



SIREN

AN INTERACTIVE TOOL FOR MINING AND VISUALIZING GEOSPATIAL REDESCRIPTIONS

Esther Galbrun & Pauli Miettinen

galbrun@cs.helsinki.fi, Helsinki Institute for Information Technology
Department of Computer Science, University of Helsinki, Finland
pmiettinen@mpi-inf.mpg.de, Max Planck Institute for Informatics
Saarbrücken, Germany

► REDESCRIPTION MINING

Given a **set of entities** with **two sets of characterizing variables** find a **pair of queries** that describe approximately the **same entities**.

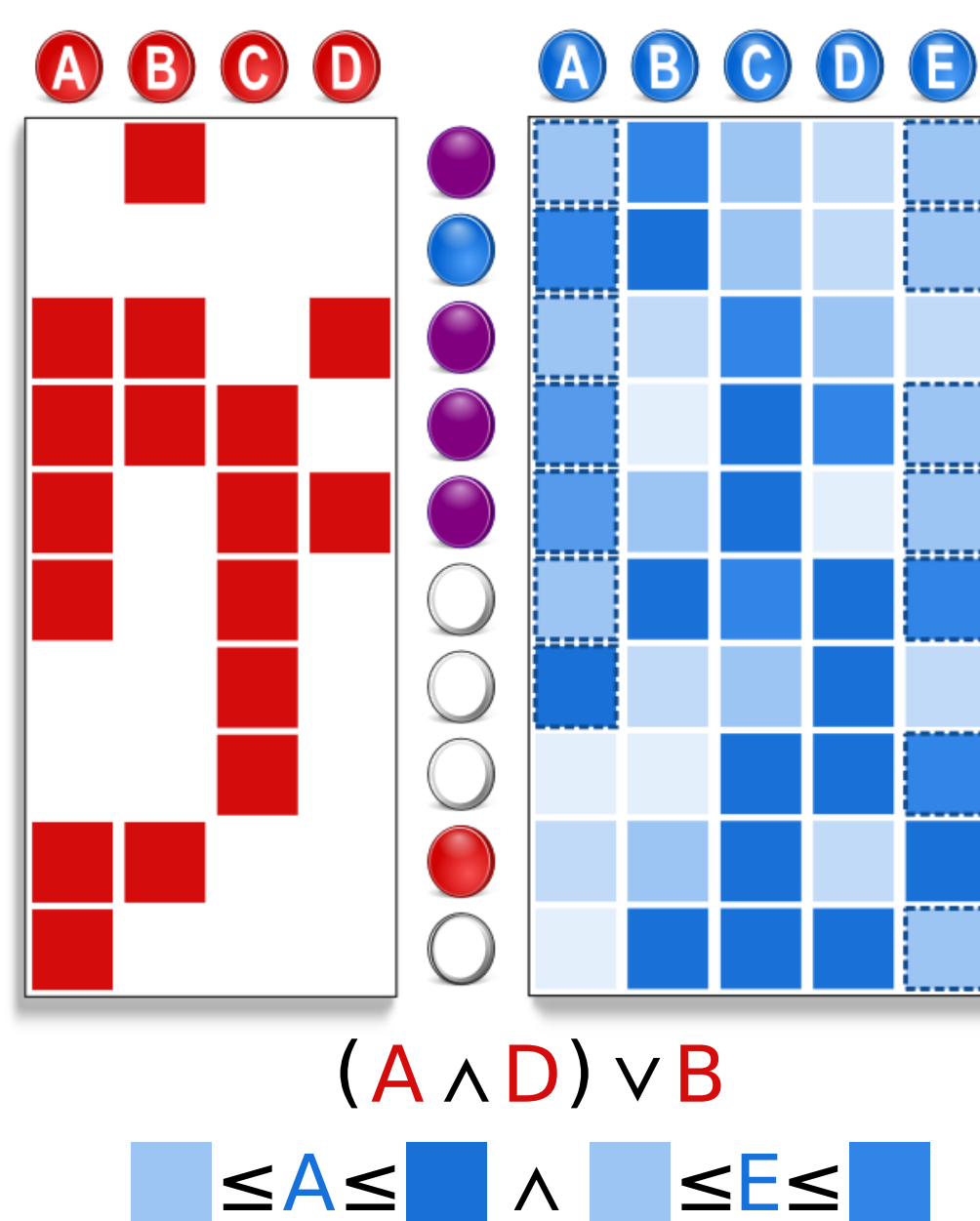
Finding **alternative descriptions of the same entities** is a problem that appears in many areas of science, for example biology (see below).

Redescription mining is a **powerful data analysis tool**.

Two points of view:

- Coherent subsets of **entities**, which can be described in two ways.
- **Variables and conditions** appearing together in the queries.

Our **ReReMi** algorithm [1] extends this problem to **real-valued data**.



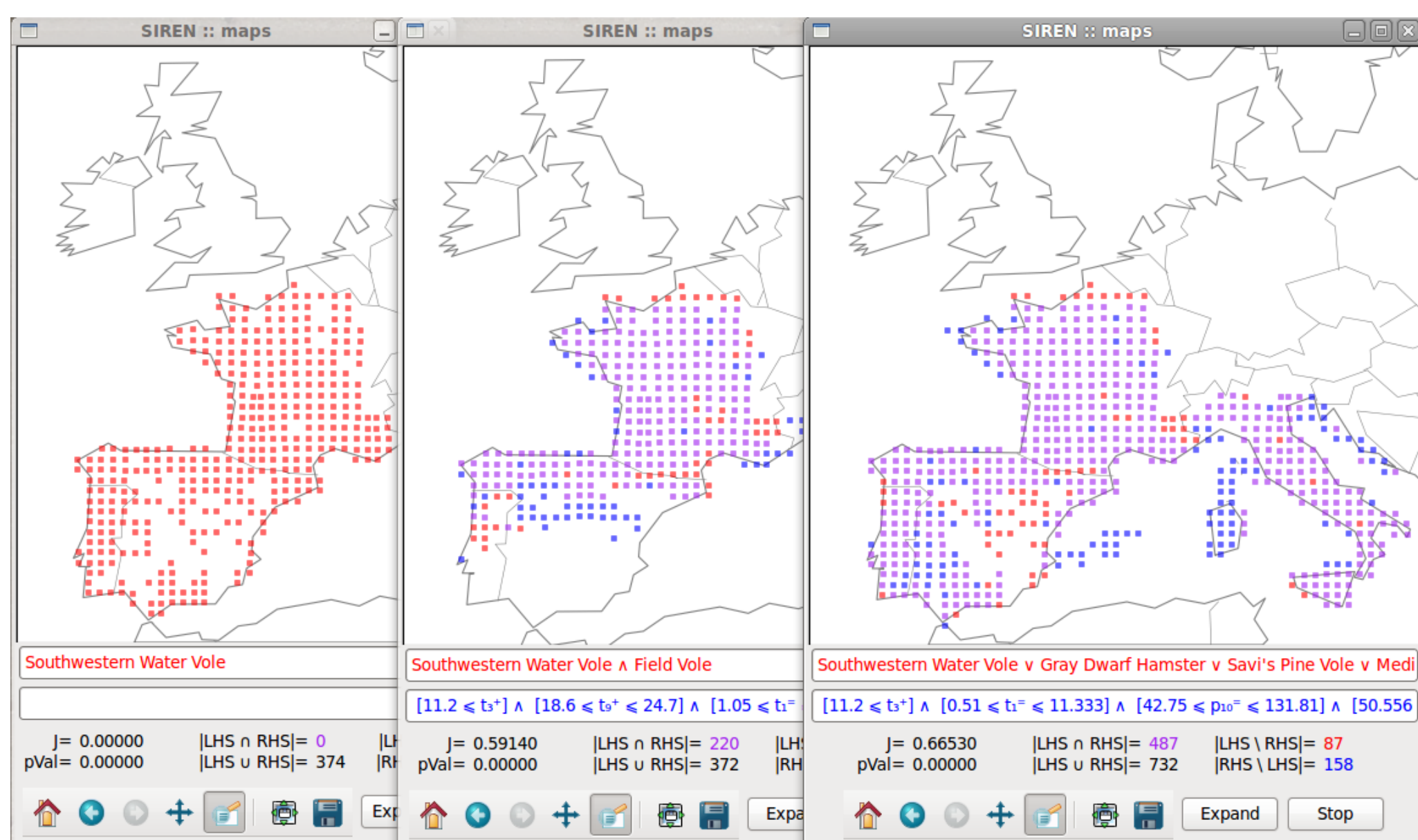
▼ INTERACTIVE DATA MINING

Mining data is generally an **iterative process**.

Results obtained at one step **give rise to hypotheses** which will be tested at a further step.

Siren improves the analysis by providing means to easily **interact with the redescription mining process**.

