

A Smell is Worth a Thousand Words: Olfactory Information Extraction and Semantic Processing in a Multilingual Perspective

Sara Tonelli 

Digital Humanities research Unit, Fondazione Bruno Kessler, Trento, Italy

Abstract

More than any other sense, smell is linked directly to our emotions and our memories. However, smells are intangible and very difficult to preserve, making it hard to effectively identify, consolidate, and promote the wide-ranging role scents and smelling have in our cultural heritage. While some novel approaches have been recently proposed to monitor so-called urban smellscape and analyse the olfactory dimension of our environments (Quercia et al., [1]), when it comes to smellscape from the past little research has been done to keep track of how places, events and people have been described from an olfactory perspective. Fortunately, some key prerequisites for addressing this problem are now in place. In recent years, European cultural heritage institutions have invested heavily in large-scale digitisation: we hold a wealth of object, text and image data which can now be analysed using artificial intelligence. What remains missing is a methodology for the extraction of scent-related information from large amounts of texts, as well as a broader awareness of the wealth of historical olfactory descriptions, experiences and memories contained within the heritage datasets. In this talk, I will describe ongoing activities towards this goal, focused on text mining and semantic processing of olfactory information. I will present the general framework designed to annotate smell events in documents, and some preliminary results on information extraction approaches in a multilingual scenario. I will discuss the main findings and the challenges related to modelling textual descriptions of smells, including the metaphorical use of smell-related terms and the well-known limitations of smell vocabulary in European languages compared to other senses.

2012 ACM Subject Classification Applied computing → Document analysis; Information systems → Digital libraries and archives

Keywords and phrases olfactory information extraction, smellscape, multilingual annotation

Digital Object Identifier 10.4230/OASICS.LDK.2021.2

Category Invited Talk

References

- 1 Daniele Quercia, Rossano Schifanella, Luca Maria Aiello, and Kate McLean. Smelly maps: the digital life of urban smellscape. In *Ninth International AAAI Conference on Web and Social Media*, 2015.



© Sara Tonelli;

licensed under Creative Commons License CC-BY 4.0

3rd Conference on Language, Data and Knowledge (LDK 2021).

Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch; Article No. 2; pp. 2:1–2:1

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany