PrefLib and the ComSoC Tools

Simon Rey

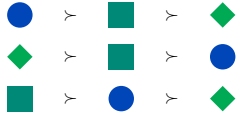
January 18, 2023

Universiteit van Amsterdam

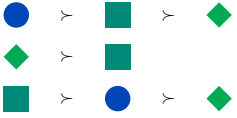
1. PrefLib

Library of Reported Preferences

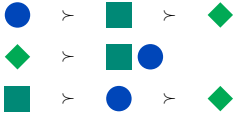
Strict Order Complete



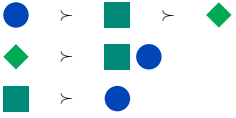
Strict Order Incomplete



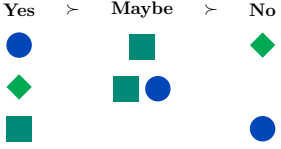
Tied Order Complete



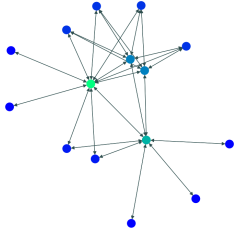
Tied Order Incomplete



Categorical



Weighted Matching (WMD)



58 Datasets, 13 734 Data Files

2002 French Presidential Election

00026

Election Politics

The 2002 French Presidential Election Dataset was collected by Jean-Francois Laslier and Karine Van der Straeten. It consists of 2,597 approval ballots collected in parallel to the actual election in 6 different districts in France.

The approval votes were collected at a set of polling stations in France during the first round of voting in the 2002 French National Election. Voters in these districts were informed prior to the election that they would have the ability to cast an approval ballot along with their normal ballot for the election. Overall, over 75% of those who turned up to vote participated in the experiment. Each of the files represent one district voting on the same election. There are between 367 and 476 voters (2,597 in all) and 16 candidates. Additional details the method used to collect the data and results of analysis can be found in the required citation for the use of this dataset.

Consists of 6 data files: CyfasNonains Orsay1 Orsay5 Orsay6 Orsay7 Orsay12

Details

Download [zip, 38.7 KB]

AAMAS Bidding Data

00037

Matching

This dataset contains the bids of reviewers over papers from the 2015, 2016 and 2021 Autonomous Agents and Multiagent Systems Conference.

For the years 2015 and 2016, inclusion in these data sets were explicitly opt-in; 2015 contains 9,817 bids of 201 reviewers over 613 papers; this represents about 40% of the actual 22,360 bids of 281 reviewers over 670 papers. The 2016 data contains 161 out of 393 reviewers with bids over 442 out of 550 papers. For the year 2021, 526 submissions, 71 SPC members, and 596 regular PC members passed the checks (not opting-out, etc...).

The bidding language for these conferences is **yes/maybe/no/conflict**. In order to make these more useful for PrefLib users, we have converted them to categorical data of the form **{yes} > {maybe} > {no response} > {no} > {conflict}**. Note that not all years have the same categories. We are deeply grateful to the IFAAMAS board and *Rafael Bordini*, *Edith Elkind*, *John Thangarajah*, and *David Shield* for approving, coordinating, and providing this dataset. The 2021 data has been generously provided by *Ulle Endriss*.

Consists of 3 data files: AAMAS 2015 AAMAS 2016 AAMAS 2021 with roles

Details

Download [zip, 701.7 KB]

AGH Course Selection

00009

Election

This dataset contains the results of surveying students at *AGU University of Science and Technology* about their course preferences. Each student provided a rank ordering over all the courses with no missing elements. There are 9 courses to choose from in 2003 and 7 in 2004.

The data on this page has been donated by *Piotr Faliszewski*.

Consists of 2 data files: 2003 Registration 2004 Registration

Details

Download [zip, 3.4 KB]

APA Election Data

00028

Election

This dataset contains the results of the elections of the American Psychological Association between 1998 - 2009. The voters are allowed to rank any number of the 5 candidates without ties. Each of these elections have 5 candidates and between 13,318 and 20,239 voters.