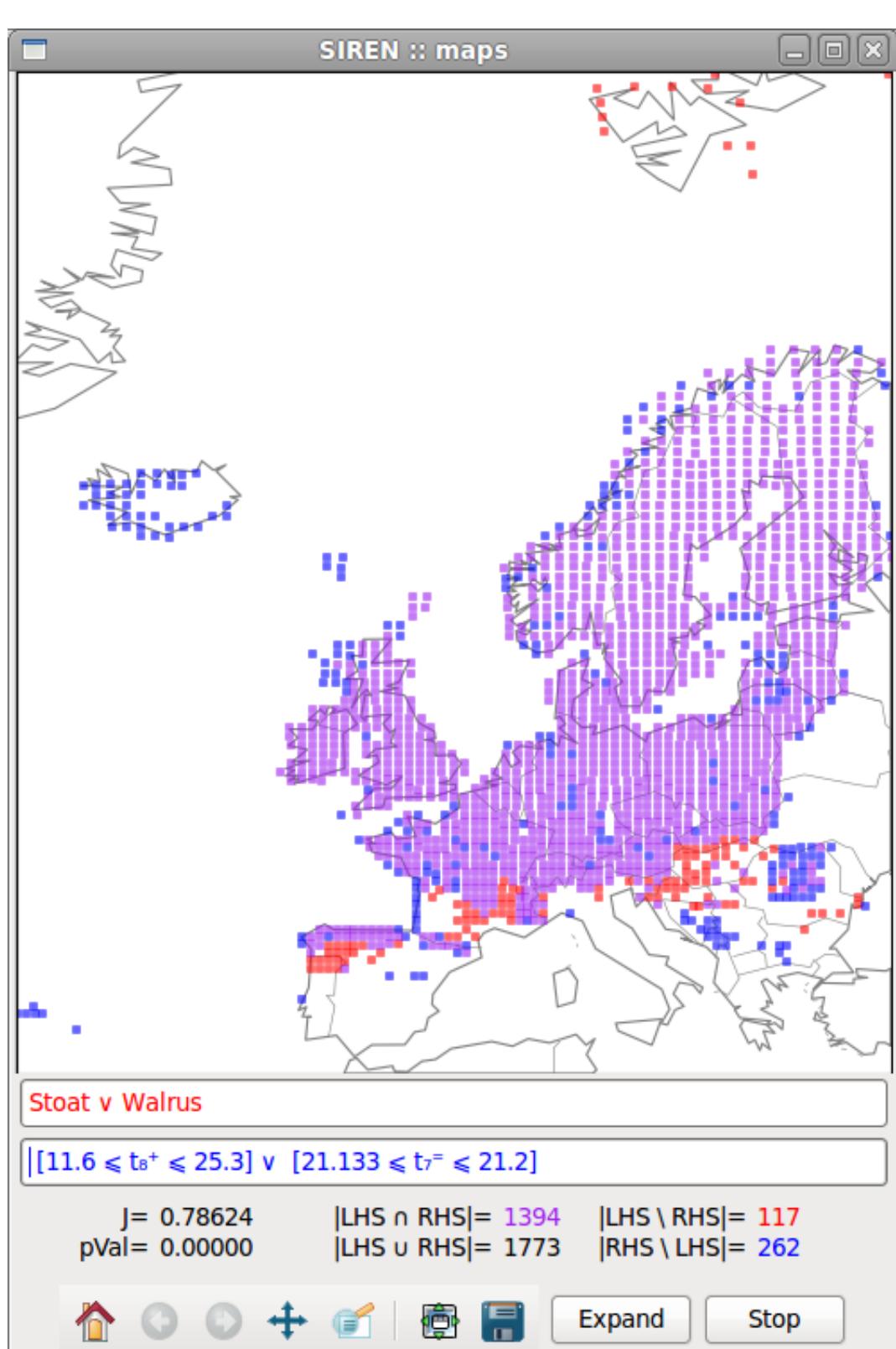
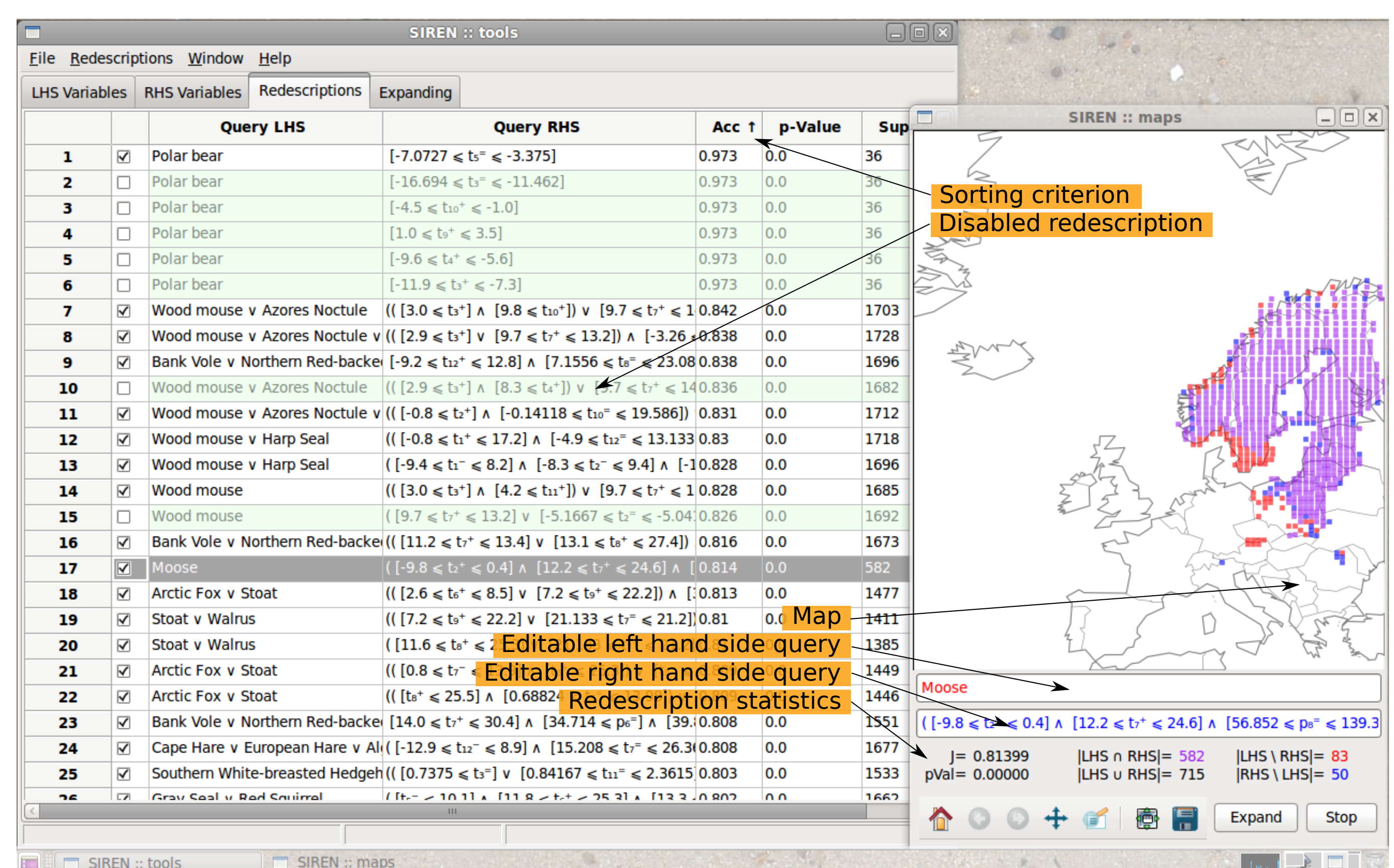
Query LHS	Query RHS	Acc	p-Value
1 Polar bear	$[1.0 \leq t_1 \leq 3.5]$	0.973	0.0
2 Moose	$[-9.8 \leq t_2 \leq 0.4] \wedge [12.2 \leq t_2 \leq 24.6] \wedge [56.852 \leq p_2 \leq 139.33]$	0.797	0.0
3 Gray Seal v Red Squirrel	$[t_2 \leq 10.1] \wedge [11.8 \leq t_2 \leq 25.3] \wedge [13.3 \leq t_2 \leq 27.3]$	0.781	0.0
4 Gray Seal v Red Squirrel	$[t_2 \leq 10.0] \wedge [12.0 \leq t_2 \leq 25.0] \vee [23.9 \leq t_2 \leq 24.6]$	0.801	0.0
5 Moose	$[-9.8 \leq t_2 \leq 0.4] \wedge [12.2 \leq t_2 \leq 24.6] \wedge [56.852 \leq p_2 \leq 139.33]$	0.797	0.0



