SceneMaker: Creative Technology for Digital StoryTelling

Murat Akser[1], Brian Bridges[1], Giuliano Campo[1], Abbas Cheddad[6], Kevin Curran[2], Lisa Fitzpatrick[1], Linley Hamilton[1], John Harding[1], Ted Leath[4], Tom Lunney[2], Frank Lyons[1], Minhua Ma[5], John Macrae[3], Tom Maguire[1], Eileen McClory[1], Victoria McCollum[1], Paul Mc Kevitt[1], Adam Melvin[1], Paul Moore[1], Eleanor Mulholland[1], Karla Muñoz[7], Greg O'Hanlon[1], and Laurence Roman[1]

[1] Ulster University, Magee,
School of Creative Arts & Technologies,
BT48 7JL, Derry/Londonderry, Northern Ireland
p.mckevitt@ulster.ac.uk,
WWW home page: http://www.paulmckevitt.com
[2] Ulster University, Magee,
School of Computing & Intelligent Systems,
BT48 7JL, Derry/Londonderry, Northern Ireland
[3] Ulster University, Jordanstown,
Research & Innovation,
BT37 OQB, Newtownabbey, Northern Ireland
[4] Ulster University, Coleraine,
Information Services Directorate,
BT52 1SA, Coleraine, Northern Ireland
[5] University of Huddersfield,
Department of Art & Communication,
Queensgate, Huddersfield HD1 3DH, England
[6] Blekinge Institute of Technology (BTH),
Department of Computer Science & Engineering,
SE-371 79, Karlskrona, Sweden
[7] BijouTech, CoLab, Letterkenny,
F92 H292, Co. Donegal, Ireland

Abstract. The School of Creative Arts & Technologies at Ulster University (Magee) has brought together the subject of computing with creative technologies, cinematic arts (film), drama, dance, music and design in terms of research and education. We propose here the development of a flagship computer software platform, *SceneMaker*, acting as a digital laboratory workbench for integrating and experimenting with the computer processing of new theories and methods in these multidisciplinary fields. We discuss the architecture of SceneMaker and relevant technologies for processing within its component modules. SceneMaker will enable the automated production of multimodal animated scenes from film and drama scripts or screenplays. SceneMaker will highlight affective or emotional content in digital storytelling with particular focus on character body posture, facial expressions, speech, non-speech audio, scene composition, timing, lighting, music and cinematography. Applications of SceneMaker include automated simulation of productions and education and training of actors, screenwriters and directors.

Key words: SceneMaker, natural language processing, speech processing, artificial intelligence (AI), affective computing, computer graphics, cinematography, 3D visualisation, digital storytelling, storyboards, film, drama, dance, music technology, design

1 Introduction

The School of Creative Arts & Technologies at Ulster University (Magee) has brought together the subject of computing with creative technologies, cinematic arts (film), drama, dance, music and design in terms of research and education. This paper proposes the development of a flagship computer software platform, *SceneMaker*, acting as a digital laboratory workbench for integrating and experimenting with the computer processing of new theories and methods in these multidisciplinary fields. The production of plays or movies is an expensive process involving planning, rehearsal time, actors and technical equipment for lighting, sound and special effects. It is also a creative process which requires experimentation, visualisation and communication of ideas between everyone involved, e.g., playwrights, directors, actors, cameramen, orchestra, managers and costume and set designers. *SceneMaker* will assist in this production process and provide a facility to test and visualise scenes before finally implementing them. Users will input a natural language text scene and automatically receive output multimodal 3D visualisations. The objective is to give directors or animators a draft idea of what a scene will look like. Users will have the ability to refine the automatically created output through a script and 3D editing interface, also accessible over the internet and on mobile devices. Thus, SceneMaker will be a collaborative tool for script writers, animators, directors and actors, sharing scenes online. Such technology could be applied in the training of those involved in scene production without having to utilise expensive actors and studios.

This work focuses on three research questions: How can affective and emotional information be computationally recognised in screenplays and structured for visualisation purposes? How can emotional states be synchronised in presenting all relevant modalities? Can compelling, life-like and believable multimodal animations be achieved? Section 2 of this paper gives an overview of current research on computational, multimodal and affective scene production. In section 3, the design and architecture of SceneMaker is discussed. SceneMaker is compared to related work in section 4 and section 5 concludes.

