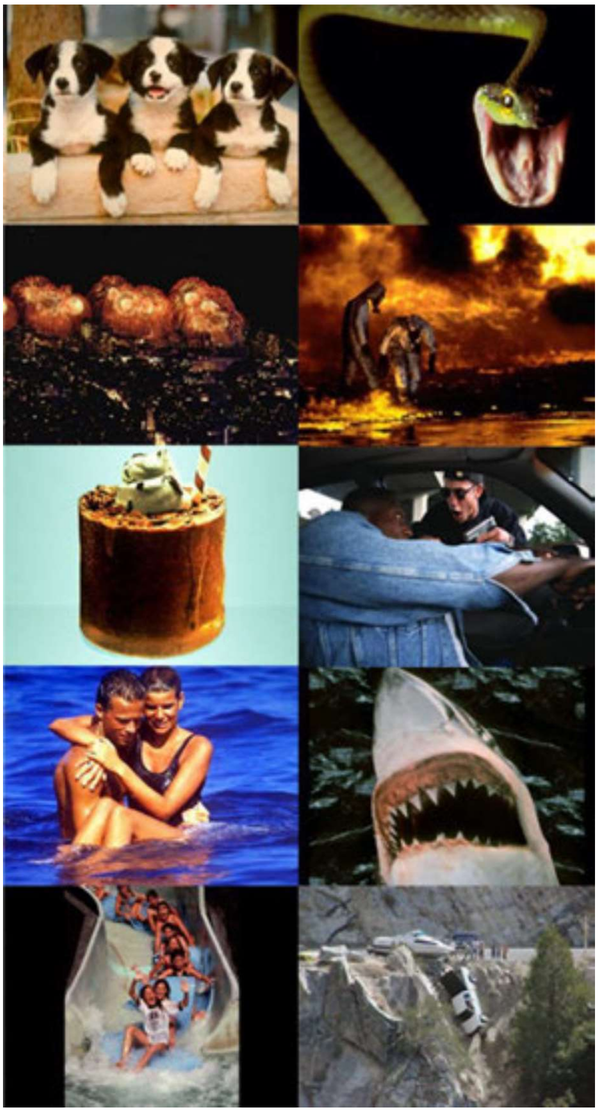

Figure 2. A sample of ten IAPS pictures used as emotion eliciting video clips and images. Happiness (left) and fear dominant emotion stimuli (right column).

The dominant emotions provoked in the experiment purposely had opposite polarity: happiness and fear. The selected stimuli were considered the most likely to induce measurable emotional responses in the participants' population.

Each participant was exposed in two separate sessions or series. Each session consisted of one happiness sequence, one fear inducing sequence, and also of one neutral sequence. A single sequence consisted of either 10 pictures or 10 video-clips (Fig. 3). The length of each stimulus was exactly 15 seconds after which the participant was shown a blank neutral screen and had to write down his affective judgments in a self-assessment questionnaire (SAQ). The neutral blank screen only showed teal color which has an optimal ratio of stimulating positive and negative emotions [29]. When the participant was finished he could resume the sequence by himself (i.e. with a mouse click).

Immediately before the start of the experiment each participant was separately introduced to the stimulation protocol with a neutral sequence. The neutral sequence consisted of one low arousal and valence picture and one video-clip without dominant emotions.

The experimental protocol demanded that half of the participants were first exposed to happiness sequences, and then fear sequences, and the other half of the participants the opposite: they first watched still images and then videos. To prevent the unwanted drift of physiological signals (cardiac and respiratory) before nonneutral sequences participants were exposed to a neutral stimulus until their baseline response was established.

The participants' emotional responses were recorded by two methods: 1) self-assessment responses in pauses between stimuli i 2) real-time monitoring of physiological signals during exposure to stimuli. After each exposure to a stimulus participants filled out a self-assessment questionnaire. The physiological signals – skin conductance, electrocardiogram (ECG), respiration and skin temperature – were monitored using the AcqKnowledge software connected with BIOPAC MP150 physiology acquisition system. The whole setup was synchronized with SuperLab tool for presentation of stimuli to the participants. The tool uses a propriety format to store information about the presentation sequence.

An unspecified long rest period followed each exposure session during which participants relaxed. This was verified by examining the physiological signal parameters that were visualized in real time. The exposure could resume only after the baseline signal levels were reestablished.

The StimSeqOnt individuals stored in ABox for modelling of the described presentation sequence are depicted in the Figure 3. The sequence is designed for stimulation of happiness.

