Associating Colors to Emotional Concepts Extracted from Unstructured Texts

Alberto Fernández-Isabel, Antonio F. G. Sevilla, Alberto Díaz

NIL**, Facultad de Informática, Universidad Complutense de Madrid
c/ Profesor José García Santesmases 9, 28040, Madrid, Spain
afernandezisabel@ucm.es, afgs@ucm.es, albertodiaz@fdi.ucm.es

Abstract. Software creativity is a pervasive branch inside Artificial Intelligence where multiple approaches have place. In the case of developing a software artist the main difficulty resides in retrieving and expressing the feelings of the context properly. This paper addresses that issue introducing a framework able to produce color palettes and abstract paintings based on processing unstructured text. To achieve it, the tool extracts the concepts related to human mood from the text once it is analyzed. Then, they are checked to identify the color that represents their psychological meaning using related literature. The resulting picture is built according to the dimension of the canvas and the amount of colors from the palette obtained. A case study illustrates the applicability of the proposal using two texts selected from Wikipedia. They describe disparate concepts as love or death. A related work is considered to situate the approach in the context and establish comparisons.

Keywords: Natural language processing, concept extraction, abtract painting, painter software, semantic expression, information retrieval

1 Introduction

Abstract painting presents difficulties to produce objective evaluations. These are related to its foundations as subjective field. This subjectivity resides in the feelings, emotions and mood an individual (i.e., the artist) wants to manifest or illustrate, and the evaluation made by others with different backgrounds or sensibilities [10]. In order to alleviate these problems there are multiple approaches that evaluate the implicit psychology of colors and their possible meanings [5, 12]. These associations (though they are culture-specific [1]) allow establishing certain connections and relationships among the abstract art and the expressed feelings and their representations.

This literature can be adopted by intelligent software tools in order to reproduce the inherent creativity [21]. This fact might lead these tools to generate abstract paintings based on forms or colors. These can be organized into the picture for simulating some specific techniques used by important human artists

** Group of Natural Interaction based on Language

[14] or distributed over the canvas randomly focusing on more generic abstract painting processes [20].

Our proposal introduces the framework called CALyPSe (Conceptual Abstract and Lyric Painting System) that is able to produce color palettes and abstract paintings. The resulting pictures are generated from concepts related to human feelings and emotions extracted from an input text. These are checked to predefined notions which are associated to a specific color following the relationships introduced in [12]. The final painting is built according to the dimensions of the canvas and the color palette. This latter is generated considering the amount of concepts identified and harmony color techniques [4]. The organization of pixels which compose the complete image follows a randomly distribution.

The framework is composed by three different modules: *concept extraction*, *concept storage* and *painting generation*. They count with their respective manager which is responsible for connecting their internal tools and make them work together. The first module takes as input an unstructured text and extract the concepts from it through Freeling [18] and a *semantic analyzer*. In the second one its manager stores these concepts that can be related to a specific color in the *knowledge database*. These relationships might be produced directly or through synonyms. These latter are obtained using WordNet [15]. Optionally, the concepts gathered can be searched on Wikipedia, processing the first paragraph of their definition if the web page exists. The third module is oriented to develop the final painting according to the information collected.

A case study illustrates an experiment where abstract paintings based on obtaining concepts from texts are generated. Two different texts gathered from Wikipedia related to human emotions and feelings are selected. The first one describes *love* while the second talks about *death*. The tool generates two abstract paintings according to them. These are analyzed both qualitative and quantitative, focusing on two specific points: the colors used to produce the pictures with each one of the selected texts, and the amount of different concepts identified and their proportion.

The rest of the paper is organized as follows. Section 2 compares this proposal with related work. The CALyPSe framework is introduced in Section 3 delving into their modules and managers that are responsible for them. The case study in Section 4 shows the application of the approach. Finally, Section 5 discusses some conclusions and future work which concern to the issue.

2 Related work

The framework presented in this approach links natural language processing and information retrieval with the expression of emotions through abstract painting. This section discusses the existing tools and its foundations comparing them with their alternatives.