These data were donated by Michal Regenwetter and Anna Popova from the University of Illinois at Urbana-Champaign. The work that analyzed this data was supported by National Science Foundation grants SES # 08-20009, ICES # 1216016 (PI: M. Regenwetter), the University Library at the University of Illinois at Urbana-Champaign (PI: A. Popova), and the Basic Research Program at HSE (S. Popov). We thank the American Psychological Association for permitting access to its election ballot data. More information and tools for analyzing this data can be found at the website of the *Decision Making Laboratory at the University of Illinois*.

Consists of 12 data files: APA_1998 APA_1999 APA_2000 APA_2001 APA_2002 APA_2003 APA_2004 and *3 more*.

Details

Download [zip, 14.4 KB]

PrefLib

Brief History

{PrefLib}: A Library for Preferences

Main **About** **Papers** **Data Formats** **Data By Domain** **Data By Type** **Tools**

A reference library of preference data and links assembled by Nicholas Mattei and Toby Walsh. We currently house over 3,000 datasets for use by the community.

We want to provide a comprehensive resource for the multuple research communities that deal with preferences, including computational social choice, recommender systems, data mining, machine learning, and combinatorial optimization, to name just a few.

Please see the about page for information about the site, contacting us, and our citation policy. We rely on the support of the community in order to grow the usefulness of this site. To contribute, please contact Nicholas Mattei at: nicholas**{dot}**mattei@nicta.com.au

$a > b > c > d$

$a > b, c, d > e$

$\frac{1}{2} : a > b > c$
$\frac{1}{4} : c > b > a$
$\frac{1}{4} : b > c > a$

Supported By:

NICTA

web.archive.org/web/20130817164743/http://www.preflib.org/

PrefLib: A Library for Preferences

PrefLib

PrefLib is a reference library of preference data and links assembled by *Nicholas Mattei*, *Toby Walsh* and lately *Simon Rey*. This site and library is proudly supported by the *Algorithmic Decision Theory group* at *Data61* and the *The COMSOC Group at the University of Amsterdam*.

We want to provide a comprehensive resource for the multiple research communities that deal with preferences, including computational social choice, recommender systems, data mining, machine learning, and combinatorial optimization, to name just a few.

For more information on PrefLib and some helpful tips on using it, please see Nick's Tutorial *Slides and Code* from *EXPLORE 2014*. Check out the *data type* page to learn more about the kind of data we provide.

Please see the *about* page for information about the site, contacting us, and our citation policy. We rely on the support of the community in order to grow the usefulness of this site. To contribute, please contact *Nicholas Mattei* at: nsmattei{at}gmail or *Simon Rey* at: s.j.rey{at}uva{dot}nl.

In Brief

We currently host:

- 11 types of data
- 38 datasets
- 3668 data files
- More than 3.37 Gb. of data

Other Links

Here are some links that you might find relevant as well.

- *DEMOCRATIX:* A Declarative Approach to Winner Determination

PrefLib

Major Recent Changes

New Data Format

```
  1 | 5
  2 | 0,Candidate 1
  3 | 1,Candidate 2
  4 | 2,Candidate 3
  5 | 3,Candidate 4
  6 | 4,Candidate 5
  7 | 17956,17956,202
  8 | 2428,3
  9 | 788,1
 10 | 565,3,1,4,2
 11 | 468,3,1
 12 | 439,1,3,4,2
 13 | 437,1,4,3,2
 14 | 436,3,2,1,4
 15 | 417,3,2
 16 | 409,3,4,1,2
 17 | 403,3,2,4,1
... | ...
```

Old Format

New Data Format

```
  1 | 5
  2 | 0,Candidate 1
  3 | 1,Candidate 2
  4 | 2,Candidate 3
  5 | 3,Candidate 4
  6 | 4,Candidate 5
  7 | 17956,17956,202
  8 | 2428,3
  9 | 788,1
 10 | 565,3,1,4,2
 11 | 468,3,1
 12 | 439,1,3,4,2
 13 | 437,1,4,3,2
 14 | 436,3,2,1,4
 15 | 417,3,2
 16 | 409,3,4,1,2
 17 | 403,3,2,4,1
... | ...
```

