Vol.2019-1
2019/01/10

感情的な文章分析と雰囲気に適したサウンドトラックの半自動選択をサポートする電子書籍作成ツール

E·book authoring to support affective text analysis and semi-automatic selection of mood-compatible soundtracks

Debaene Florian

Supervised by Prof. Michael Cohen

The first aim of this project is to analyze to what extent applying mood compatible soundtracks to a story can enhance the reading experience. This is an opportunity, using an eBook application able to create and read stories, to give a new face to literacy. This project feasibility is granted by capabilities of the digital nature of eBooks. The second aim of this project is to create an emotional analysis algorithm, to support users of the eBook application in choosing the music. This research tries to explore the capabilities of e-readers to enrich the reading experience on an emotional level. To do so, an application was created in order to conduct subjective tests regarding the improvement of the reading; and writing experience.

1.Introduction

1.1 Motivation

Firstly, nowadays, there are many applications and websites allowing sharing of creations such as music (SoundCloud[1]), drawing (DeviantArt[2], Pinterest[3]), text (different forums). Of these, the only one with a platform allowing editing and upload at the same time is drawing (with the app MediBang[4]). eBooks are well implemented on the market now, it is no surprise that people tend to create new content using eBooks capabilities. It's a hundred of books that needs one's backpack for only shelf. And it also is an immense playground to scientific experimentations.

Secondly, if soundtracks can enhance the experience of watching a movie, TV show, anime or even playing games, then why can't they also enhance the reading experience? The same way unsettling music can turn a merely creepy scene in a movie into a real nightmare, wouldn't it be able to turn a creepy scene in a book into a nightmare?

1.2 Improving the reading experience

In literacy, genres and style are evolving constantly, yet the main skill used is about structuring words. It remains in the spectrum of the written language. The ability to immerse the reader, to have him feeling emotions, is a responsibility borne by the author of the story only. What is beautiful with written language is that it is a framework which the reader fills in with personal images. The reader can imagine the story freely. But with new technologies, new opportunities emerged, enabling eventual new reading paradigms.

The research done on this paper is based on the fact that eBooks are a perfect medium to enable playing well timed soundtracks to enhance reading experience. But before going into details about the eBook applications, it is necessary to answer the questions raised above. Are soundtracks able to enhance the reading experience?

There's no doubt that music as used in movies, TV shows, anime, visual novels or even by itself, has the power to evoke emotions. It is especially interesting to realize that music both

represents emotions and elicits them in the audience. It is also proven that music evokes stronger emotions when combined with visual media.

In order to get an idea about how soundtracks can improve reading experience, we conducted two main experiments.

1.3 Related research

Audiobooks are probably the earliest example of trying to change the act of reading. Instead of reading yourself, you listen to someone's voice recorded reading the book. More usually used when you're in a situation you can't read the normal way, this is simply replacing the act of reading by hearing. In that sense here, the audio part is not there to enhance the experience but simply to "read" the story.

There are only a few projects that allow things such as playing background music while reading BOOKS to enhance the reading experience. One attempt to mix these two media involved adding soundtracks to online spoken books (Booktrack [5]). Booktrack also attempted to create a similar experience as the one proposed by this project. This website allows authors to write (to a certain extent) stories, and add BGMs (Background Music) and SFX (Sound Effects) to the story. In order to synchronize BGMs and SFXs with text, it uses an algorithm that estimates reading speed to change BGM and play SFX at the right time. After many attempts, the reason becomes more and more evident why this function isn't used widely and isn't really popular: It is clumsy. This algorithm simply can't handle random events like the fact that sometimes one's reading speed goes up or down, or one is interrupted by some external event. This ends up changing BGMs to wrong timings or even worse, plays SFX when not needed. It's highly disruptive for the reader's experience, and that's definitely something to avoid. This was proved during Holenderski, K., & Hu, J. research about Enriching reading experience with dramatic soundtracks [6]

1.4 Uniqueness of this application: Emotion analysis

Not everyone can choose soundtracks accurately; though many have enough artistic sense to do so, there are also many who do not. There's also the eventuality of an author being "lazy" and wishing to find soundtracks an easy way. Therefore, we implement a function that allow an "author" to invite the application, for a chosen part of the text (the "page"), to recommend soundtracks that might fit the mood of the scene.

Of course, a computer is not artistically perfect, so it will not be allowed to arbitrarily choose the soundtrack, and will only give advice by showing a list of recommended soundtracks. Since we're talking about art, related to subjective senses, it is likely that the computer might choose discordant soundtracks. But overall the computer will be able to help authors find some soundtracks that synergise with the story they are writing. This can result in enhancement of the experience, much like watching a movie in which one's brain processes the visuals and dialog, and the software handles the sound atmosphere.