The framework accomplishes the extraction of concepts from an input text according to its dependency analysis through Freeling [18]. This tool automatizes the required steps using different layers of linguistic analysis. A similar alterna-

tive consists of the Stanford parser [11] which is also able to generate their own type of specific relations and dependencies applying stochastic analysis. Both tools are designed to ease their integration into a more complex pipeline.

The semantic information retrieval and lexical support for obtaining synonyms is provided by WordNet [15]. It is the standard lexical knowledge base where certain related items can be collected (e.g., verbs, nouns or adjectives).

Regarding the painting artist frameworks, there are multiple approaches but three main perspectives are related to our proposal: frameworks that mimic existing human artists styles, creative painting based on feelings and emotions, and collage generation.

Mimic human styles is a complex issue due to each painter develops a personal style. Nevertheless there are proposals that try to evoke some notions or standards related to a specific painter [14] or painting techniques [3]. Others more generic are focused on generating images simulating their own style as painters [6] or rendering images to simulate certain strokes techniques in paintings [7].

Creative painting is oriented to simulate human emotions through an intelligent software that follows some rules or background [9]. One of the famous frameworks in this field is The Painting Fool [8]. It is based on creativity associated to the elaboration of pictures through non-photorealistic rendering.

Collage generation consists of producing a picture integrating different images that could have sense together. The information to represent can be obtained from different sources. One of the most common is the processing of unstructured texts [13]. The different images which are added to the painting are usually extracted from the web.

There also are frameworks similar to CALyPSe. For instance, [17] is able to produce images according to a set of adjectives that describe the input.

CALyPSe framework generates its paintings taking as a reference the abstract art, while it identifies the concepts related to thoughts and emotions from an input text associating them to colors. The conception process of these paintings uses a random algorithm (instead of positioning or organizing the pixels) in order to simulate modern art guidelines [20].

3 Associating colors to emotional concepts

This approach presents CALyPSe which is a framework oriented to produce color palettes and abstract pictures that express the emotions and feelings included in an unstructured text. To accomplish it, the tool has a set of modules in order to classify the different tasks to perform (e.g., concept extraction or color comparisons) until the picture is generated.

It has an structure based on three main modules (see Fig. 1): *concept extraction*, *concept storage* and *painting generation*. The first processes the input texts and organizes the concepts captured from them by sentences. The second has the storage structures and implements the connection to WordNet database[15]. The third is in charge of collecting the knowledge stored and producing the resulting abstract picture.

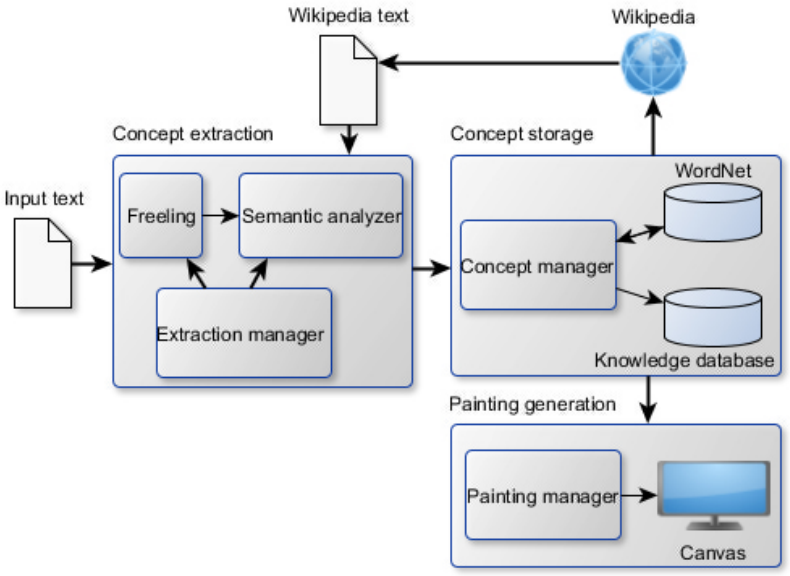

Fig. 1. Excerpt of the framework structure.

Section 3.1 introduces the *concept extraction* module, while the *concept storage* is described in Section 3.2. The issues related to the *painting generation* module are addressed in Section 3.3.

3.1 Concept extraction

This module is in charge of processing the current input text identifying its concepts. To achieve this operation it uses two items: Freeling [18] and a *semantic analyzer*. The *extraction manager* synchronizes both in order to generate the appropriate result (see Fig. 1).

