



# Odeuropa

## Wikifier and Smell Tracker

Inna Novalija, Dunja Mladenčić, Janez Brank

November 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004469.







# JSI Wikifier: Introduction

**Tool for Wikification** of textual documents in 100 languages

## Functionalities:

- Wikification
- Part-of-speech tagging
- Cosine similarity measure
- Neighborhood subgraph
- Concept information
- REST API

The screenshot shows the Wikifier web interface. At the top, there is a dark blue header with the text "Wikifier" and "Semantic Annotation Service for 100 Languages". Below the header, there are links for "Documentation", "About", and "Register". The main interface is light blue and contains a "Language:" dropdown menu set to "Autodetect" and a "Load sample document" button. Below this is another dropdown menu labeled "Where available, also show concept names in:" set to "English". A text input area is labeled "Enter or paste your document below:" with a "Submit" button. The text area contains a sample paragraph about St Paul's Cathedral. At the bottom, there are "Advanced options:" with a text input field containing "(for testing purposes)", and a "Host:" dropdown set to "posta" and a "Port:" dropdown set to "8095", both with "Submit" buttons.



# JSI Wikifier: Demo



<https://wikifier.org>

Wikifier results

Show less  Show more

🔊 Language autodetected as English (en; #0). [Show details...](#)

### Text

Visitors to [St Paul's Cathedral Dean and Chapter Library](#), a eighteenth-century space designed by [Christopher Wren](#), often register their appreciation of the aroma of the space in the guest book: "City guide trainees we all loved the smell and beautiful [library](#)"; "Amazing place! I can inhale the [knowledge](#)"; "We can smell the [history](#), the fragrance of [heritage](#) and our [communion](#) with souls of the past". In this case, in addition to the explicit communal value, the smell becomes [culturally significant](#) by its association with a [heritage](#) space and its collection.

### Annotations

PR	Annotation	Annotation (en)	
0.0466	<a href="#">Christopher Wren</a> <span style="font-size: 0.8em;">w d</span>	<a href="#">Christopher Wren</a>	>>
0.0171	<a href="#">St Paul's Cathedral</a> <span style="font-size: 0.8em;">w d</span>	<a href="#">St Paul's Cathedral</a>	>>
0.0128	<a href="#">Eucharist</a> <span style="font-size: 0.8em;">w d</span>	<a href="#">Eucharist</a>	>>
0.0121	<a href="#">Saint</a> <span style="font-size: 0.8em;">w d</span>	<a href="#">Saint</a>	>>
0.0104	<a href="#">Dean (Christianity)</a> <span style="font-size: 0.8em;">w d</span>	<a href="#">Dean (Christianity)</a>	>>
0.0104	<a href="#">Library</a> <span style="font-size: 0.8em;">d</span>	<a href="#">Library</a>	>>
0.0104	<a href="#">History</a> <span style="font-size: 0.8em;">d</span>	<a href="#">History</a>	>>
0.0101	<a href="#">Culture</a> <span style="font-size: 0.8em;">d</span>	<a href="#">Culture</a>	>>
0.0099	<a href="#">Cultural heritage</a> <span style="font-size: 0.8em;">d</span>	<a href="#">Cultural heritage</a>	>>
0.0098	<a href="#">Soul</a> <span style="font-size: 0.8em;">d</span>	<a href="#">Soul</a>	>>
0.0094	<a href="#">Chapter (religion)</a> <span style="font-size: 0.8em;">d</span>	<a href="#">Chapter (religion)</a>	>>

### Support

**PR** **P(|p)** **Index** **Phrase**

Select an annotation to show which phrases in the text support it.

### Link Targets

In the Wikipedia, where do links with a particular anchor text point to?

Anchor text: **Christopher Wren** ([Where does it occur?](#))  
P(link | phrase) = 0.6834

PR	PR (lin.)	Score	Cosine	Target
0.0466	1.0000	0.0466	0.2483	<a href="#">Christopher Wren</a>
0.0034	0.6852		0.1941	<a href="#">Christopher Wren (priest)</a>
0.0034	0.6820		0.1536	<a href="#">Christopher Wren the Younger</a>

Anchor text: **Wren** ([Where does it occur?](#))  
P(link | phrase) = 0.0395

**Note:** 8 candidates were ignored because they were linked to less than 2 times (the minLinkFrequency parameter) with this phrase as the anchor text.

