

4 Sample Use-Cases

4.1 Finnish 2011 parliamentary elections

We provide a prepared dataset about the Finnish 2011 parliamentary elections. Get the data (non-geospatial), try out *Siren* and learn about the Finnish political scene! (More details on the main webpage.)

To illustrate the use of *Siren*, we present example use-cases from different application domains.

4.2 Biological niche-finding

One use-case concerns niche-finding, i.e. the problem of finding species' bioclimatic envelope, an important task in biology.

4.3 DBLP Computer Science Bibliography

We also consider non-geospatial data, namely from the DBLP Computer Science Bibliography.

5 Resources

References

- [A] Esther Galbrun and Pauli Miettinen. A Case of Visual and Interactive Data Analysis: Geospatial Redescription Mining. In *Instant Interactive Data Mining Workshop @ ECML-PKDD*. 2012.
- [B] Esther Galbrun and Pauli Miettinen. Siren: an interactive tool for mining and visualizing geospatial redescrptions. In *KDD*, 1544–1547. ACM, 2012.
- [C] Esther Galbrun and Pauli Miettinen. From black and white to full color: extending redescription mining outside the Boolean world. *Statistical Analysis and Data Mining*, 5(4):284–303, 2012.
- [D] Esther Galbrun and Pauli Miettinen. From Black and White to Full Colour: Extending Redescription Mining Outside the Boolean World. In *SDM*, 546–557. 2011.
- [E] Esther Galbrun. Methods for Redescription Mining. PhD thesis, University of Helsinki, 2014. <http://urn.fi/URN:ISBN:978-952-10-9431-6>



Siren Demo at SIGMOD 2014

Release 2.1.1

Esther Galbrun and Pauli Miettinen

June, 2014

This document summarizes information about the demonstration of the Siren interactive interface for redescription mining presented at the 2014 ACM SIGMOD/PODS Conference, Snowbird, Utah, USA on June 22-27, 2014.¹

Esther Galbrun and Pauli Miettinen. Interactive Redescription Mining. In SIGMOD. 2014.

More details can be found on the main Siren [webpage](#)² or in the [user guide](#).³

Redescription mining is a powerful data analysis tool that aims at finding alternative descriptions of the same entities.

For example, in biology, an important task is to identify the bioclimatic constraints that allow some species to survive, that is, to describe geographical regions in terms of both their bioclimatic conditions and the fauna that inhabit them.

Siren is a tool for interactive mining and visualization of redescriptions. It is based on the ReReMi mining algorithm.

¹<http://sigmod2014.org>

²<http://www.cs.helsinki.fi/u/galbrun/redescriptors/siren/main/>

³<http://www.cs.helsinki.fi/u/galbrun/redescriptors/siren/help/>

1 Download

Siren is a multi-platform software. *Siren* and its core mining algorithm *ReReMi* are implemented in Python. It has been used on MacOS and Ubuntu Linux.

Siren and *ReReMi* are licensed under the Apache License, Version 2.0.

- **Source code** (packaged using Python distutils)⁴
- **OS X**⁵
- **Linux** (deb)⁶
- **Windows** (installation executable)⁷

2 Data Formats

In *Siren*, data include:

- **Variables:** The variables describing the entities are divided in two sets. They can be of three types:
 1. Boolean,
 2. categorical,
 3. or real-valued.

Obviously, this is required.

- **Entities names:** Optional additional information, providing names for the entities.
- **Variable names:** Optional additional information, providing names for the variables.
- **Coordinates:** Optional location information, i.e. geographic coordinates of the entities. This makes the data geospatial.

Data can be imported to *Siren* via the interface menu *File* → *Import* → *Import Data*.

Data can be imported into *Siren* as CSV files. The program expects a pair of files, one for either side in character-separated values, as can be imported and exported to and from spreadsheet programmes, for instance.

In particular, the data can be stored as a table with one column for each variable and one row each entity. The first row can contain the names of the variables. The entities names can be included as columns named *ids*. Similarly the coordinates can be included as a pair of columns named *longitudes* and *latitudes*, respectively.

⁴<http://www.cs.helsinki.fi/u/galbrun/redescriptors/code/siren/python-siren-2.1.1.tar.gz>

⁵http://www.cs.helsinki.fi/u/pamietti/data/siren/Siren_2_1_1.dmg

⁶http://www.cs.helsinki.fi/u/galbrun/redescriptors/code/siren/python-siren_2.1.1_all.deb

⁷http://www.cs.helsinki.fi/u/galbrun/redescriptors/code/siren/install_siren_2.1.1.exe

3 Functionalities

Using *Siren*, a user can explore data of his interest by interactively mining, visualizing and editing redescrptions.

3.1 Mining

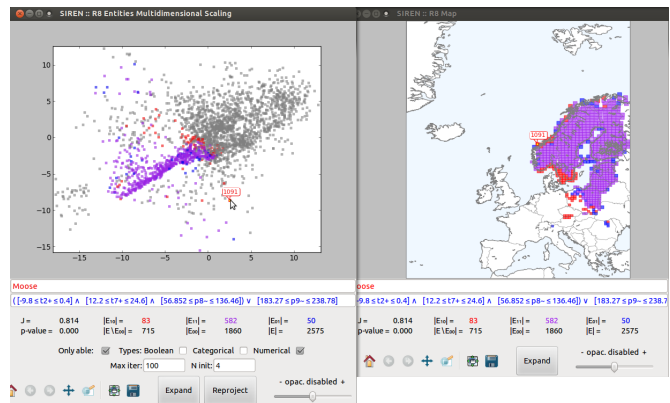
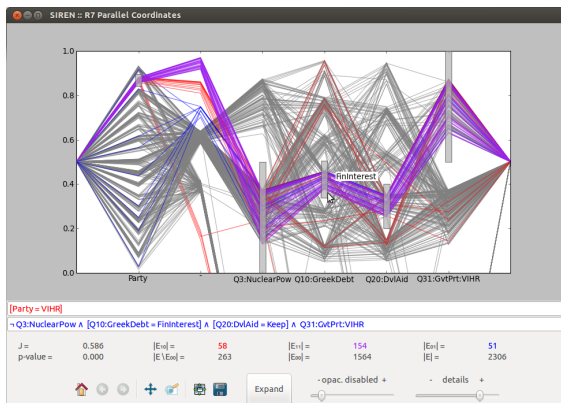
At the core of *Siren* is the *ReReMi* redescription mining algorithm. Various modes of interaction with the mining algorithm are possible through the interface.

- Mine redescrptions from the data automatically.
- Mine extensions of an existing redescription, on both sides or selectively on one side.
- It is possible to select a subset of variables for use in mining/expanding, by disabling some variables and to specify a subset of entities of interest to emphasize.

3.2 Vizualizing

Siren offers a number of different visualizations.

- A *parallel coordinates* plot represents the values taken by the entites for the variable appearing in the queries. It allows to easily visualize the impact of the queries conditions on the support of the redescription.
- A number of *data projections* from the scikit-learn package allow to highlight different aspects of the data.
- A *map* (for geospatial data) to show the locations where both queries hold, only the left hand side query holds and only the right hand side query holds.



3.3 Editing and Selecting

Existing redescrptions can be edited and the visualization and statistics will automatically recomputed. It is also possible to build a new redescription from scratch.