MUSTI - Multimodal Understanding of Smells in Texts and Images at MediaEval 2022

Ali Hürriyetoğlu[1,*], Teresa Paccosi[2], Stefano Menini[2], Mathias Zinnen[3], Pasquale Lisena[4], Kiymet Akdemir[1], Raphaël Troncy[4] and Marieke van Erp[1]

[1]*KNAW Humanities Cluster DHLab, the Netherlands*
[2]*Fondazione Bruno Kessler, Italy*
[3]*Pattern Recognition Lab, Friedrich-Alexander-Universität, Germany*
[4]*EURECOM, France*

Abstract

MUSTI aims to collect information about smell from digital text and image collections from the 17th to 20th century in a multilingual setting. More precisely, MUSTI studies the relatedness of evocation of smells (smell sources being identified, objects being detected, gestures being mentioned or recognized) between texts and images. The main task is a binary classification task and entails identifying whether a pair of image and a text snippet contains the same smell source independent of what is the smell source. An optional sub-task is the determination of the smell sources that make the respective pair related.

1. Introduction

To make sense of digital heritage collections, one can go beyond an oculo-centric approach and engage with the olfactory dimension as it offers a powerful and direct entry to our emotions and memories. TheMultimodal Understanding of Smells in Texts and Images (MUSTI[1]) task at MediaEval 2022[2] aims to accelerate the understanding of olfactory references in multilingual texts and images as well as the connection between these modalities. As recent and ongoing exhibitions at Mauritshuis[3] in The Hague, Museum Ulm[4] in Ulm, and the Prado Museum[5] in Madrid demonstrate, museums and galleries are keen to enrich museum visits with olfactory components - either for a more immersive experience or to create a more inclusive experience for visitors presenting visual impairments.

Reinterpreting historical scents is attracting attention from various research disciplines [1] leading in some cases to novel collaborations with perfumers such as the *Scent of the Golden Age* candle developed after a recipe by Constantijn Huygens in a collaboration between historians and a perfume maker.[6] To ensure that such enrichments are grounded in historically correct

MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online
*Corresponding author.
ali.hurriyetoglu@dh.huc.knaw.nl (A. Hürriyetoğlu); tpaccosi@fbk.eu (T. Paccosi); menini@fbk.eu (S. Menini); mathias.zinnen@fau.de (M. Zinnen); pasquale.lisena@eurecom.fr (P. Lisena); kiymet.akdemir@dh.huc.knaw.nl (K. Akdemir); raphael.troncy@eurecom.fr (R. Troncy); marieke.van.erp@dh.huc.knaw.nl (M. v. Erp)
© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

[1]https://multimediaeval.github.io/editions/2022/tasks/musti/
[2]https://multimediaeval.github.io/editions/2022/
[3]https://www.mauritshuis.nl/en/press-releases/smell-the-art-fleeting-scents-in-colour/
[4]https://museumulm.de/veranstaltung/der-nase-nach/
[5]https://www.museodelprado.es/en/whats-on/exhibition/the-essence-of-a-painting-an-olfactory-exhibition/07849a71-d9b0-faeb-94c1-689f2614f8d0
[6]https://www.huygens.fr/en/essential-oil-candles/2523-826-constantijn-huygens-candle.html

contexts, language and computer vision technologies can aid in finding olfactory relevant examples in the collections and related sources.

While the sense of smell is of vital importance in our day-to-day lives, it has received little attention within the natural language processing and computer vision communities. While there are some olfactory lexicons [2, 3], the Odeuropa text benchmark dataset is the first multilingual, cross-domain text dataset focused on smell references [4]. Similarly, for computer vision, no prior datasets existed until the ODOR challenge dataset [5, 6]. The MUSTI task brings these modalities together, inviting the research community to explore parallels and complementarities in the way smells are described and depicted in different modalities. MUSTI offers texts in English (EN), German (DE), French (FR), and Italian (IT).

The remainder of this paper is structured as follows: we provide details on the task in Section 2. The steps followed to prepare data for training and evaluation are provided in Section 3. The evaluation methodology is described in Section 4 before concluding in Section 5.[7]

2. Task description

The manner in which humans engage with smell is a prime example of intangible cultural heritage: the way smells are created, in what situations they are used, but also how they are appreciated are highly culturally-dependent. By engaging with expressions of smells in texts and images across multiple genres and multiple languages over a long period of time, we expect to gain insights into how smells have affected human interactions through time.

Smell is an underrepresented dimension of many multimedia analysis and representation tasks. The goal of MUSTI is to advance the understanding of how smells are described and depicted by recognizing and connecting references to smells in texts and images. In this shared task, participants are provided with multilingual texts and images, from the 17th to the 20th century, that pertain to smell (i.e. selected because they evoke smells).