2 Background and Literature Review

Automatic and intelligent production of film/theatre scenes with characters expressing emotional or affective states involves four development stages and this section reviews current advances in these areas which provide a sound basis for SceneMaker.

2.1 Detecting Emotions and Personality in Film/Play Scripts

All modalities of human interaction express emotional states and personality, such as voice, word choice, gesture, body posture and facial expression. In order to recognise emotions in script text and to create life-like characters, psychological theories for emotion, mood, personality and social status are translated into computable methods, e.g., Ekman's 6 basic emotions [4], the Pleasure-Arousal-Dominance (PAD) model [5] with intensity values or the OCC model (Ortony-Clore-Collins) [6] with cognitive grounding and appraisal rules. Word choice is a useful indicator of the personality of a story character, their social situation, emotional state and attitude. Different approaches to textual affect sensing are able to recognise explicit affect words such as keyword spotting and lexical affinity [7], machine learning methods [8], hand-crafted rules and fuzzy logic systems [9], and statistical models [8]. Common knowledge-based approaches [10], [11] and a cognitive inspired model [12] include emotional context evaluation of non-affective words and concepts.

2.2 Modelling Affective Embodied Characters

Research aimed at automatic modelling and animating virtual humans with natural expressions faces challenges not only in automatic 3D character manipulation/transformation, synchronisation of face expressions, e.g., lips and gestures with speech, path finding and collision detection, but furthermore in the execution of actions. SCREAM (Scripting Emotion-based Agent Minds) [13] is a web-based scripting tool for multiple characters which computes affective states based on the OCC-Model [6] of appraisal and intensity of emotions, as well as social context. ALMA [14] (A Layered Model of Affect) implements AffectML, an XML-based modelling language which incorporates the concept of short-term emotions, medium-term moods and long-term personality profiles. The OCEAN personality model [15], Ekman's basic emotions [4] and a model of story character roles are combined through a fuzzy rule-based system [9] to decode the meaning of scene descriptions and to control the affective state and body language of the characters. The high-level control of affective characters in [9] maps personality and emotion output to graphics and animations. Embodied Conversational Agents (ECA) are capable of real-time face-to-face conversations with human users or other agents, generating and understanding natural language and body movement. Greta [16] is a real-time 3D ECA with a 3D model of a woman compliant with the MPEG-4 animation standard. Two standard XML languages, FML-APML (Function Markup Language, Affective Presentation Markup Language) for communicative intentions and BML (Behavior Markup Language) for behaviours enable users to define Greta's communicative intentions and behaviours.

2.3 Visualisation of 3D Scenes and Virtual Theatre

Visual and auditory elements involved in composing a virtual story scene, the construction of the 3D environment or set, scene composition, automated cinematography and the effect of genre styles are addressed in complete text-to-visual systems such as WordsEye [17], ScriptViz [19], CONFUCIUS [1] and NewsViz [20], and the scene directing system, CAMEO [22]. WordsEye depicts non-animated 3D scenes with characters, objects, actions and environments. A database of graphical objects holds 3D models, their attributes, poses, kinematics and spatial relations in low-level specifications. ScriptViz renders 3D scenes from natural language screenplays immediately during the writing process, extracting verbs and adverbs to interpret events and states in sentences. The time and environment where a story takes place, the theme the story revolves around and the emotional tone of films, plays or literature classify different genres with distinguishable presentation styles. Genre is reflected in the detail of a production, exaggeration and fluency of movements, pace (shot length), lighting, colour and camerawork. Cinematic principles in different genres are investigated in [21]. Dramas and romantic movies are slower paced with longer dialogues, whereas action movies have rapidly changing, shorter shot length. Comedies tend to be presented in a large spectrum of bright colours, whereas horror films adopt mostly darker hues.

CONFUCIUS [1] produces multimodal 3D animations of single sentences. 3D models perform actions, dialogues are synthesised and basic cinematic principles determine camera placement. NewsViz [20] gives numerical emotion ratings to words calculating the emotional impact of words and paragraphs, facilitating the display mood of the author over the course of online football reports and tracks the emotions and moods of the author, aiding reader understanding.

A high-level synchronised Expression Mark-up Language (EML) [18] integrates environmental expressions like cinematography, illumination and music as a new

modality into the emotion synthesis of virtual humans. The automatic 3D animation production system, CAMEO, incorporates direction knowledge, like genre and cinematography, as computer algorithms and data to control camera, light, audio and character motions. A system which automatically recommends music based on emotion is proposed in [23]. Associations between emotions and music features in movies are discovered by extracting chords, rhythm and tempo of songs.

2.4 Multimodal Interfaces and Mobile Applications

Technological advances enable multimodal human-computer interaction in the mobile world. System architectures and rendering can be placed on the mobile device itself or distributed from a server via wireless broadband networks. SmartKom Mobile [24] deploys the multimodal system, SmartKom, on mobile devices. The user interacts with a virtual character (Smartakus) through dialogue. Supported modalities include language, gesture, facial expression and emotions through speech emphasis. Script writing tools assist the writing process of screenplays or play scripts, like Final Draft [25] for mobile devices.

3 Design and Architecture of SceneMaker

The software prototype platform, SceneMaker will use Natural Language Processing (NLP) methods applied to screenplays to automatically extract and visualise emotions, moods and film/play genre. SceneMaker will augment short 3D scenes with affective influences on the body language of characters and environmental expression, like illumination, timing, camera work, music and sound automatically directed in respect of the genre style.

3.1 Architecture of SceneMaker

SceneMaker's architecture is shown in Fig. 1. The key component is the *scene production module* including modules for understanding, reasoning and multimodal visualisation situated on a server. The *understanding module* performs natural language processing, sentiment analysis and text layout analysis of input text utilising algorithms and software from CONFUCIUS [1], NewsViz [20] and 360-MAM-Affect [3]. The *reasoning module* interprets the context based on common, affective and cinematic knowledge bases, updates emotional states and creates plans for actions, their manners and representation of the set environment with algorithms and software from Control-Value Theory emotion models [2] and CONFUCIUS [1]. StoryTelling techniques developed in [26] will also be employed here and gender and affect in performance [27] will also be important. The *visualisation module* maps these plans to 3D animation data, selects appropriate 3D models from the graphics database, defines their body motion transitions, instructs speech synthesis, selects non-speech sound and music files from the audio database and assigns values to camera and lighting parameters. The visualisation module synchronises all modalities into an animation manuscript. Techniques for automated multimodal presentation [28], embodied cognition music [29], visual and audio synchronisation [30], and analysing sound in film [31] will be employed here. The online *user interface*, available via desktop computers and mobile devices, consists of two parts. The input module provides assistance for film and play script writing and editing and the output module renders the 3D scene according to the manuscript and allows manual scene editing to fine-tune the automatically created animations.

