

Reproducibility Review of: Mining Meaningful Facets in Spatial



Information Retrieval with Spatial Relevance Feedback

Item	Value
Title	Mining Meaningful Facets in Spatial Information Retrieval with Spatial Relevance Feedback
Authors	Martin Werner
Ref. paper	Werner, M.: Mining Meaningful Facets in Spatial Information Retrieval with Spatial Relevance Feedback, AGILE GIScience Ser., 6, 12, https://doi.org/10.5194/agile-giss-6-12-2025 , 2025.
Codechecker	Sophie Teichmann
Date of Check	2025-05-19
Summary	Partial reproduction
Repository	https://github.com/reproducible-agile/spatial_facet_search
Ref. certificate	DOI 10.17605/OSF.IO/956A8

Table 1: Reproduction metadata

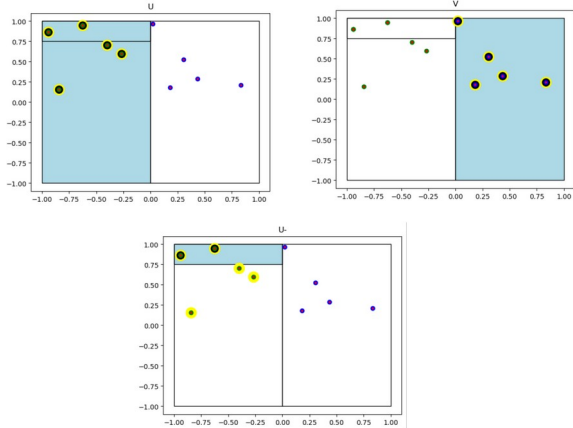
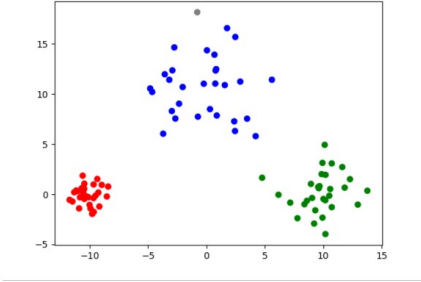
Output	Comment
Figure 1  Figure 2 	Figure 1 Reproduction of the supervised spatial facet search with the notebook 'SimpleExperiments.ipynb' Figure 2 Reproduction of the unsupervised spatial facet search with the notebook 'SimpleExperiments.ipynb'

Table 2: Summary of output files generated

Summary

The author provided clear steps for the partial reproduction. However, due to computational and privacy constraints, only two basic examples could be reproduced. Since those two minimal examples demonstrate both new methods introduced in the paper, I would consider the partial reproduction to be successful.

Reproducibility reviewer notes

The following section contains the notes taking during the review. The author provided a lot of additional information the well documented GitHub page (https://github.com/tum-bgd/spatial_facet_search)

Installation prerequisites and computational environment

The author provides an additionally README for reproduction. There the set-up is made easy with a Docker container and a jupyter note book. Additionally, I tested the Quickstart set-up in the README, which was carried out successfully. The reproduction is carried out on a machine with Ubuntu 24.04.2 LTS.

Data preparation

In the created python environment with the quick start I ran the script 'contrib/simple.py' successfully. This script sets up an easy database. Additionally, I was able to download the OSM dataset. The twitter dataset is not available due to legal and privacy concerns.

Running the code

The main notebook for reproduction is 'SimpleExperiments.ipynb'. There the different sections in the paper are discussed in terms of reproducibility. The individual sections of the paper are discussed in terms of reproducibility. After adjusting the paths, the notebook ran without any problems within the provided docker container.

Outputs and results

Reproduction for the supervised facets (Section 2.2.1) is shown in **Figure 1**. The light blue shaded area shows the area included in the querying. The inner colors depict the colors assigned to the points. The black circle denotes the data included in the supervised facet. The yellow circle denotes the data points selected after augmentation. Even in the case of U- all the fitting data points are found even though not all are in the original area. The approach extrapolates from the existing area. It is a similar question as displayed in **Figure 1(a)** of the original paper.