▲ GEOSPATIAL REDESCRIPTIONS

Visualizing the redescriptions on a map is a key toward **interpreting the results** when the entities are connected to **geographical locations**.

A meaningful geospatial redescription should define **coherent areas** using **expressive queries**.



◀ BIOCLIMATIC NICHE-FINDING

An instance of **geospatial redescription mining** – and an important task in biology – is to **identify the bioclimatic constraints** that allow some **species** to survive.

► FUNCTIONALITIES

The functionalities of **Siren** include:

- Mining initial redescriptions
- Extending redescriptions
- Editing redescriptions
- Selecting variables
- Filtering results

LHS Variables	RHS Variables	Redescriptions	Expanding	Id	Name	Type	Min	Max
				1	t_1	numerical	-23.6	11.3
				2	t_2	numerical	-23.2	11.9
				3	t_2	numerical	-22.5	11.8
				4	t_2	numerical	-21.1	12.9
				5	t_2	numerical	-19.6	13.3
				6	t_2	numerical	-17.61	14.225
				7	t_2	numerical	-17.2	15.5
				8	t_2	numerical	-17.113	14.725
				9	t_2	numerical	-16.85	14.9

Query LHS	Query RHS	Acc	p-Value	Support
1 Polar bear	$[1.0 \leq t_1 \leq 3.5]$	0.973	0.0	36
2 Polar bear	$[-9.6 \leq t_2 \leq -5.6]$	0.973	0.0	36
3 Polar bear	$[-7.0727 \leq t_2 \leq -3.375]$	0.973	0.0	36
4 Polar bear	$[-4.5 \leq t_2 \leq -1.0]$	0.973	0.0	36
5 Polar bear	$[-16.694 \leq t_2 \leq -11.462]$	0.973	0.0	36
6 Polar bear	$[-11.9 \leq t_2 \leq -7.3]$	0.973	0.0	36
7 Wood mouse v Azores Noctule	$[(3.0 \leq t_1) \wedge [9.8 \leq t_2]] \vee [9.7 \leq t_2 \leq 1.0.842]$	0.0	1703	
8 Wood mouse v Azores Noctule v Harp Seal	$[(2.9 \leq t_1) \vee [9.7 \leq t_2 \leq 13.2]] \wedge [-3.26 \leq p_2 \leq 139.33]$	0.0	1728	
9 Bank Vole v Northern Red-backed Vole v Steppe Mole	$[-9.2 \leq t_2 \leq 12.8] \wedge [7.1556 \leq t_2 \leq 23.080.838]$	0.0	1696	
10 Wood mouse v Azores Noctule	$[(2.9 \leq t_1) \wedge [8.3 \leq t_2]] \vee [9.7 \leq t_2 \leq 140.836]$	0.0	1682	
11 Wood mouse v Azores Noctule v Harp Seal	$[(1.0 \leq t_1) \wedge [0.14118 \leq t_2 \leq 19.586]] \wedge [0.831]$	0.0	1712	