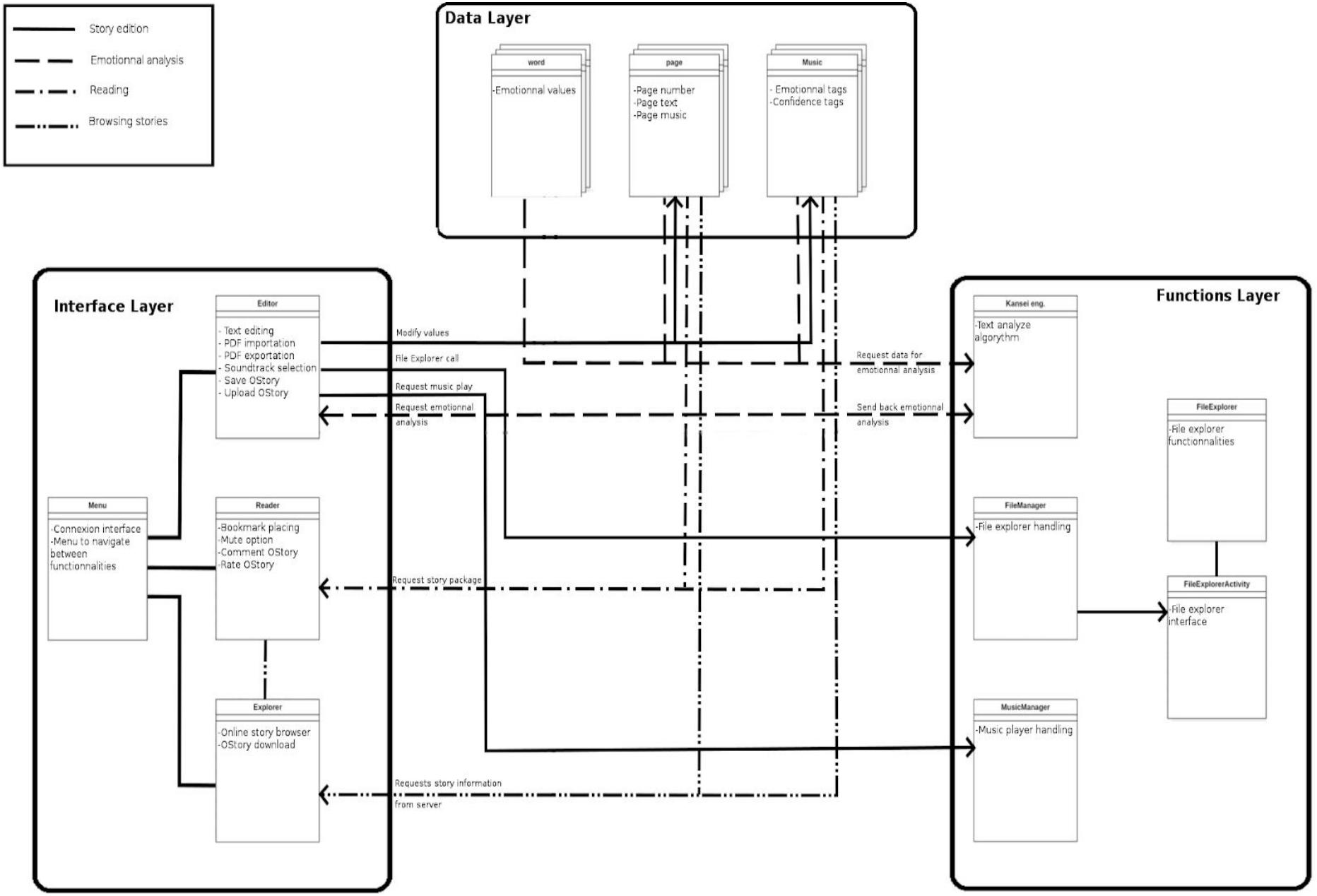
1.5 Aim of this project

Here are the main objectives of this application:

- Allow creating and editing of soundtrack-enhanced stories that can be hosted locally, then uploaded to a server once finished.
- Allow import of PDF files for those who do not wish to write a full story on a smartphone.
- Allow easy synchronisation of soundtracks (in the form of '.mp3' files for now) to stories in progress or PDF-imported stories.
- Assisted Soundtrack choice, by implementing affective computing algorithms able to analyze the mood of some text.
- Machine learning: the application learns about about words' emotional values from user feedback.
- Allow story-sharing, rating, and comments; on the application using a server that keeps all uploaded stories (with a short description of the book and cover art if applicable).