Figure 1

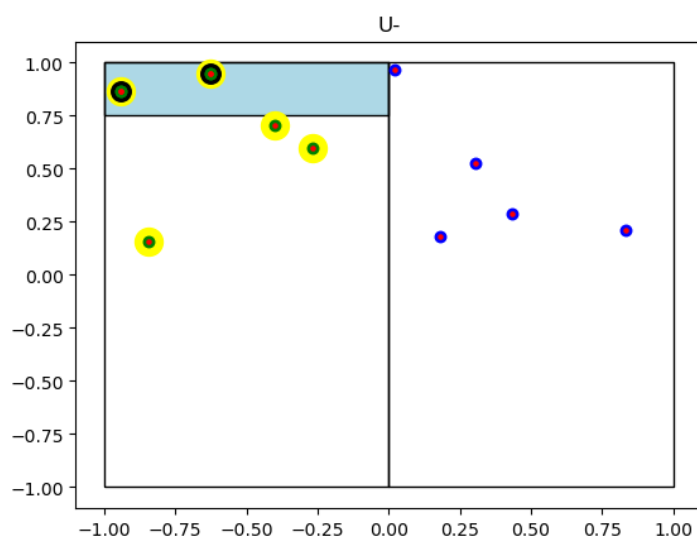
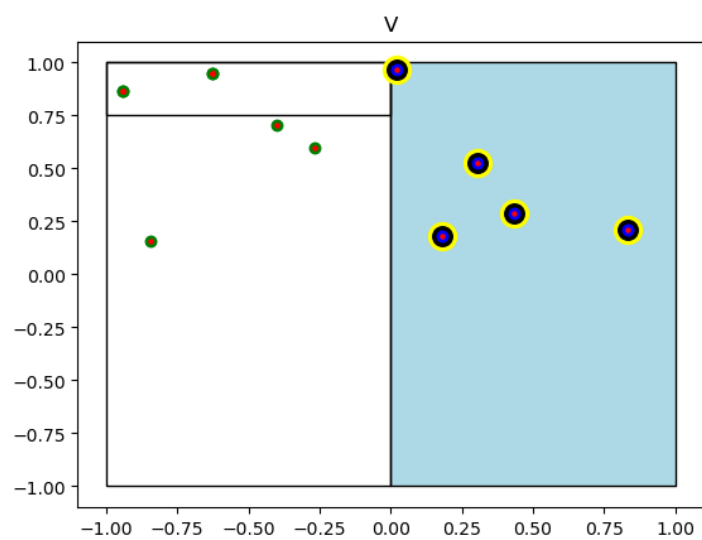
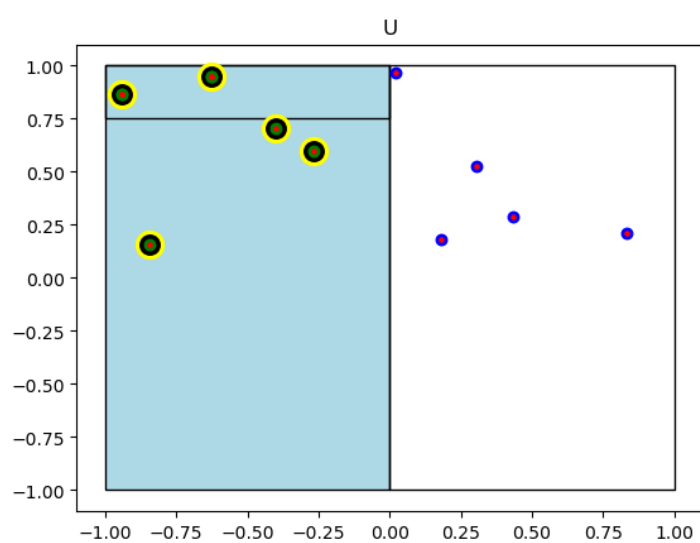
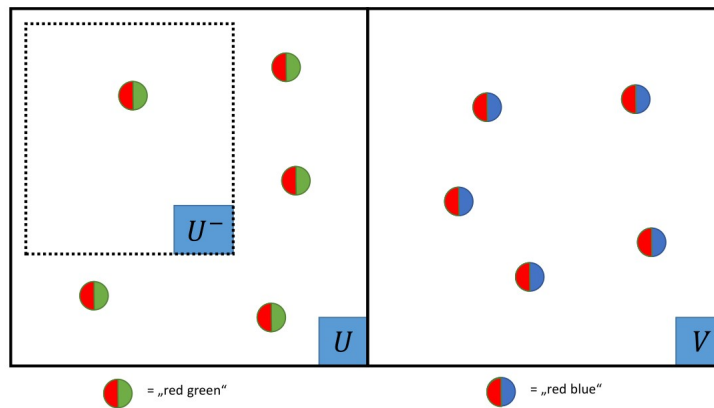