Task participants should develop language and image recognition technologies to predict whether a text passage and an image evoke the same smell source or not. In a subsequent optional sub-task, the participants are also asked to identify common smell source(s) such as the person, object or place that has a specific smell, or that produces odours (e.g. plant, animal, perfume, human), between the text passages and the images.

The "Quest for insight" part, which goes beyond quantitative evaluation by posing questions that endorse a deeper understanding of the challenge, including data and the strengths and weaknesses of particular types of approaches, of MUSTI consists of the following questions:

1. What does it mean for a text passage and an image to be related in terms of smell?
2. Do different text and image genres reference smell differently?
3. Do different languages reference smell differently?
4. How do references to smell in texts and images change over time?
5. How do relationships between smell references in texts and images change over time?

3. Data preparation and Release

The data consists of copyright-free texts (historical books and documents) and partly copyrighted images from open repositories.[8] We offer sentences that should be matched with images, which are selected from RKD, Bildindex der Kunst und Architektur, Museum Boijmans, Ashmolean

[7]Task related materials can be found at https://github.com/Odeuropa/musti_mediaeval2022.
[8]For copyrighted materials, the source URL is shared with participants.

Museum Oxford, Plateforme ouverte du patrimoine and annotated with 80+ categories of smell objects and gestures such as flowers, food, animals, sniffing and holding the nose.[9]

3.1. Candidate creation

In a first step, we search for texts potentially related to the images used for the task. For English, the text are extracted from the British Library corpus, the Early English Books Online (EEBO), and from Project Gutenberg. For Italian, the sources are Project Gutenberg, Liber Liber, and Wikisource. For French, the sources include Gutenberg, Gallica, and the ARTFL Project. Finally, Berlin State Library OCRs and Deutsches Textarchiv are utilized for German text data.

Each sentence in the corpora is lemmatized. Next, candidate extraction is based on the presence in the text of words from three different image metadata fields (when available):

- *Title*: the nouns (lemmatized) in the title, representing the subjects/objects in the painting.
- *Categories*: labels of visible smell sources, identified by the annotators within the paintings.
- *Keywords*: list of keywords associated with images in their respective collections.

We only keep sentences sharing content with at least two of the three fields. Additionally, we identified if each sentence does or does not contain a *Smell Word*, e.g. *stink, smell, odor, sniff, fetid, smelly*. Sentences containing a *Smell Word* are more likely to be instances where images and texts represent the same smell while sentences without *Smell Word* are usually only about the same subject without evoking any smell related to it.[10]

3.2. Annotation methodology

We run our sentence extractor based on *categories*, *keywords*, and *titles* to create text-image pairs. Using the sentence extractor, we look for pictures containing more categories, keywords or nouns from the title to find texts as relevant as possible to the image.

For **Subtask 1**, we annotate matches (pairs of texts and images) as evoking the same smell (YES) or not (NO). The images have already been selected as either explicitly or implicitly evoking a smell. An example of a YES instance is the text "*Having secured rooms in this establishment, we started for a walk through the town. The first thing that strikes a stranger upon his arrival at Accra is* ***a strong, all - pervading smell of pig***" matching with a picture representing pigs.

The annotation NO is used when there is no match in terms of the smell represented in the text and the image. An example is the text "*How I do hate tobacco and that disgusting habit of smoking, cried Geraldine, who was a very fastidious little lady, and could not endure the* ***smell of a pipe or even a cigar***" with a picture in which there is no pipes, smoke, or cigars.

The NO examples have two possible characteristics that make the matching more difficult: i) a *negation* in the text and ii) the case when the text and the image contain the same object(s) but not the same smell. We consider an image-text pair a negation if the object which evokes the smell represented in the picture is negated in the text. An example of negation is the text "*They had* ***no censers*** *to perfume the air extinguishing the morning fragrance, nor bore they their diamond crosiers through the streets*" with a picture in which there is one or more censers. The other complex cases are the ones in which the smell-evoking object is mentioned in the text but does not evoke the same smell represented in the picture. These are metaphorical or similarity sentences like "*Will you not believe these miracles? Which however of themselves they may shine* ***like a candle*** *lighted up*" with a picture representing candles. Sentences mentioning the same

[9]The taxonomy is described in https://odeuropa.eu/wp-content/uploads/2022/05/D2_1_TaxonomyOfOlfactoryPhenomena.pdf.

[10]The translations of the English taxonomy in respective languages are utilized.

smell source represented in the picture but with a different acceptation, e.g. "*The barracks and all buildings were heaps of ruins,* ***the fires*** *still burning,* ***the smoke and stench from which*** *were offensive and suffocating*" with a picture representing a different kind of smoke such as the smoke of a pipe. Finally, sentences mentioning the same object as the picture but in a different state, e.g. "*The Sunday or visiting dress of Mrs. Yates consisted of a thick sort of silk [...]* ***printed*** *in a large chintz pattern on a white ground , in which* ***butterflies and flowers, of unknown and fantastic varieties, predominated***." are considered as tricky.