Freeling [18] evaluates the input text achieving a dependency analysis. This allows producing its syntax structure and the identification of the different types of words (e.g., nouns, verbs or adjectives). This structure is captured by the *extraction manager*.

The *semantic analyzer* receives as input the syntax structure stored by the *extraction manager*. It examines the different elements and extracts the lemmas of the concepts discarding the pronouns and the articles (they are not appropriated to be compared to the concepts related to colors). The manager is also in charge of producing the final result of the module, organizing these lemmas by sentence following the order of precedence provided by the text.

Note that the module obtains the input texts through three different ways according to the features provided by the framework. It is able to analyze unstructured text, read files and collect their texts, or process a web page selecting

Key	Value (RGB)
Innocence	(253,233,242)
Love	(250,63,66)
...	...
Purity	(237,237,245)

Table 1. Excerpt of the *color structure* in the *knowledge database*.

Sympathy	Harmony	...	Love
3	2		3
Sympathy	Harmony		Love
Dog	Landscape		Girl
Friend			Dog

Table 2. Excerpt of the *concept structure* in the *knowledge database* associated to the text: *"The girl feels love and sympathy for her dog." Her friend also has sympathy for the dog. The landscape generates much harmony in them."*

only its paragraphs. This latter uses techniques related to web scraping [16] that allow extracting only the raw text.

3.2 Concept storage

It is the module where the information collected from the input texts is processed and stored. It is composed by a *knowledge database*, a WordNet [15] database connection and the *concept manager* that control the interactions among these elements (see Fig. 1).

The *knowledge database* presents two internal structures: *color structure* and *concept structure*. The first one stores 179 relations among concepts (i.e., keys) and colors collected from the literature that concerns to psychology of colors [12] (see Table 1). The second contains a set of head concepts which are the keys with an associated color in the first structure. Each one of the positions of the *concept structure* stores how many times the head element has been related to the concepts from the input text and a list of items. This list might contain the same concept of the head element or associated concepts. The list supports the simulation of associative learning based on words in sentences [19]. In order to achieve it, when a notion does not match with any head element or item of lists but others of the same sentence does, the unmatched concept is stored in the list of these latter (see Table 2). This establishes a semantic relationship among the concept and the rest of notions that could be identified in its same sentence.

WordNet [15] is used by the *concept manager* to find synonyms. It allows associating more notions from the text with the head elements of the *concept structure*.

The module presents an optional set of items related to Wikipedia in order to increment the color enrichment of the final painting. It is able to extend each

one of the concepts extracted from the text in order to find a description of it in its related Wikipedia web page. In the case the page exists, it is scraped and the first paragraph is gathered. This text is sent to the *concept extraction module* where its concepts are obtained. These are stored in the *concept structure* if they produce matches with its head elements. In order to avoid infinite cycles the concepts selected from this part has an special identification.

3.3 Painting generation

This module produces the color palette and the resulting picture. It is accomplished using both structures contained in the *knowledge database* (i.e., *concept structure* and *color structure*).

The *painting manager* is responsible for achieving the painting generation. It applies a set of rules related to color harmony [4] in order to generate the color palette. These rules are implemented through a basic filter. It allows discarding mixtures among combinations of similar colors (where the most important prevail) that do not produce an appropriate contrast to human eye (e.g., red color does not match to pink color).

Then, the *painting manager* produces the final composition obtaining the number of elements of the positions of the *concept structure* that are related to the filtered colors. This is obtained comparing the head elements to the keys of the *color structure*. Thus, the painting is built randomly according to the proportion between the number of elements and the dimension of the canvas.

4 Case Study

The case study illustrates how the tool produces two different palettes and pictures based on them from unstructured texts selected from Wikipedia. These texts introduce information about two specific human feelings or thoughts: *love* and *death*. The images obtained from them are compared in a visual way and the amount of concepts detected during the process are analyzed. The Wikipedia optional feature provided by the framework (see Section 3.2) is not considered in this case.

The framework starts reseting the *concept structure* of the *knowledge database*. Then the *extraction manager* scraps the corresponding web pages obtaining the input texts. These texts are provided to Freeling [18] in order to generate their dependency analysis. This output is processed by the *semantic analyzer* which is in charge of obtaining the lemmas of the identified concepts.

In the next step the concepts are analyzed in the *second module*. In it, the framework tries to link them to the head elements or to the items of their lists provided by the *concept structure*. In order to increment the matches, the tool uses WordNet [15]. It obtains synonyms that might be linked to the head elements. If the current concept is directly related or conversely one of its synonyms, the first one is stored in the list of items of the specific head element.

Input text	Head elements	Items in lists	Synonyms	% Matches
Love	8	1433	5	11%
Death	2	48	0	5%

Table 3. Summarization of matches in texts related to *love* and *death*.

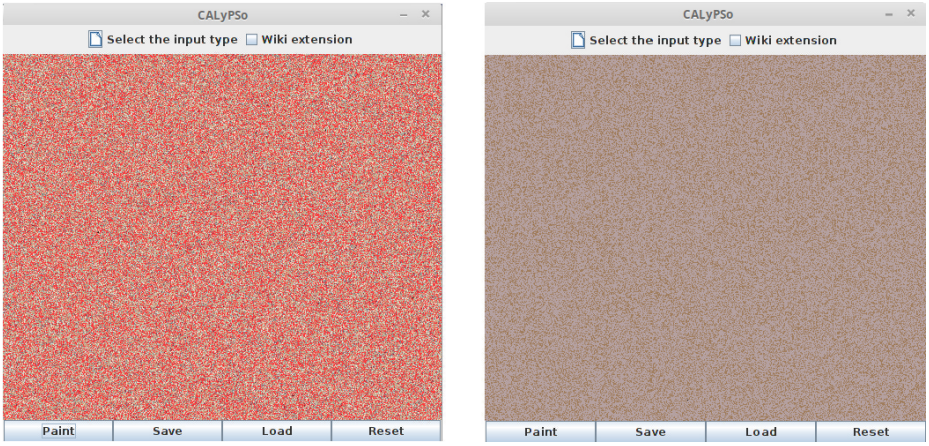

Fig. 2. Pictures generated by the framework related to *love* and *death* texts.

In the case of *love* text, some of its concepts can be checked easily (e.g., *affection* is associated to a head element that contains the *love* concept). The text that concerns to *death* has similar situations with other concepts (e.g., *solitude*) but no with the *death* concept. It happens because of the *death* concept is not considered in the *color structure* as a key (i.e., it is not evaluated in the literature [12]). The associative learning [19] solves this situation inserting the *death* notion into the list of items of the concepts related to head elements of its same sentence. Thus, when *death* appears again it will be matched.

Once the concepts are stored in the *concept structure*, the palette is produced. In both cases there are not similar colors to filter so that every head element is selected in order to generate the paintings. *Love* painting is composed by eight different colors (i.e., head elements) while the *death* presents only two colors (see Table 3). Nevertheless, the amount of pixels in both pictures suggests a high number of matching with the same head elements and lists of items.

Regarding the visual effect, the picture that depicts the *love* text is noticeably red with several spots of different colors (see Fig. 2). In the second picture concerned to *death* text, there are multiple spots with shades of gray. This makes the painting sad and dull. Therefore, it could be said that visually both pictures seem to illustrate some of the meanings of the concepts introduced in the texts.

Finally, delving into a quantitative analysis it can be found that in the text concerned to *love* there are more matches related to the head elements of the *concept structure* (i.e., more colors are used to draw the painting) than in the

death text. The same occurs in the case of the concepts of the lists of items (see Table 3). The amount of related synonyms is low in both cases or even irrelevant (*death* text does not provide concepts where their synonyms match to head elements). An enriched *color structure* in the *knowledge database* with a wider list of concepts associated to colors could enhance the problem in future experiments.

5 Conclusions

This paper has introduced the CALyPSe framework to generate color palettes and abstract pictures. These are produced from unstructured texts which are processed to gather their emotional concepts and associate them to colors using related literature [12].