Figure 1(a) (from original paper)



(a) A simple example of a dataset. To the left, within the set U , documents contain the two colors as terms, and similarly on the set V

The reproduction for the unsupervised facets is shown in **Figure 2**. There the approach shall uncover unknown keyword. For that first a database is generated with three normal distributed clusters. Each of the clusters has an individual label and the label black. **Figure 2** illustrates the results of the reconstruction. The additional keywords are identified. The original results are displayed in **Figure 1(b)** of the original paper.

Figure 2

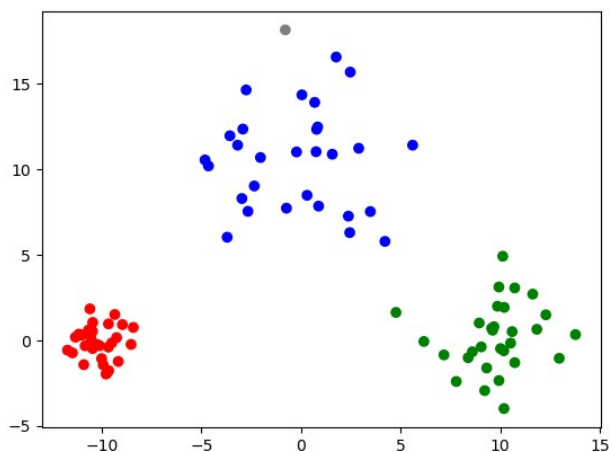
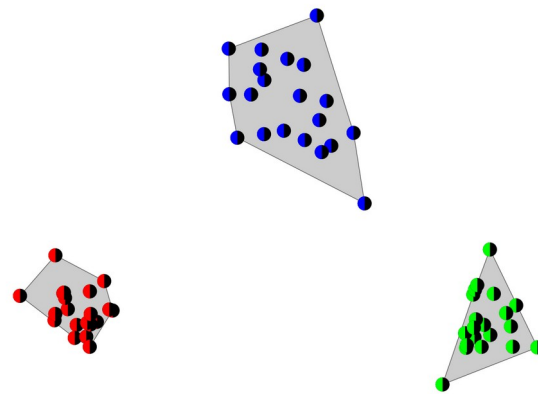


Figure 1(b) (from original paper)



(b) Result of Facet Clustering in the Unsupervised Facet Mining Example

The rest of the paper cannot be easily reproduced because to cite the author 'It is extremely compute and integration-intense to reproduce the given figures. ' and 'This is practically impossible to reproduce outside our infrastructure. ' The steps provided in the notebook seem reasonable. I will not be able to recreate any of the steps. In GitHub folder there are additional scripts to set up databases for ones own data.

Recommendations

I do not have any recommendations for the author.

Acknowledgments

I want to the thank to author for the helpful exchange and well documented code he provided. Additionally, I want to thank the AGILE reproducibility review chair for organizing the reviews and allowing me to participate.

Citing this document

This report is part of the reproducibility review at the AGILE conference. For more information see <https://reproducible-agile.github.io/>. This document is published on <https://osf.io/956a8/> at DOI 10.17605/OSF.IO/956A8 .

To cite the report use:

DOI 10.17605/OSF.IO/956A8