2. Methodology

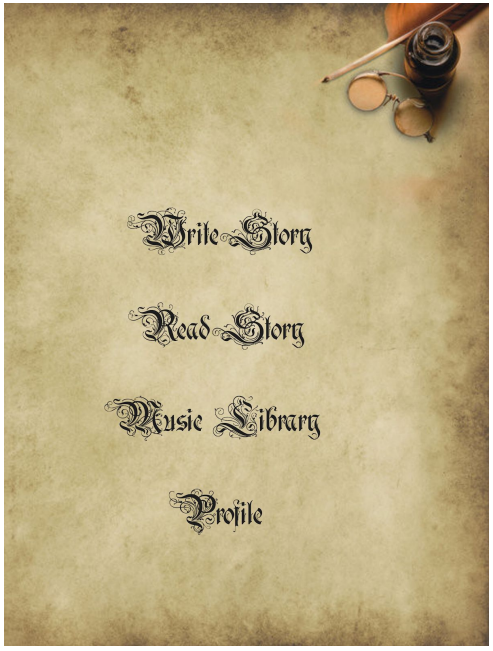
The following schematic details how the application runs, each aspect is detailed as follow. This project was developed respecting the MVC (Model View Controller) principles.


Story edition
Emotionnal analysis
Reading
Browsing stories
Data Layer
word
-Emotionnal values
page
-Page number
-Page text
-Page music
Music
- Emotionnal tags
-Confidence tags
Interface Layer
Editor
- Text editing
- PDF importation
- PDF exportation
- Soundtrack selection
- Save OStory
- Upload OStory
Menu
-Connexion interface
-Menu to navigate between functionnalities
Reader
-Bookmark placing
-Mute option
-Comment OStory
-Rate OStory
Explorer
-Online story browser
-OStory download
Modify values
File Explorer call
Request music play
Request emotionnal analysis
Request story package
Requests story information from server
Request data for emotionnal analysis
Send back emotionnal analysis
Functions Layer
Kansei eng.
-Text analyze algorythm
FileManager
-File explorer handling
MusicManager
-Music player handling
FileExplorer
-File explorer functionnalities
FileExplorerActivity
-File explorer interface

2.1 Overview explanation

2.1.1 Interface layer

The interface layer is composed of all classes of the applications related to the main features, the Editor, the Reader and the Explorer parts. Following here, a screenshot of the menu leading to these different parts. The menu uses a technique known as skeuomorphism, an emulation of a traditional “interface” style (Which is the page of a book), to immerse a little bit more the user into the fact he is reading a story. This is in this layer the user will be able to interact with OSTories uploaded on the server. (OSTory being the combination of the two terms “OST” and “Story”)



2.1.2 Data layer

The data layer is composed of every database or data structure used by this application. Information such as words’ emotional values,and stored stories, and stored music belong to this layer.

2.1.3 Functions layer

The function layer is composed of functions that run in the background, yet are necessary for the functioning of the application, such as the file manager, emotion analysis, music manager.

3.Affective computing algorithm

3.1 Word classification

As for the Affective computing algorithm, the choice was made not to use any API or already existing system but rather to use existing sources as inspiration and create a personal one. A database, a server able to host words database, and a link between this database and our application is deployed.

As for sources used as inspiration, the idea was to mix the “Tone Analyzer” API[7] with the PAD (Pleasure, Arousal, Dominance) Model (Russell circumplex model)[8].

There are two significantly different models for representing emotions: the categorical model and dimensional model.

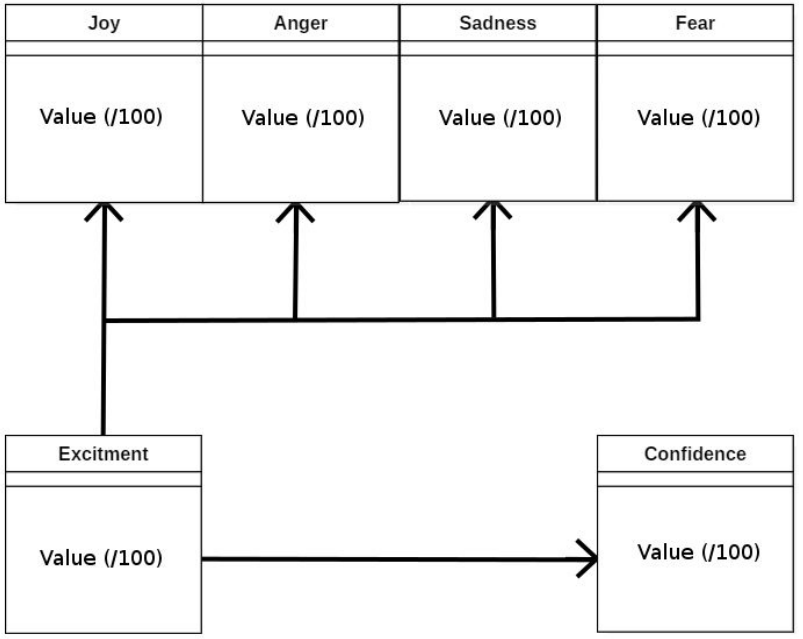
The categorical model assumes that there are discrete emotional categories (Ekman’s model[9] and its six basic emotions -anger, disgust, fear, joy, sadness, and surprise-) This model implies a specialization of the model, toward a specific field (such as teaching or storytelling).