**Note:** 1 candidate was ignored because of their WikiData class memberships.

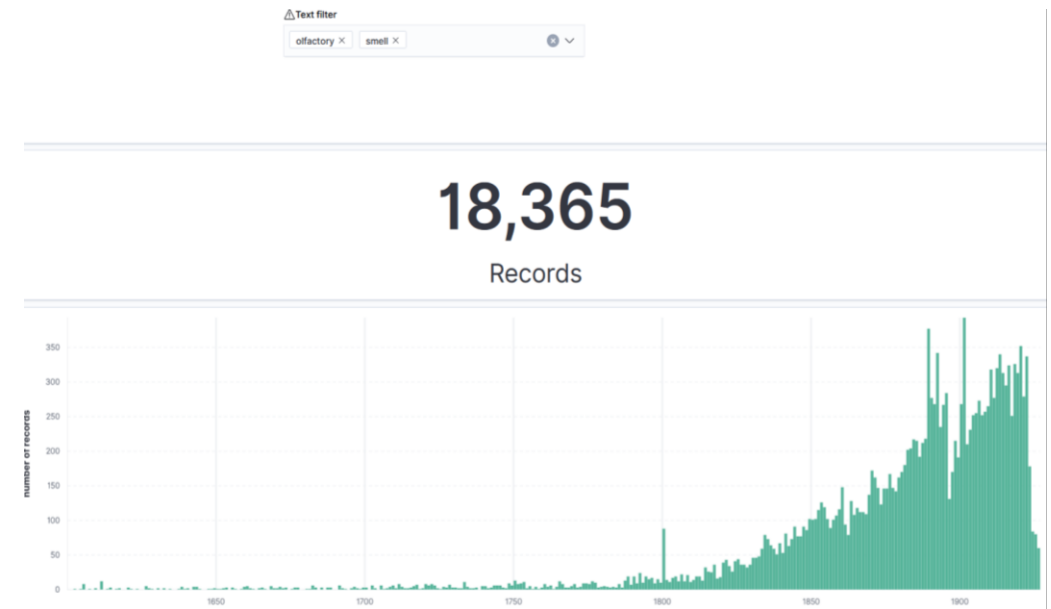
PR	PR (lin.)	Score	Cosine	Target
0.0466	1.0000	0.0466	0.2483	<a href="#">Christopher Wren</a>
0.0047	0.7934		0.1127	<a href="#">Women's Royal Naval Service</a>
0.0021	0.5311		0.1422	<a href="#">Eurasian wren</a>
0.0019	0.5016		0.0505	<a href="#">Wren</a>
0.0018	0.4820		0.1135	<a href="#">Wren, Ohio</a>
0.0014	0.4328		0.1450	<a href="#">Wren, Oregon</a>



# Data Exploration with Smell Tracker



- **The Smell Tracker** tool helps users to find information about different aspects related to smells coming from research, news, and multilingual historical text datasets.
- The tool consists of several sections, each responsible for a **data source**. For each data source, there is a landing page explains more in details about each data source. Smell Tracker specifically covers **Slovenian data sources**.



Gutenberg collection analysis, filters:  
“Olfactory”, “Smell”

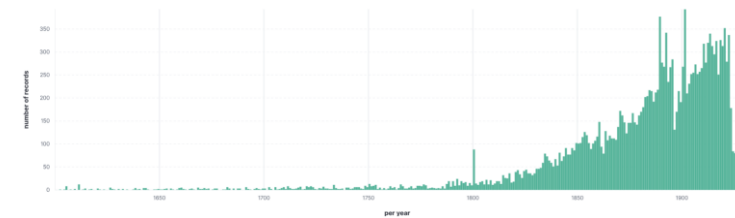
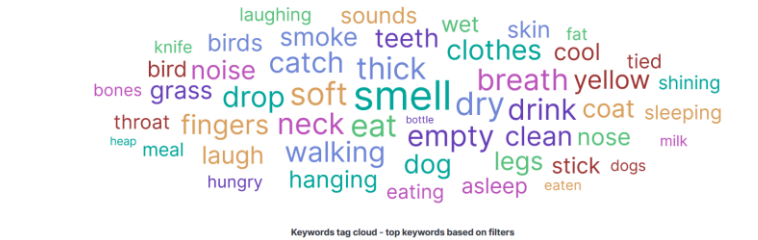
# Smell Tracker: Selected Dashboards



- **Live news** (recent news about the Odeuropa project in live mode)
- **Gutenberg dashboard** (visualizations and analysis of text corpus from Gutenberg project)
- **dLib dashboard** (visualizations and analysis of Slovenian text corpus)
- **Tales dashboard** (visualizations and analysis specifically addressing emotion detection from tales)

# Smell Tracker: Visualizations

- **Concept** tag cloud - top concepts based on filters.
- **Keywords** tag cloud - top keywords based on filters
- **Timeline**
- **Filtering** functionalities



△ Concept  
Select...  
Apply changes Cancel changes Clear form

△ Text filter  
olfactory x smell x

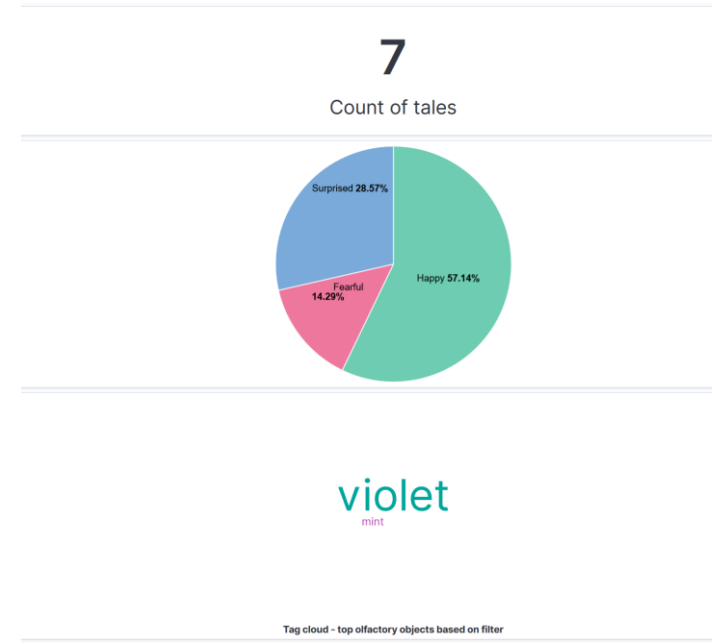
Gutenberg collection analysis, filters:  
“Olfactory”, “Smell”



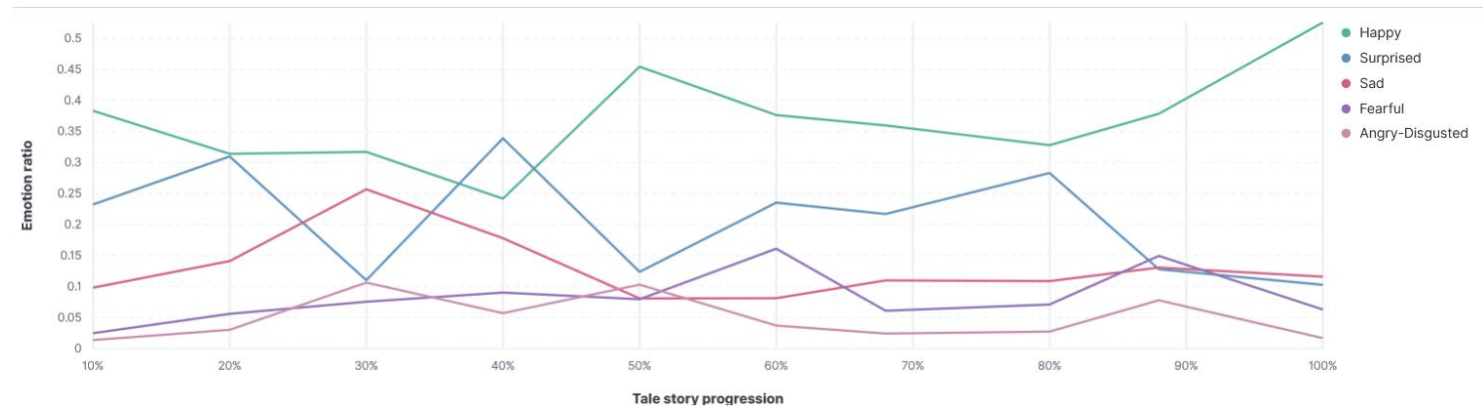
# Smell Tracker: Emotions



- **Top emotions** pie chart
- **Tag cloud** - top **olfactory objects** based on filter
- Tag cloud - top **concept** based on filter
- Tag cloud - top **significant words** based on filter
- **Tale story** progression
- **Dominant emotions** tales



Tales dashboard, filter: "Violet"







# Questions?

