Data Mining and Knowledge Graphs as a Backbone for Advanced Olfactory Experiences

PASQUALE LISENA, EURECOM, France
MARIEKE VAN ERP, KNAW Humanities Cluster, the Netherlands
CECILIA BEMBIBRE, University College London, United Kingdom
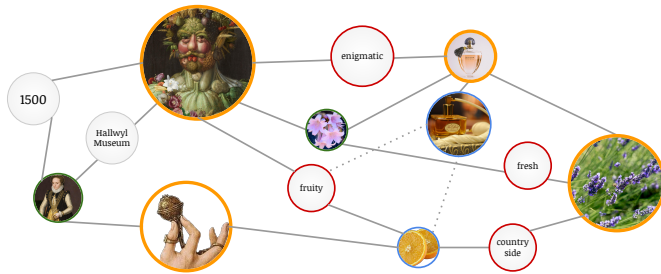INGER LEEMANS, KNAW Humanities Cluster, the Netherlands

Fig. 1. A graph of olfactory data

Our senses are the gateways to our memories and emotions and can provide rich experiences when used in museum or retail settings. However, information about sensory experiences is often dispersed in different data sources, or not yet machine processable. In this position paper, we present our vision on mining olfactory data from text and integrating it with structured olfactory sources (e.g. chemistry databases, industry odour descriptors) to advance our understanding of how smells are represented in texts and structured data, as well as to leverage this information to create more advanced multisensory experiences. In particular, we propose Knowledge Graph technologies for enabling data re-using, enrichment with background information, and AI applications. In addition, we review the challenges of curating, interpreting and presenting smells in instances where the source or/and the original situational context are not present and discuss the relevance of a scent significance preservation framework.

CCS Concepts: • **Computing methodologies** → **Object detection**; **Information extraction**; **Semantic networks**; **Ontology engineering**.

Additional Key Words and Phrases: olfactory mining, olfactory representation, semantic web, natural language processing, computer vision, knowledge engineering, olfactory heritage

1 INTRODUCTION

Our vocabulary for smells, which we use to name, interpret and communicate olfactory experiences, is often contained within different expert groups. For example, perfume makers deal with the molecular and the sensory aspect of fragrance and ingredients, and have developed a vocabulary to discuss and evaluate their creations [17]; wine sommeliers and coffee tasters use domain-specific words to communicate their assessments [6], chemists, including those working in close connection with historic artefacts and buildings, focus on the volatile organic compounds (VOC) responsible for odours and study how these relate to aspects of sensory perception [3]. In addition, the study of olfactory experiences, especially historical ones, involves interdisciplinary approaches to understand, contextualise and convey meaning. Sensory historians, art historians with knowledge of olfactory art, psychologists, and linguists studying connections between smell and memory, emotions, behaviour and other aspects connected to human olfaction contribute to our understanding of smells. This knowledge is often domain-specific and not easily accessible.

As part of the European research project *Odeuropa*,[1] our goal is to bring together olfactory information from texts, images, olfactory databases, and chemical analyses to (re)enable people, from the general public as well as olfactory experts, to access, understand, and interpret olfactory knowledge through the use of Artificial Intelligence (AI) techniques. We focus in particular on historical scent mining from Europe from the 17th to early 20th century but aim to make our approaches generalisable to other locales and time periods. In this position paper, we explain our general goal, as well as the three sub-problems we see and our proposed solutions.

2 OBSTACLES TOWARDS OLFACTORY KNOWLEDGE SHARING

The task we set ourselves is basically to capture smells, something profoundly non-digital, in a digital form, such that the knowledge can be shared, analysed, and compared.

2.1 Terminology

Whereas experts have created vocabularies within their domains to exchange olfactory information, this jargon is often specific to their domain and does no aid in discussing olfactory elements across domains [6]. To communicate information beyond any one domain, a way to harmonise these vocabularies is needed, with explicit explanations of the terminology used.

2.2 Finding scent references

In the past, people engaged with 'smell' communally and privately in their daily lives, and the sense of smell was acknowledged as an instrument of knowledge, for instance in medicine and urban planning [15]. Today, this shared olfactory knowledge and these experiences have been largely forgotten, although there is a rekindled interest that involves studying the varieties of sensory experience, including smells, across and within cultures and understanding their impact [12]. Efforts are being made to recognise and safeguard sensory knowledge associated with specific places or cultures as part of the heritage, documenting smells as part of the process [10, 16].

A quick analysis shows that our lexicon for smells was more elaborate in the past, evidenced by the number of entries in the Historical Thesaurus of English (90 terms are listed at the top level of the category 'Smell/odour' with various subcategories) versus the number of entries in the current Cambridge Dictionary of English (21). Whilst the natural language processing community has developed tools to recognise individuals' names and emotions amongst other things in a multitude of languages, tools to automatically recognise references to scents have not received much attention, which is why we set out to create these ourselves [4].

---

[1]https://www.odeuropa.eu

The field of computer vision received a big impulse with the development of ImageNet [7]. But its use for art/historic images is limited, as transformations from contemporary photos to drawings and painting need to be made, and contemporary photos do not contain as many still lives, snuff boxes and pomanders to create reliable algorithms to recognise such objects.

2.3 Interpreting and communicating historical smells

Experiencing what the world smelled like in the past enriches our knowledge of it [2] and allows us to develop an emotional connection to history [11]. However, once we have identified relevant smells from history and developed methods that allow for them to be recorded, there are scientific, curatorial and interpretative considerations associated with bringing back historical smells in a way that makes sense to contemporary audiences. We need to think about issues such as how the scent was perceived originally (and how it can be perceived by our noses today) and its connection to the source and practices, places, people, or artefacts that made it meaningful. Conveying not only the character of the scent but also its significance is one of the challenges we face when translating recorded information into meaningful sensory experiences. Disclosing the methodology used to reconstruct or recreate historic smells, establishing its identity via authorship are steps towards a framework to preserve and communicate significance, avoiding potential feelings in the audience of being manipulated [9]. Finally, interpreting a historical smell in a new context to a public perhaps unfamiliar with the odour requires identifying and developing relevant storylines that contextualise the value of the scent, and finding relevant channels to communicate it to museum audiences [18].

3 TOWARDS AN OLFACTORY KNOWLEDGE GRAPH

The above mentioned challenges in referring and describing smells call for new ways to represent the olfactory information and experience. Previous work has shown that Knowledge Graphs (KG) are suitable to represent and exploit the domain information in fields such as cultural heritage [5, 13], history [14], and art [1, 8]. The graph paradigm ensures at the same time explicit semantics[2] and a high expressiveness granted by a dynamic data structure. Knowledge Graphs, in combination with other Semantic Web technologies, provide methods and standards for integrating different data sources, promoting the re-use of resources through shared identifiers (URIs) and the interlinking with other datasets in the Web of Data. In this way, information about the chemical composition of a smell, the structured description of its perception by a person, data about the odorant object or place, and the contextual spatio-temporal information can cohabit in the same graph. In addition, it enables advanced use of domain knowledge, e.g. in smart search, reasoning and AI applications.

For this reason, we envision the realisation of a large KG collecting olfactory-related data from different sources. To do so, the design of a suitable representation model is required, relying upon and extending existing ontologies and introducing classes and relations specific to olfactory and perception in general. We envision this model use specialised vocabularies in machine-readable formats, in which each term can group cross-language synonyms and has explicit logical relationships with other terms – e.g. a *neroli scent* can be part of a *fruity scent* family, and was historically believed to help *combat the plague*. The vocabularies can be derived by the existing thesauri mentioned in Section 2.1, as well as be automatically extracted from textual corpora. Finally, proper tools are required for populating this graph from textual and visual sources and for interlinking it with the Web of Data.

Such a KG would support the digital preservation of our olfactory heritage, organise the sparse information, capture the involved logical relationships, and enable their access, search, and reuse. In addition, the KG can provide specialised knowledge to feed AI applications, leveraging the data

---

[2]Classes and relations in a KG are formally defined and conceptualised in machine-readable schemas called *ontologies*.

to automatically group similar odours, infer the dominant olfactory features of an era or a place, answer smell-related questions, improve the search of olfactory references in texts and images.

4 CONCLUDING REMARKS

Looking to historical text and images to extract olfactory references is a way of capturing sensory knowledge that would otherwise be either forgotten or limited to isolated domains of study. KGs are a suitable way of integrating smell-related data from different sources, which, once interpreted and contextualised to convey key meanings via interdisciplinary collaborative research, can enable contemporary audiences to access the fragrant past.