In **Subtask 2**, we add a second layer of annotations to the image-text pairs annotated with YES in the Subtask 1. In this case, the annotation indicates the person, object or place (multiple annotations are possible) related to the evoked smell appearing both in the text and in the image. For instance in *"Having secured rooms in this establishment, we started for a walk through the town. The first thing that strikes a stranger upon his arrival at Accra is a strong, all - pervading smell of pig."*, the smell evoking element appearing both in the text and in an image retrieved from RKD[11] is *"pig"*.

We split the data in a training and a test sets, with a similar proportion of YES and NO labels. The YES instances are labeled for subtask 2 as well. The number of instances in the training and test sets are respectively 795 and 200 for EN, 480 and 213 for DE, 300 and 200 for FR, and 799 and 201 for IT.

4. Evaluation

We utilize the binary F1-macro score as provided by Scikit-learn library [12] for Subtask 1. We propose a naive baseline F1-macro score, based on the majority label, that reached 27.14% on the whole test set (42.85% for EN, 42.89% for DE, 22.07% for FR, and 42.73% for IT). Subtask 2 is evaluated by averaging F1-macro scores calculated for predictions of each instance. The F1-macro calculation is performed on the overlap between gold and predicted labels. The label mismatch between gold and predicted labels is resolved by creating a label set by unifying and alphabetically sorting what appears both in gold and predicted labels for each prediction.

5. Conclusion

We propose the MUSTI challenge for detecting text-image pairs that contain references to the same smell source(s) as two subtasks, i.e. prediction of the existence of the relationship as a binary classification task and detecting the smell-related tokens that are connected through this relationship. The texts and images are extracted from historical archives. Moreover, we identified difficult examples of text during the annotation phase. Finally, a baseline based on prediction as majority label was calculated for the test data released. The system performances on the MUSTI dataset are described in [7] and [8].

Acknowledgements

This work has been partially supported by European Union's Horizon 2020 research and innovation programme within the Odeuropa project (grant agreement No. 101004469). We would like to thank Hang Tran (Friedrich-Alexander-Universität Erlangen-Nürnberg) and Marta Sandri (University of Pavia) who significantly contributed to the annotation effort.

[11] https://rkd.nl/nl/explore/images/193733

[12] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

References

[1] B. Huber, T. Larsen, R. N. Spengler, N. Boivin, How to use modern science to reconstruct ancient scents, Nature Human Behaviour 6 (2022) 611–614. URL: https://doi.org/10.1038/s41562-022-01325-7.

[2] S. S. Tekiroğlu, G. Özbal, C. Strapparava, Sensicon: An automatically constructed sensorial lexicon, in: Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1511–1521. URL: https://aclanthology.org/D14-1160.

[3] S. Menini, T. Paccosi, S. S. Tekiroğlu, S. Tonelli, Building a multilingual taxonomy of olfactory terms with timestamps, in: 13th Language Resources and Evaluation Conference (LREC), European Language Resources Association, Marseille, France, 2022, pp. 4030–4039. URL: https://aclanthology.org/2022.lrec-1.429.

[4] S. Menini, T. Paccosi, S. Tonelli, M. Van Erp, I. Leemans, P. Lisena, R. Troncy, W. Tullett, A. Hürriyetoğlu, G. Dijkstra, F. Gordijn, E. Jürgens, J. Koopman, A. Ouwerkerk, S. Steen, I. Novalija, J. Brank, D. Mladenic, A. Zidar, A multilingual benchmark to capture olfactory situations over time, in: 3rd Workshop on Computational Approaches to Historical Language Change, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 1–10. URL: https://aclanthology.org/2022.lchange-1.1.

[5] M. Zinnen, P. Madhu, R. Kosti, P. Bell, A. Maier, V. Christlein, ODOR: The ICPR2022 ODeuropa Challenge on Olfactory Object Recognition, in: 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 4989–4994.

[6] M. Zinnen, P. Madhu, R. Kosti, P. Bell, A. Maier, V. Christlein, Odeuropa dataset of smell-related objects, 2022. URL: https://doi.org/10.5281/zenodo.6367776.

[7] K. Akdemir, A. Hürriyetoğlu, R. Troncy, T. Paccosi, S. Menini, M. Zinnen, V. Christlein, Multimodal and Multilingual Understanding of Smells using VilBERT and mUNITER, in: MediaEval Benchmarking Initiative for Multimedia Evaluation, 2022.

[8] Y. Shao, Y. Zhang, W. Wan, J. Li, J. Sun, Multilingual Text-Image Olfactory Object Matching Based on Object Detection, in: MediaEval Benchmarking Initiative for Multimedia Evaluation, 2022.