12 Wood mouse v Harp Seal	$[(1.0 \leq t_1) \wedge [4.2 \leq t_2]] \vee [9.7 \leq t_2 \leq 1.0.828]$	0.0	1685	
13 Wood mouse	$[(3.0 \leq t_1) \wedge [4.2 \leq t_2]] \vee [9.7 \leq t_2 \leq 1.0.828]$	0.0	1685	
14 Wood mouse	$[(9.7 \leq t_2 \leq 13.2) \vee [-5.1667 \leq t_2 \leq -5.04] \wedge [0.826]$	0.0	1692	
15 Wood mouse	$[(1.2 \leq t_1) \wedge [0.6882] \wedge [13.1 \leq t_2 \leq 27.4]] \wedge [0.816]$	0.0	1673	
16 Bank Vole v Northern Red-backed	$[(11.2 \leq t_1 \leq 13.4) \wedge [13.1 \leq t_2 \leq 27.4]] \wedge [0.816]$	0.0	1673	
17 Moose	$[-9.8 \leq t_2 \leq 0.4] \wedge [12.2 \leq t_2 \leq 24.6] \wedge [0.814]$	0.0	582	
18 Arctic Fox v Stoat	$[(2.6 \leq t_1 \leq 8.5) \wedge [7.2 \leq t_2 \leq 22.2]] \wedge [0.813]$	0.0	1477	
19 Stoat v Walrus	$[(17.2 \leq t_1 \leq 22.2) \wedge [21.133 \leq t_1 \leq 21.2]] \wedge [0.813]$	0.0	1411	
20 Stoat v Walrus	$[(11.6 \leq t_1 \leq 22.2) \wedge [21.133 \leq t_1 \leq 21.2]] \wedge [0.813]$	0.0	1385	
21 Arctic Fox v Stoat	$[(1.8 \leq t_1) \wedge [0.6882] \wedge [13.1 \leq t_2 \leq 27.4]] \wedge [0.816]$	0.0	1449	
22 Arctic Fox v Stoat	$[(1.8 \leq t_1) \wedge [0.6882] \wedge [13.1 \leq t_2 \leq 27.4]] \wedge [0.816]$	0.0	1446	
23 Bank Vole v Northern Red-backed	$[14.0 \leq t_1 \leq 30.4] \wedge [34.714 \leq p_1] \wedge [139.0.808]$	0.0	1551	
24 Cape Hare v European Hare v All	$[-12.9 \leq t_2 \leq 8.9] \wedge [15.208 \leq t_2 \leq 26.310.808]$	0.0	1677	
25 Southern White-breasted Hedgh	$[(10.7375 \leq t_1) \wedge [0.84167 \leq t_2 \leq 2.3615 \wedge 0.803]$	0.0	1533	
26 Gray Seal v Red Squirrel	$[t_2 \leq 10.1] \wedge [11.8 \leq t_2 \leq 25.3] \wedge [13.3 \leq t_2 \leq 27.3]$	0.0	1663	

REFERENCES

[1] E. Galbrun & P. Miettinen. From Black and White to Full Colour: Extending Redescription Mining Outside the Boolean World. In Statistical Analysis and Data Mining, in press, online version available, 2012.

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