ACKNOWLEDGMENTS

REFERENCES

[1] Manel Achichi, Pasquale Lisena, Konstantin Todorov, Raphaël Troncy, and Jean Delahousse. 2018. DOREMUS: A Graph of Linked Musical Works. In *The Semantic Web – ISWC 2018*. Springer International Publishing, Cham, 3–19.

[2] John P. Aggleton and Louise Waskett. 1999. The ability of odours to serve as state-dependent cues for real-world memories: Can Viking smells aid the recall of Viking experiences? *British Journal of Psychology* 90, 1 (1999), 1–7.

[3] Cecilia Bembibre and Matija Strlič. 2017. Smell of heritage: a framework for the identification, analysis and archival of historic odours. *Heritage Science* 5, 2 (2017), 1–11. Number: 2.

[4] Ryan Brate, Paul Groth, and Marieke van Erp. 2020. Towards Olfactory Information Extraction from Text: A Case Study on Detecting Smell Experiences in Novels. In *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. International Committee on Computational Linguistics, Online, 147–155. https://www.aclweb.org/anthology/2020.latechclfl-1.18

[5] Valentina Anita Carriero, Aldo Gangemi, Maria Letizia Mancinelli, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. 2019. ArCo: The Italian Cultural Heritage Knowledge Graph. In *The Semantic Web – ISWC 2019*. Springer International Publishing, Cham, 36–52.

[6] Ilja Croijmans and Asifa Majid. 2016. Not all flavor expertise is equal: The language of wine and coffee experts. *PLoS One* 11, 6 (2016), e0155845.

[7] Jia Deng, R. Socher, Li Fei-Fei, Wei Dong, Kai Li, and Li-Jia Li. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New Jersey, NY, USA, 248–255.

[8] Chris Dijkshoorn, Lizzy Jongma, Lora Aroyo, Jacco van Ossenbruggen, Guus Schreiber, Wesley ter Weele, and Jan Wielemaker. 2018. The Rijksmuseum collection as Linked Data. *Semantic Web Journal* 9 (2018), 221–230. 2.

[9] Jim Drobnick. 2014. The museum as smellscape. In *The Multisensory Museum: Cross-Disciplinary Perspectives on Touch, Sound, Smell, Memory, and Space*. Vol. 17. Rowman & Littlefield, Lanham, MD, USA, Chapter 12, 177–196.

[10] Jack Guy. 2021. France has passed a law protecting the sounds and smells of the countryside. *CNN* (22 01 2021). https://www.cnn.com/travel/article/france-rural-noise-law-scli-intl/index.html

[11] Rachel S. Herz, Corrente Schankler, and Sophia Beland. 2004. Olfaction, Emotion and Associative Learning: Effects on Motivated Behavior. *Motivation and Emotion* 28, 4 (01 Dec 2004), 363–383.

[12] David Howes. 2006. Charting the Sensorial Revolution. *The Senses and Society* 1, 1 (March 2006), 113–128.

[13] Antoine Isaac and Bernhard Haslhofer. 2013. Europeana Linked Open Data – data.europeana.eu. *Semantic Web Journal* 4 (2013), 291–297. 3.

[14] Mikko Koho, Esko Ikkala, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Eero Hyvönen. 2021. WarSampo knowledge graph: Finland in the Second World War as Linked Open Data. *Semantic Web Journal* 12 (2021), 265–278. 2.

[15] William Tullett. 2019. *Smell in Eighteenth-Century England: A Social Sense*. Oxford University Press, Oxford, UK.

[16] UNESCO. 2018. The skills related to perfume in Pays de Grasse: the cultivation of perfume plants, the knowledge and processing of natural raw materials, and the art of perfume composition. https://ich.unesco.org/en/lists

[17] Charles S Van Toller and George H Dodd. 1993. *Fragance: The Psychology and Biology of Perfume*. Vol. 2. Springer Science & Business Media, Berlin, Germany.

[18] Caro Verbeek. 2016. Presenting Volatile Heritage: Two Case Studies on Olfactory Reconstructions in the Museum. *Future Anterior* 13, 2 (2016), 33–42.