A second approach is the dimensional model, which represents affect in a dimensional form (Russell 2003). Emotional states are related each other by a common set of dimensions (valence or arousal) and are generally defined in a two- or three-dimensional space. One might call this Valence-Arousal graph representation a trigonometric representation of emotions. Each emotion has specific coordinates in the graphic. Having two main axis, valence and arousal, respectively ranging the positiveness and excitement, it sometimes comes with a third dimension called dominance. This third dimension is used to differentiate situations under, or out, of control.

The two approaches both have their own pros and cons. And in this new model, the application will possess, tries to take advantages from these two well known methods and combine these advantages, resulting in a fairly accurate emotional analysis. If the word “fairly” is used there, it is obviously because no matter how advanced and intelligent a machine is, chances are extremely low that it will ever be able to understand completely how humans emotions work. Concepts such as irony, sarcasm are layers that even some humans lack appreciation of.

So in order to compile these two methods together, we chose to use a set of six values, quantified from 0 to 100. Four of these emotions, joy, anger, fear, and sadness, which are common emotions in three datasets commonly used for emotion analysis: SemEval-2007 “Affective Text”[10], ISEAR (International Survey on Emotion Antecedents and Reactions), and children’s fairy tales. Even though they won’t be used it is worth mentioning that, since it proves how important and basic these four emotions are.

As Values used along these main emotions, excitement and confidence were chosen. These two values are slightly different from the four main emotions. First, the confidence value is used the same way “dominance” is used in the valence-arousal-dominance representation. This value will be used to range how under control the situation seems to be in the current paragraph. Second, the excitement, quantized from 0 to 100, will weight all previous values, to make them more detailed and accurate (Ex: Not only fear, but more distress or panic). This will be explained below.



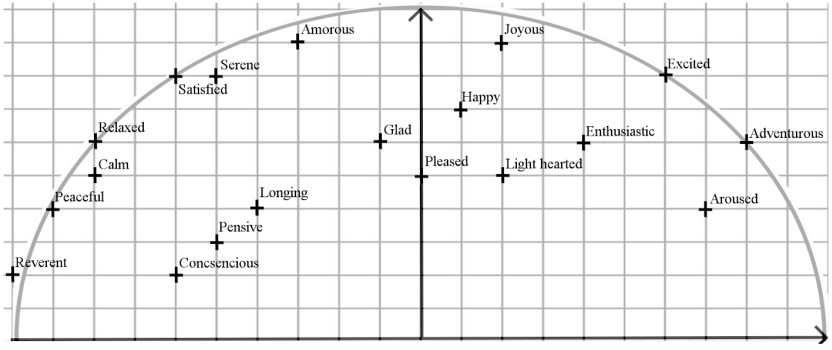
3.2 From values to tags

Using these data, two tags are issued for the full paragraph to analyze: first the combination emotion-excitement, then the combination confidence-excitement. Using these two dataset, the application now is able to cover a large amount of feelings. For instance, fear added to a high excitement and low confidence would be characterized as terror. And on the other side, fear added to low excitement and moderately high confidence would be characterized as an uneasy moment.

As for soundtrack characterization, tags are better suited than a dataset; a graph will be set for these emotions, so one may, for each emotion, determine a word corresponding to the values given by the text.

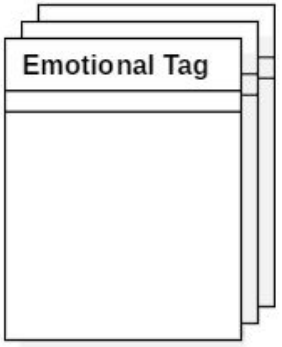
Here are these graphs for each emotion,

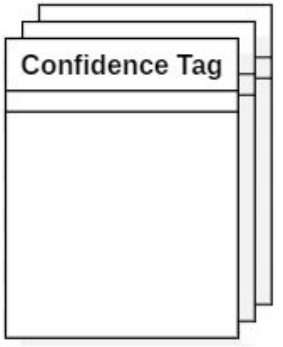
Here is an example for these graphs (Joy vs Excitement),



3.3 Soundtrack classification

In order to sort soundtracks, a simple system of tags is used. Each soundtrack has at least two tags. The first being the confidence tag, and the second being an emotional tag. Soundtracks can have as many tags as needed.





