OBSERVING ODOR-RELATED INFORMATION IN ACADEMIC DOMAIN

Inna Novalija
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia
inna.koval@ijs.si

Dunja Mladenić
Jožef Stefan Institute and Jožef Stefan
International Postgraduate School
Jamova cesta 39, Ljubljana, Slovenia
dunja.mladenic@ijs.si

Daniel Schwabe
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia
daniel.schwabe@ijs.si

M. Besher Massri
Jožef Stefan Institute and Jožef Stefan
International Postgraduate School
Jamova cesta 39, Ljubljana, Slovenia
besher.massri@ijs.si

Marko Grobelnik
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia
marko.grobelnik@ijs.si

Janez Brank
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia
janez.brank@ijs.si

ABSTRACT

In this paper, we demonstrate an approach for observing olfactory related information in an academic publications environment (such as Microsoft Academic Graph) based on semantic technologies. We present an Odor Observatory tool that enables several usage scenarios, such as observing odor-related papers and topics, viewing institutions conducting olfactory research, defining top journals and key countries in the olfactory domain.

Validation of the proposed approach on a collection of academic publications from 1800 until 1925 confirms applicability of the proposed approach on large data collections with a wide span of time. In usage scenarios we observed the odor-related publications in Microsoft Academic Graph by topic, discovered the journals with historical olfactory publications and found that the most popular terms in odor-related research content are: method, olfactory, odor, device, invention, smell, preparation, utility model.

KEYWORDS

Odor, Olfactory information, Microsoft Academic Graph (MAG), Data mining.

1. INTRODUCTION

Olfaction, or the sense of smell, is the sense through which smells (or odors) are perceived [1]. Olfactory science involves studying olfaction and odor-related topics, the sensory system, physiology, and pheromone signals.

The Odeuropa project [2] gathers and integrates expertise in sensory mining and olfactory heritage. The project partners are developing novel methods to collect information about smell from (digital) text and image collections.

The Odeuropa project partners apply state-of-the-art AI techniques to text and image datasets in order to identify and trace how 'smell' was expressed in different languages, with what places it was associated, what kinds of events and practices it characterized, and to what emotions it was linked.

In this paper we present an approach for mining olfactory information from scientific research collections, such as the Microsoft Academic Graph (MAG) [3].

The olfactory mining approach combines data processing, modelling and visualization methods in order to develop applicable tools for data analysis.

We present an Odor Observatory tool [4] targeted at several visualization scenarios. In particular, the Odor Observatory allows exploring olfactory related papers from the MAG over time, and along with current data, provides historical information starting with the early XIX century.

The data-driven functionalities of Odor Observatory are:

- Possibility of exploring top ranked topics in the olfactory academic domain;
- Possibility of exploring top ranked institutions conducting olfactory research;
- Possibility of exploring key countries and defining top ranking journals in the olfactory academic domain;
- Odor-related search functionalities;
- Word cloud visualization for odor-related terms.

2. RELATED WORK

Olfactory science covers different aspects of research related to odors, therefore exploring odor related information and data can be viewed as complex multidisciplinary area.

Lötsch et al. [5] considered machine learning approaches for human olfactory research. The authors state that the complexity of the human sense of smell is reflected in complex and high-dimensional data, which supports the applicability of machine learning and data mining techniques. The use of machine learning in human olfactory research includes the following aims:

1. The study of the physiology of pattern-based odor detection and recognition processes;
2. Pattern recognition in olfactory phenotypes;
3. The development of complex disease biomarkers including olfactory features;

4. Odor prediction from physico-chemical properties of volatile molecules, and
5. Knowledge discovery in publicly available large databases.

The authors provide review of key concepts of machine learning and summarizes current applications on human olfactory data.

At the same time, linguistic and semantic communities focused on studying the language of smell [6]. Iatropoulos et al. developed a computational method to characterize the olfaction-related semantic content of words in a large text corpus of internet sites in English. They also introduced novel metrics, such as olfactory association index (OAI) and olfactory specificity index (OSI).

Tonelli [7] describes olfactory information extraction and semantic processing from a multilingual perspective. The author states that in several studies it was found that languages seem to have a smaller vocabulary to describe smells as compared to other senses.

In our work we apply data mining and machine learning, as well as semantic approaches for enriching textual data. We use data from Microsoft Academic Graph and our methodologies can be regarded as being in the context of semantic and text processing research. Our approaches can cover cross-lingual and multilingual data and allow for tracking olfactory trends in time.

3. PROBLEM DEFINITION

3.1 DATA SOURCES

The Microsoft Academic Graph (MAG) [3] is a heterogeneous graph containing scientific publication records, citation relationships between those publications, as well as authors, institutions, journals, conferences, and fields of study.

Since this research is conducted in line with the Odeuropa project (targeted at olfactory heritage), the time frame used for MAG data is set to range from the early publications in the 19th century to the present time. The Odeuropa project is interested in particular in the data available up to 1925. Though the project is focused on the historical datasets, the developed Odor Observatory tool allows users to explore recent olfactory publications as well. The dataset is updated on a monthly basis and new available data is uploaded into the observatory.
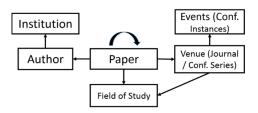

Figure 1: The Conceptual Schema for MAG

The Microsoft Academic Graph data schema is based on the list of following entity types: publication, author, author affiliation (institution), publication venue (journals and conferences), field of study (topic). It contains information about publication dates, as well as citation pairs and co-authorship data (see Figure 1).

Figure 2 illustrates an entry in MAG for a historical publication tagged with several odor-relevant topics.

Figure 2 illustrates an entry in MAG for a historical publication tagged with several odor-relevant topics.
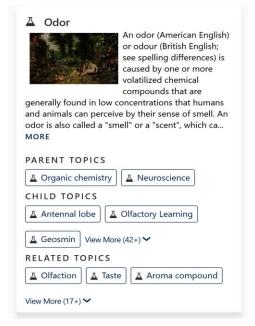

Figure 2: Publication in MAG

Figure 3: Odor in the MAG Taxonomy

Figure 3 shows a representation of Odor in the MAG taxonomy, with parent topics (Organic chemistry and Neuroscience) and child topics (Olfactory learning, Geosmin etc.)

An important functionality while exploring the literature is the ability to expand searches by looking at related topics to a topic of interest. Figure 4 displays topics about/related to Olfaction/Odor/Smell in MAG taxonomy.

RELATED TOPICS

Olfaction | Taste | Aroma compound
Flavor | Pheromone | Essential oil
Organic compound | Bibcode
Olfactory receptor | Alcohol | Insect
Olfactory system | Solvent
Ammonia | Carbon dioxide | Bacteria
Olfactory bulb | Hydrogen sulfide
Major histocompatibility complex
Odor detection threshold | View Less ∧

Figure 4: Odor-related Topics in MAG

3.2 METHODOLOGY

The methodology for observing olfactory related information from academic publication resources includes a number of steps:

- Using the MAG taxonomy, obtain the list of research papers that corresponds to odor-related topics. Papers were filtered to those containing the topics: Olfaction, Odor, Fragrance, Fragrance ingredient, as well as the "smell" keyword;
- Ingest the extracted corpus into the Elastic Search tool[1];
- Provide visualization functionalities, such as MAG time series per term.

The key challenges of the development techniques include:

- Interpretability and explainability of the results – the aim is for the visualizations to be able easily interpretable by humans;
- Given the large scale of the incoming data streams, it is essential that building visualizations are scalable. The MAG contains more than 265 million records (August 2021), including several types of publication, such as journal articles, conference papers, books, book chapters, and papers from other repositories. In addition, MAG also indexes a large corpus of patents.

3.3 USAGE SCENARIOS

We present a couple of usage scenarios for the Odor Observatory tool, cast as questions asked by scholars studying the field.

1. What are the historical trends in odor-related publications?

Figure 5 shows the number of odor-related historical publications in MAG over time. This scenario assumes observing trends in different olfactory topics throughout a time interval.

It is possible to observe that the highest number of publications are in the domains of biology and psychology.
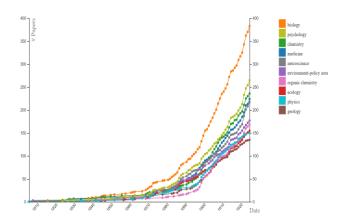

Figure 5: Odor-related Publications in MAG (from year 1800 until year 1925, cumulative) by topic

2. What are the most popular terms used in odor-related publications?

This use case helps the user to visualize term usage by displaying a word cloud with the most popular olfactory terms used in the publications in the period of interest (see Figure 6).
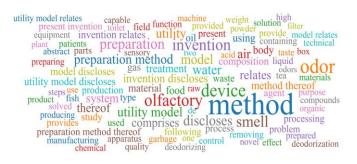

Figure 6: Odor Terms Word Cloud in MAG

3. Which venues were mostly used when publishing odor-related research articles?

This use case shows a number of journals that had historical publications about smells (see Figure 7).