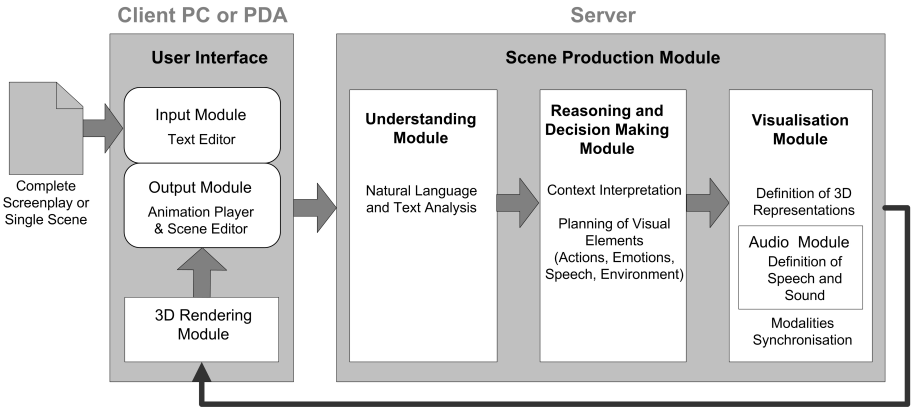

Fig. 1. Architecture of SceneMaker

3.2 Implementation of SceneMaker

Multimodal systems automatically mapping text to visual scenes face challenges in interpreting human natural language which is usually variable, ambiguous, imprecise and relies on shared knowledge between the communicators. Enabling a machine to understand a natural language text involves feeding the machine with grammatical structures, semantic relations and visual descriptions to be able to match suitable graphics. Existing software tools fulfilling sub-tasks will be modified, combined and extended for the implementation of SceneMaker. For the interpretation of input scripts, SceneMaker will build upon the NLP modules of CONFUCIUS [1] and 360-MAM-Affect [3] with GATE [32], but a pre-processing tool will first deconstruct the layout structure of the input screenplay/play script. The syntactic knowledge base parses input text and identifies parts of speech, e.g., noun, verb, adjective, with the Connexor Part-of-Speech Tagger [33] and determines the constituents in a sentence, e.g., subject, verb and object, with Functional Dependency Grammars [34]. The semantic knowledge base (WordNet [35] and LCS database [37]) and temporal language relations will be extended by an emotional knowledge base, e.g., WordNet-Affect [36], emotion processing with 360-MAM-Affect [3], EmoSenticNet [38] and RapidMiner [39] and Control-Value Theory emotional models [2], and context reasoning with ConceptNet [11] to enable an understanding of the deeper meaning of the context and emotions.

In order to automatically recognise genre, SceneMaker will identify keyword co-occurrences and term frequencies and determine the length of dialogues, sentences and scenes/shots. The visual knowledge of CONFUCIUS, such as object models and event models, will be related to emotional cues. CONFUCIUS' basic cinematic principles will be extended and classified into expressive and genre-specific categories. EML [18] is a comprehensive XML-based scripting language for modelling expressive modalities including body language and cinematic annotations. Resources for 3D models are H-Anim models [40] which include geometric or physical, functional and spatial properties. For speech generation from dialogue text, the speech synthesis module used in CONFUCIUS, FreeTTS [41], will be tested for its suitability in SceneMaker with regard to mobile applications and the effectiveness of emotional prosody. An automatic audio selection tool, as in [23], will be incorporated for intelligent, affective selection of sound and music in relation to the theme and mood of a scene.

Test scenarios will be developed based on screenplays of different genres and animation styles, e.g., drama films, which include precise descriptions of set layout and props versus comedy, which employs techniques of exaggeration for expression.

The effectiveness and appeal of the scenes created in SceneMaker will be evaluated against hand-animated scenes and existing feature film scenes. The functionality and usability of SceneMaker's components and the GUI will be tested in cooperation with professional film directors, comparing the process of directing a scene traditionally with real or virtual (SceneMaker) actors.

4 Relation to Other Work

Research implementing various aspects of modelling affective virtual actors, narrative systems and film-making applications relates to SceneMaker. CONFUCIUS [1] and ScriptViz [19] realise text-to-animation systems from natural language text input, but they do not enhance visualisation through affective aspects, the agent's personality, emotional cognition or genre specific styling. Their animations are built from well-formed single sentences but not extended texts or scripts. SceneMaker will facilitate animation modelling of sentences, scenes or whole scripts. Single sentences require more reasoning about default settings and more precision will be achieved from collecting context information from longer passages of text. No previous storytelling system controls agent behaviour through integrating all of personality, social status, narrative roles and emotions. EML [18] combines multimodal character animation with film making practices based on an emotional model, but it does not consider personality types or genre. CAMEO [22] relates specific cinematic direction, for character animation, lighting and camera work, to the genre or theme of a given story, but genre types are explicitly selected by the user. SceneMaker will introduce a new approach to automatically recognise genre from script text with keyword co-occurrence, term frequency and calculation of dialogue and scene length.

5 Conclusion

This paper proposes the development of a flagship computer software platform, SceneMaker, acting as a digital laboratory workbench for integrating and experimenting with the computer processing of new theories and methods in multidisciplinary fields. SceneMaker will contribute to believability and aesthetic quality of automatically produced animated multimedia scenes. SceneMaker, which automatically visualises affective expressions of screenplays, aims to advance knowledge in the areas of affective computing, digital storytelling and expressive multimodal systems.

Existing systems solve partial aspects of NLP, emotion modelling and multimodal storytelling. Thereby, we focus on semantic interpretation of screenplays or play scripts, the computational processing of emotions, virtual agents with affective behaviour and expressive scene composition including emotion-based audio selection. SceneMaker's mobile, web-based user interface will assist directors, drama students, writers and animators in the testing of their ideas. Accuracy of animation content, believability and effectiveness of expression and usability of the interface will be evaluated in empirical tests comparing manual animation, feature film scenes and real-life directing with SceneMaker. In conclusion, SceneMaker will automatically produce multimodal animations with heightened expressivity and visual quality from screenplay or play script input.