3.4 Soundtrack recommendation

In order to recommend a soundtrack to an author, the application proceeds as follows: it returns soundtracks corresponding to a percentage of each feeling, matching it with the confidence and emotional tag of these feelings.

To explain more clearly, let's use an example. A sample text returns these values:

Joy: 0
Fear: 80
Anger: 10
Sadness: 10
Confidence: 20

Excitement: 85

The application uses its different charts to find tags corresponding to these values when balanced by the excitement.

Joy: (null)
Fear: (Afraid)
Anger: (Bellicose)
Sadness: (Distressed)
Confidence: (Confused)

Now that the application has these 2 different datasets, it'll search for a soundtrack that matches the situation. First, it determines the quantity of soundtracks it picks up in each main emotion using the percentage. (Up to 20 soundtracks)

So keeping with the example before, it gives something like this:

Joy: 0 soundtrack
Fear: 16 soundtracks
Anger: 2 soundtracks
Sadness: 2 soundtracks.

Then it uses the tags to determine which soundtracks it picks in these different emotions, comparing the emotional and confidence tag with music tags from the chart. Let's take the example of fear. The application will look for music files which have tags "Afraid" and "Confused" and return them as suggestions. If not enough music files correspond to these, then it will first look for tags close to "panic" in the fear chart.

4. Results

Now that main functions related to editing and reading are written and works, the remaining work are to put these together in a single application, in order to deploy it on the android store later. Then to add the accounting part and the catalog part before release (Since these two parts are related). Once deployed, this application will allow a user to share his stories to the world, and get feedbacks about his story.

We also opened, as booktrack did, a new interesting path for story-telling, using soundtracks as a way to enhance the experience of the reader.

This was also the opportunity of creating an emotional recognition system in order to create an OSTs recommendation system that can actually help people that struggle with music selection, giving them advices based on the text they've written.

5. Discussion

This opens discussions to many features that could be added to such a program. Especially in term of machine learning, If the user was able to choose a word, and ask the application what emotion it think it is related to, then the author should be able to recommend an update for this word. If he thinks the application is wrong, he could choose an emotion that is more suitable, then the application would learn from all authors feedbacks and after a while, be quite accurate about emotion recognition. This might even become a tool that will be used for more things than this particular application.

The application could also be adapted to work on iOS (for iPhones and iPads) or even as a desktop computer application, facilitating the writing part, sacrificing the "read whenever you want" part.

5. References

[1] - SoundCloud: https://soundcloud.com/

[2] - DeviantArt: https://www.deviantart.com/

[3] - Pinterest: https://www.pinterest.com/

[4] - Medibang Paint: https://medibang.com/

[5] - Booktrack (Related project, Hypermedia): https://booktrack.com/

[6] Holenderski, K., & Hu, J. (2013). Enriching reading experience with dramatic soundtracks. Adv. Multimedia Technol, 2(1), 24-31.

[7] - Tone analyser (emotion analysis): https://tone-analyzer-demo.ng.bluemix.net

[8] - Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161.

[9] - Ekman, P. (1992). An argument for basic emotions. Cognition & emotion, 6(3-4), 169-200.

[10] - Strapparava, Carlo, and Rada Mihalcea. "Semeval-2007 task 14: Affective text." Proceedings of the 4th international workshop on semantic evaluations. Association for Computational Linguistics, 2007.

- Madefire (Hypermedia, animated, dubbed comics): https://www.madefire.com/

- Blizzard Entertainment's animated comics (Published on madefire website) https://www.madefire.com/

- Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pages 62–70, Los Angeles, California, June 2010.

- Cohen, A. J. (2001). Music as a source of emotion in film. Music and emotion: Theory and research, 249-272.

- Michael Cohen. "From Killing trees to Executing Bits: A survey of Computer-Enabled Reading Enhancements for Evolving Literacy". In: Proc. VSMM: Int. Conf. on Virtual systems and Multimedia. Hong Kong, Dec. 2014. ISBN: 978-1-4799-7227-2. DOI: 10.1109/VSMM.2014.7136686.

- Kindle (ebooks, Spoken books, TTS)