The tool consists of three modules (*concept extraction*, *concept storage* and *painting generation*) and their respective managers. The first one is in charge of processing unstructured text using Freeling [18] and a *semantic analyzer*. The second stores the concepts related to colors (i.e., *color structure*) and the matchings among these and the notions coming from the input text (i.e., *concept structure*) in its *knowledge database*. It presents an optional feature that eases the acquisition of extra information. It is based on searching each concept on Wikipedia. The third is focusing on producing the color palette through color harmony techniques [4] and drawing the painting.

The case study shows the viability of the proposal using two different texts that have been selected from Wikipedia. They describe concepts as *love* and *death*. Their visual representations are compared observing the differences among them. A quantitative analysis is achieved focusing on the amount of concepts extracted from the texts that match to the head elements provided by the *concept structure*, or which ones are stored in the lists of items. Synonyms associated to the concepts are also considered showing a low hit rate.

More experiments support the proposal but it is still ongoing work with open issues. New concepts from literature related to the psychology of colors have to be inserted in the *color structure* of the *knowledge database*. This will allow obtaining higher percentages in the matches among concepts from texts and head elements. Another point to consider consists of adopting some painting techniques and image rendering [2] in order to produce more realistic pictures.

Acknowledgements

This work is funded by ConCreTe. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733.

This research is funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (TIN2015-66655-R (MINECO/FEDER)).

References

1. Aslam, M.M.: Are you selling the right colour? a cross-cultural review of colour as a marketing cue. Journal of marketing communications **12**(1), 15–30 (2006)
2. Baxter, B., Scheib, V., Lin, M.C., Manocha, D.: Dab: interactive haptic painting with 3d virtual brushes. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 461–468. ACM (2001)
3. Baxter, W., Wendt, J., Lin, M.C.: Impasto: a realistic, interactive model for paint. In: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering, pp. 45–148. ACM (2004)
4. Burchett, K.E.: Color harmony. Color Research & Application **27**(1), 28–31 (2002)
5. Cherry, K.: Color psychology: How colors impact moods, feelings, and behaviors. About Education (2015)
6. Cohen, H.: The further exploits of aaron, painter. Stanford Humanities Review **4**(2), 141–158 (1995)
7. Colton, S.: Stroke matching for paint dances. In: Proceedings of the Sixth international conference on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 67–74. Eurographics Association (2010)
8. Colton, S.: The painting fool: Stories from building an automated painter. In: Computers and creativity, pp. 3–38. Springer (2012)
9. Colton, S., Pease, A., Charnley, J.: Computational creativity theory: The face and idea descriptive models. In: Proceedings of the Second International Conference on Computational Creativity, pp. 90–95 (2011)
10. Csikszentmihalyi, M., Robinson, R.E.: The art of seeing: An interpretation of the aesthetic encounter. Getty Publications (1990)
11. De Marneffe, M.C., MacCartney, B., Manning, C.D., et al.: Generating typed dependency parses from phrase structure parses. In: Proceedings of LREC, vol. 6, pp. 449–454 (2006)
12. Heller, E.: Psychology of colour. Yedam (2005)
13. Krzeczkowska, A., El-Hage, J., Colton, S., Clark, S.: Automated collage generation-with intent. In: Proceedings of the 1st international conference on computational creativity, p. 20 (2010)
14. Lee, S., Olsen, S., Gooch, B.: Simulating and analysing jackson pollock's paintings. Journal of Mathematics and the Arts **1**(2), 73–83 (2007)
15. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)
16. Munzert, S., Rubba, C., Meißner, P., Nyhuis, D.: Automated data collection with R: A practical guide to web scraping and text mining. John Wiley & Sons (2014)
17. Norton, D., Heath, D., Ventura, D.: Establishing appreciation in a creative system. In: Proceedings of the international conference on computational creativity, vol. 301, pp. 26–35. Citeseer (2010)
18. Padró, L., Stanilovsky, E.: Freeling 3.0: Towards wider multilinguality. In: LREC2012 (2012)
19. Prior, A., Bentin, S.: Word associations are formed incidentally during sentential semantic integration. Acta psychologica **127**(1), 57–71 (2008)
20. Schönlieb, C.B., Schubert, F.: Random simulations for generative art construction–some examples. Journal of Mathematics and the Arts **7**(1), 29–39 (2013)
21. Wiggins, G.A.: Searching for computational creativity. New Generation Computing **24**(3), 209–222 (2006)