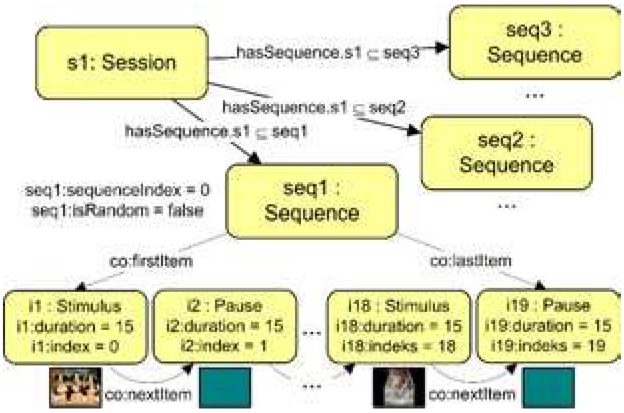

Figure 3. An illustration of real multimedia presentation sequences modelled in StimSeqOnt for the emotion provoking experiment in [32]. Significant individuals stored in ABox are shown.

In the figure above seq1 is the individual representing the happiness provoking sequence, seq2 fear provoking and seq3 the neutral stimuli list. All sequences are related to the single session individual s1. Order of items in the sequences is not randomized. Individuals seq1 and seq2 each include 20 items: 10 emotion provoking multimedia documents and 10 pauses with emotionally neutral images (blank screen in teal color). The first stimulus in seq1 can be fetched using seq1.firstItem data relationship, and the last with seq2.lastItem or, alternatively, by following co:nextItem relationship starting from the first item in the list. All sequences and items have zero-based indices.

## VI. Conlusion and Future Work

The StimSeqOnt ontology presented in the article offers a number of advantages over other possible formats for knowledge representation of emotionally annotated multimedia sequences. We demonstrated how the proposed ontology model can be applied to capture all metadata relevant for construction of a stimuli sequence to be used in an emotion elicitation experiment. This formal approach should facilitate an increase in efficiency of multimedia information retrieval. Although ontologies require more complicated prerequisites, such as reasoning engines, than markup languages or keyword-based annotations, they are more advantageous than the simpler methods. The StimSeqOnt's properties enable a formal, consistent, systematic and expressive model of the affective multimedia sequences. The ontology enables DL-based reasoning about the aggregated content and document metadata.

The StimSeqOnt is specifically designed to interlink with the previously developed STIMONT and Collection Ontology. These models together enable formal representation of high-level semantics, emotion and related states, document context and stimulated physiology that collectively define multimedia stimuli sequences. Both ontologies facilitate knowledge reuse, interoperability and formalization of stimuli information that are superior to the contemporary methods for representation of affective multimedia documents.

In the future, we would like to develop a more expressive upper ontology for the complete description of emotion elicitation experiments. Such even more expressive model would incorporate STIMONT and StimSeqOnt. Additionally, the ontology would add new vocabularies for formal representation about experimentation protocols, employed instrumentation and their settings, recorded and processed psychophysiological signals. This ontology would also need to enable automated reasoning about experimentation results using captured categorical and numerical time-series data. We hope that the presented StimSeqOnt is a step toward this larger goal.

## References

[1] A. Collins and D. G. Bobrow, Representation and understanding: Studies in cognitive science. Elsevier, 2017.

[2] G. A. Miller, The cognitive revolution: a historical perspective. Trends in Cognitive Sciences, 2003.

[3] P. Thagard, "Mind: Introduction to cognitive science," Thompson, Cambridge, MA: MIT press, vol. 17, pp. 811−834, 2005.

[4] R. W. Picard, Affective computing. MIT press, 2000.

[5] S. Brave and C. Nass, "Emotion in human-computer interaction," In The human-computer interaction handbook, CRC Press, pp. 103−118, 2007.

[6] M. M. Bradley, M. Codispoti, B. N. Cuthbert, and P. J. Lang, "Emotion and motivation I: defensive and appetitive reactions in picture processing," Emotion, vol. 1(3), pp. 276, 2001.

[7] M. M. Bradley, M. Codispoti, D. Sabatinelli, and P. J. Lang, "Emotion and motivation II: sex differences in picture processing," Emotion, vol. 1(3), pp. 300, 2001.

[8] D. Kukolja, S. Popović, M. Horvat, B. Kovač, and K. Ćosić, "Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications, "International journal of human-computer studies, vol. 72(10), pp. 717−727, 2014.

[9] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," IEEE transactions on affective computing, vol. 3(2), pp. 211−223, 2012.

[10] C. G. Courtney, M. E. Dawson, A. M. Schell, A. Iyer, and T. D. Parsons, "Better than the real thing: Eliciting fear with moving and static computer-generated stimuli," International Journal of Psychophysiology, vol. 78(2), pp. 107−114, 2010.

[11] J. Rottenberg, R. D. Ray, and J. J. Gross, "Emotion elicitation using films," Handbook of emotion elicitation and assessment, Oxford University Press, New York, pp. 9−28, 2007.

[12] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," Biological psychology, vol. 84(3), pp. 394−421, 2010.

[13] K. Ćosić, S. Popović, M. Horvat, D. Kukolja, B. Dropuljić, B. Kovač, and I. Fabek, "Multimodal paradigm for mental readiness training and PTSD prevention," In New tools to enhance posttraumatic stress disorder diagnosis and treatment: invisible wounds of war. IOS Press, 2013.

[14] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, "How emotion is made and measured," International Journal of Human-Computer Studies, vol. 65(4), pp. 275−291, 2007.

[15] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J., Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," In Proceedings of the 19th ACM international conference on multimodal interaction, pp. 524−528, 2017.

[16] G. L. Clore and A. Ortony, "Psychological construction in the OCC model of emotion," Emotion Review, vol. 5(4), pp. 335−343, 2013.

[17] P. C. Ellsworth and K. R. Scherer, "Appraisal processes in emotion," Handbook of affective sciences, vol. 572, V595, 2003.

[18] S. Marsella and J. Gratch, EMA: A computational model of appraisal dynamics, 2006.

[19] D. Grandjean, D. Sander, and K. R. Scherer, "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization," Consciousness and cognition, vol. 17(2), pp. 484−495, 2008.

[20] M. Horvat, S. Popović, N. Bogunović, K. Ćosić, "Tagging multimedia stimuli with ontologies," Proceedings of the 32nd International Convention MIPRO 2009, Croatian Society for Information and Communication Technology, Electronics and Microelectronics – MIPRO, Opatija, Croatia, pp. 203–208, 2009.

[21] M. Horvat, Generation of multimedia stimuli based on ontological, affective and semantic annotation (Doctoral thesis), University of Zagreb, 2013.

[22] J. Hastings, W. Ceusters, B. Smith, and K. Mulligan, "Dispositions and processes in the Emotion Ontology," 2011.

[23] F. Berthelon and P. Sander, "Emotion ontology for context awareness," In 2013 IEEE 4th International Conference on Cognitive Infocommunications, IEEE, pp. 59−64, 2013.

[24] J. M. López, R. Gil, R. García, I. Cearreta, and N. Garay, "Towards an ontology for describing emotions," In World Summit on Knowledge Society, Springer, Berlin, Heidelberg, pp. 96−104, 2008.

[25] H. H. Kim, "A semantically enhanced tag-based music recommendation using emotion ontology," In Asian Conference on Intelligent Information and Database Systems. Springer, Berlin, Heidelberg, pp. 119−128, March 2013.

[26] F. Burkhardt, C. Pelachaud, B. W. Schuller, and E. Zovato, "EmotionML," In Multimodal interaction with W3C standards. Springer, Cham, pp. 65−80, 2017.

[27] M. Horvat, N. Bogunović, and K. Ćosić, "STIMONT: a core ontology for multimedia stimuli description," Multimedia tools and applications, vol. 73(3), pp. 1103−1127, 2014.

[28] P. Ciccarese and S. Peroni, "The Collections Ontology: creating and handling collections in OWL 2 DL frameworks," Semantic Web, vol. 5(6), pp. 515−529, 2014.

[29] M. Horvat, D. Kukolja, and D. Ivanec, "Comparing affective responses to standardized pictures and videos: A study report," In

MIPRO, 2015 Proceedings of the 38th International Convention, IEEE, pp. 1394–1398, May 2015.

[30] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, “International affective picture system (IAPS): Affective ratings of pictures and instruction manual,” Technical Report A–8, University of Florida, Gainesville, FL, 2008.

[31] P. J. Lang and M. M. Bradley, “The International Affective Digitized Sounds (2nd Edition; IADS-2): affective ratings of sounds and instruction manual,” Technical report B-3, University of Florida, Gainesville, FL, 2007.

[32] M. Horvat, “A Brief Overview of Affective Multimedia Databases,” In Central European Conference on Information and Intelligent Systems (CECIIS 2017), pp. 3–11, 2017.