References

1. Ma, M., McKevitt, P.: Virtual Human Animation in Natural Language Visualisation. Special Issue on the 16th Artificial Intelligence and Cognitive Science Conference (AICS-05), Artificial Intelligence Review, 25 (1-2), 37-53 (2006)
2. Muñoz, K., Mc Kevitt, P., Lunney, T., Noguez, J. and Neri, L.: Designing and Evaluating Emotional Student Models for Game-based Learning. In: Ma, M., Oikonomou, A. and Jain, L.C. (Eds.), Serious Games and Edutainment Applications, 245-272, London, England: Springer. Also presented at the First International Workshop on Serious Games Development and Applications (SGDA-10), School of Computing, University of Derby, England, July 8th (2011)
3. Mulholland, E., Mc Kevitt, P., Lunney, T., Farren, J. and Wilson, J.: 360-MAM-Affect: Sentiment Analysis with the Google Prediction API and EmoSenticNet. In: Proc. of the 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN-2015), Politecnico di Torino, Turin (Torino), Italy, June 10-12, 1-5; also published in EUDL EAI Endorsed Transactions on Scalable Information Systems, 2(6): e5, Nov. (2015)
4. Ekman, P. and Rosenberg E. L.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System. Oxford University Press, England (1997)
5. Mehrabian, A.: Framework for a Comprehensive Description and Measurement of Emotional States. In: Genetic, Social, and General Psychology Monographs. Heldref Publishing, 121 (3), 339-361 (1995)
6. Ortony A., Clore G. L., and Collins A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, MA (1988)
7. Francisco, V., Hervás, R. and Gervás, P.: Two Different Approaches to Automated Mark Up of Emotions in Text. In: Research and development in intelligent systems XXIII: Proc. of AI-2006. Springer, 101-114 (2006)
8. Strapparava, C. and Mihalcea, R.: Learning to Identify Emotions in Text. In: Proc. of the 2008 ACM Symposium on Applied Computing. SAC '08. ACM, New York, NY, 1556-1560 (2008)
9. Su, W-P., Pham, B., Wardhani, A.: Personality and Emotion-Based High-Level Control of Affective Story Characters. IEEE Transactions on Visualization and Computer Graphics, 13 (2), 281-293 (2007)
10. Liu, H., Lieberman, H., and Selker, T.: A Model of Textual Affect Sensing Using Real-world Knowledge. In: Proc. of the 8th International Conference on Intelligent User Interfaces. IUI '03. ACM, New York, 125-132 (2003)
11. Liu, H. and Singh, P.: ConceptNet: A Practical Commonsense Reasoning Toolkit. In: BT Technology Journal. Springer, The Netherlands, 22(4), 211-226 (2004)
12. Shaikh, M.A.M., Prendinger, H. and Ishizuka, M.: A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text. In: Affective Information Processing. Springer, London, 45-73 (2009)
13. Prendinger, H. and Ishizuka, M.: SCREAM: Scripting Emotion-based Agent Minds. In: Proc. of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1. AAMAS '02. ACM, New York, 350-351 (2002)
14. Gebhard, P.: ALMA - Layered Model of Affect. In: Proc. of the 4th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 05). Utrecht University, The Netherlands. ACM, New York, 29-36. (2005)
15. De Raad, B.: The Big Five Personality Factors. In: The Psycholexical Approach to Personality. Hogrefe & Huber, Cambridge, MA, USA (2000)
16. Mancini, M. and Pelachaud, C.: Implementing Distinctive Behavior for Conversational Agents. In: Sales Dias, M., Gibet, S., Wanderley, M.M., Bastos, R. (Eds.), Gesture-Based Human-Computer Interaction and Simulation, Lecture Notes in Computer Science, Vol. 5085/2009, 163-174. Springer, Berlin/Heidelberg (2009)
17. Coyne, B. and Sproat, R.: WordsEye: An Automatic Text-to-Scene Conversion System. In: Proc. of the 28th Annual Conference on Computer Graphics and Interactive Techniques. ACM Press, Los Angeles, 487-496 (2001)
18. De Melo, C. and Paiva, A.: Multimodal Expression in Virtual Humans. Computer Animation and Virtual Worlds, 17 (3-4), 239-348 (2006)