The figure shows that JAMA and Nature journals are the most popular journals regarding historical olfactory publications.
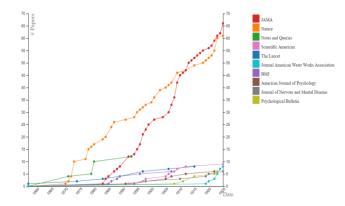

Figure 7: Journals with Olfactory Publications in MAG (from year 1800 until year 1925, cumulative)

4. Which are the publications about smell (from a contextual point of view)?

The Research Explorer tool is a search engine that enables exploring the individual articles in the corpus of odor-related publications.
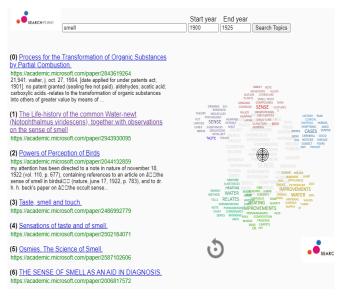

Figure 8: List of Olfactory Publications in MAG (from year 1800 until year 1925) that contains the keyword "smell"

The tool is built on Elastic Search and provides search by keyword and by date. It also supports smart navigation through the results by clustering the results and re-ranking the results by moving the focus of search through the cluster space (see Figure 8). The goal of the tool is to enhance a search engine by providing the users multiple rankings of the results for each query. It is achieved by generating topics for the given query and its result set, and visualizing these topics on the "Ranking Space" panel. When the focus is set near a given topic, results that are on or closer to that topic are ranked higher.

The figure shows a ranked list of relevant publications on the topic of "smells", in the period from 1900 to 1925. The list is modified by changing the context on the right side - the focus is changed by placing the cursor over a cluster, and publications associated with this cluster are displayed.

4. CONCLUSION

In this paper we demonstrated an approach towards observing olfactory related information in scientific publications, as recorded in the MAG.

In addition, we present an Odor Observatory tool that enables several usage scenarios for exploring historical and present olfactory research.

The future work will include the exploration of other textual datasets applicable for olfactory research, with an accent on olfactory heritage information.

In line with the Odeuropa project, the relevant information extracted from textual sources will be, following semantic web standards, aligned with the 'European Olfactory Knowledge Graph' (EOKG).

5. ACKNOWLEDGMENTS

REFERENCES

[1] Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R., Lederman, S. J., & Merfeld, D. M. (2012). Sensation & perception (3rd ed.). Sinauer Associates.

[2] Odeuropa project, https://odeuropa.eu (accessed in August, 2021).

[3] Wang K. et al. A Review of Microsoft Academic Services for Science of Science Studies, Frontiers in Big Data, 2019, doi: 10.3389/FDATA.2019.00045.

[4] JSI Odor Observatory, public service, https://odeuropa.ijs.si/dashboards/Main/Index?visualization= visualizations-MAG--top-topics# (accessed in August, 2021).

[5] Lötsch, J., Kringel, D., Hummel, T. Machine Learning in Human Olfactory Research, Chemical Senses, Volume 44, Issue 1, January 2019, Pages 11–22, https://doi.org/10.1093/chemse/bjy067.

[6] Iatropoulos, G., Herman, P., Lansner, A., Karlgren, J., Larsson, M., Olofsson, JK. The language of smell: Connecting linguistic and psychophysical properties of odor descriptors. Cognition. 2018 Sep;178:37-49. doi: 10.1016/j.cognition.2018.05.007. Epub 2018 May 12. PMID: 2976379.

[7] Tonelli, S. A Smell is Worth a Thousand Words: Olfactory Information Extraction and Semantic Processing in a Multilingual Perspective. doi: https://doi.org/10.4230/OASIcs.LDK.2021.2 https://drops.dagstuhl.de/opus/volltexte/2021/14538/pdf/OASIcs-LDK-2021-2.pdf (accessed in August, 2021).