19. Liu, Z. and Leung, K.: Script visualization (ScriptViz): a smart system that makes writing fun. Soft Computing, 10, 1, 34-40 (2006)
20. Hanser, E., Mc Kevitt, P., Lunney T., Condell, J.: NewsViz: emotional visualization of news stories. In: Inkpen, D. and Strapparava, C. (Eds.), Proc. of the NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, 125-130, Millennium Biltmore Hotel, Los Angeles, CA, USA, June 5th. (2010)
21. Rasheed, Z., Sheikh, Y., Shah, M.: On the Use of Computable Features for Film Classification. IEEE Transactions on Circuits and Systems for Video Technology, IEEE Circuits and Systems Society, 15(1), 52-64 (2005)
22. Shim, H. and Kang, B. G.: CAMEO - Camera, Audio and Motion with Emotion Orchestration for Immersive Cinematography. In: Proc. of the 2008 International Conference on Advances in Computer Entertainment Technology. ACE '08. ACM, New York, NY. 352, 115-118 (2008)
23. Kuo, F., Chiang, M., Shan, M., and Lee, S.: Emotion-based Music Recommendation by Association Discovery from Film Music. In: Proc. of the 13th Annual ACM international Conference on Multimedia, MULTIMEDIA '05. ACM, New York, 507-510 (2005)
24. Wahlster, W.: Smartkom: Foundations of Multimodal Dialogue Systems. Springer Verlag, Berlin/Heidelberg (2006)
25. Final Draft, `http://www.finaldraft.com` (Accessed: 11th April, 2016) (2016)
26. Maguire, T.: Performing Story on the Contemporary Stage. Palgrave MacMillan, Basingstoke (2015)
27. Fitzpatrick, L.: Gender and Affect in Testimonial Performance. Irish University Review, 45 (1), pp. 126-140 (2015)
28. Solon, A.J., Mc Kevitt, P. and Curran, K.: TeleMorph: A Fuzzy Logic Approach to Network-Aware Transmoding in Mobile Intelligent Multimedia Presentation Systems. In: Dumitras, A., Radha, H., Apostolopoulos, J. and Altunbasak, Y. (Eds.), Special issue on Network-Aware Multimedia Processing and Communications, IEEE Journal of Selected Topics In Signal Processing, 1(2) (August), 254-263 (2007)
29. Bridges, B. and Graham, R.: Electroacoustic Music as Embodied Cognitive Praxis: Denis Smalleys Theory of Spectromorphology as an Implicit Theory of Embodied Cognition. In: Proc. of the Electroacoustic Music Studies Network Conference, The Art of Electroaoustic Music (EMS15), University of Sheffield, England, June. Electroacoustic Music Studies Network (2015)
30. Moore, P., Lyons, F. and O'Hanlon, G.: The River Still Sings. In: Digital Arts and Humanities Conference, Ulster University, Magee, Derry/Londonderry, Northern Ireland, 10-13 September. Digital Humanities Ireland. 20 pp. (2013)
31. Melvin, A.: Sonic Motifs, Structure and Identity in Steve McQueens Hunger. The Soundtrack, 4 (1), 23-32 (2011)
32. Cunningham, H.: General Architecture for Text Engineering (GATE). `http://www.gate.ac.uk` (Accessed: 11th April, 2016) (2016)
33. Connexor, `http://www.connexor.com/nlplib` (Accessed: 11th April, 2016) (2016) (2016)
34. Tesniere, L.: Elements de syntaxe structurale. Klincksieck, Paris (1959)
35. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press. Cambridge (1998)
36. Strapparava, C., Valitutti, A.: WordNet-Affect: an Affective Extension of WordNet. In: Proc. of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, 1083-1086, 26th-28th May (2004)
37. Lexical Conceptual Structure (LCS) Database, `http://www.umiacs.umd.edu/~bonnie/LCS_Database_Documentation.html` (Accessed: 11th April, 2016) (2016)
38. Gelbukh, A.: EmoSenticNet. Available at, http://www.gelbukh.com/emosenticnet/ (Accessed: 20th February, 2014) (2014)
39. RapidMiner.: RapidMiner. Available at, `https://rapidminer.com/` (Accessed: 4th May, 2015) (2015)
40. Humanoid Animation Working Group, `http://www.h-anim.org` (Accessed: 11th April, 2016) (2016)
41. FreeTTS 1.2 - A speech synthesizer written entirely in the JavaTM programming language: `http://freetts.sourceforge.net/docs/index.php` (Accessed: 11th April, 2016) (2016)