Old Format

```
  1 | # FILE NAME: 00028-00000006.soi
  2 | # TITLE: APA_2003
  3 | # DESCRIPTION:
  4 | # DATA TYPE: soi
  5 | # MODIFICATION TYPE: original
  6 | # RELATES TO:
  7 | # RELATED FILES: 00028-00000006.toc
  8 | # PUBLICATION DATE: 2017-04-13
  9 | # MODIFICATION DATE: 2022-09-16
 10 | # NUMBER ALTERNATIVES: 5
 11 | # NUMBER VOTERS: 17956
 12 | # NUMBER UNIQUE ORDERS: 286
 13 | # ALTERNATIVE NAME 1:  Candidate 1
 14 | # ALTERNATIVE NAME 2:  Candidate 2
 15 | # ALTERNATIVE NAME 3:  Candidate 3
 16 | # ALTERNATIVE NAME 4:  Candidate 4
 17 | # ALTERNATIVE NAME 5:  Candidate 5
 18 | 2428:  4
 19 | 788:  2
 20 | 565:  4,2,5,3
 21 | 468:  4,2
 22 | 439:  2,4,5,3
 23 | 437:  2,5,4,3
 24 | 436:  4,3,2,5
 25 | 417:  4,3
 26 | 409:  4,5,2,3
 27 | 403:  4,3,5,2
... | ...
```

New Format

New Data Format

```
  1 | 5
  2 | 0,Candidate 1
  3 | 1,Candidate 2
  4 | 2,Candidate 3
  5 | 3,Candidate 4
  6 | 4,Candidate 5
  7 | 17956,17956,202
  8 | 2428,3
  9 | 788,1
 10 | 565,3,1,4,2
 11 | 468,3,1
 12 | 439,1,3,4,2
 13 | 437,1,4,3,2
 14 | 436,3,2,1,4
 15 | 417,3,2
 16 | 409,3,4,1,2
 17 | 403,3,2,4,1
... | ...
```

Old Format

```
  1 | # FILE NAME: 00028-00000006.soi
  2 | # TITLE: APA_2003
  3 | # DESCRIPTION:
  4 | # DATA TYPE: soi
  5 | # MODIFICATION TYPE: original
  6 | # RELATES TO:
  7 | # RELATED FILES: 00028-00000006.toc
  8 | # PUBLICATION DATE: 2017-04-13
  9 | # MODIFICATION DATE: 2022-09-16
 10 | # NUMBER ALTERNATIVES: 5
 11 | # NUMBER VOTERS: 17956
 12 | # NUMBER UNIQUE ORDERS: 286
 13 | # ALTERNATIVE NAME 1:  Candidate 1
 14 | # ALTERNATIVE NAME 2:  Candidate 2
 15 | # ALTERNATIVE NAME 3:  Candidate 3
 16 | # ALTERNATIVE NAME 4:  Candidate 4
 17 | # ALTERNATIVE NAME 5:  Candidate 5
 18 | 2428:  4
 19 | 788:  2
 20 | 565:  4,2,5,3
 21 | 468:  4,2
 22 | 439:  2,4,5,3
 23 | 437:  2,5,4,3
 24 | 436:  4,3,2,5
 25 | 417:  4,3
 26 | 409:  4,5,2,3
 27 | 403:  4,3,5,2
... | ...
```

Versatile metadata section

New Format

New Data Format

```
  1 | 5
  2 | 0,Candidate 1
  3 | 1,Candidate 2
  4 | 2,Candidate 3
  5 | 3,Candidate 4
  6 | 4,Candidate 5
  7 | 17956,17956,202
  8 | 2428,3
  9 | 788,1
 10 | 565,3,1,4,2
 11 | 468,3,1
 12 | 439,1,3,4,2
 13 | 437,1,4,3,2
 14 | 436,3,2,1,4
 15 | 417,3,2
 16 | 409,3,4,1,2
 17 | 403,3,2,4,1
... | ...
```

Old Format

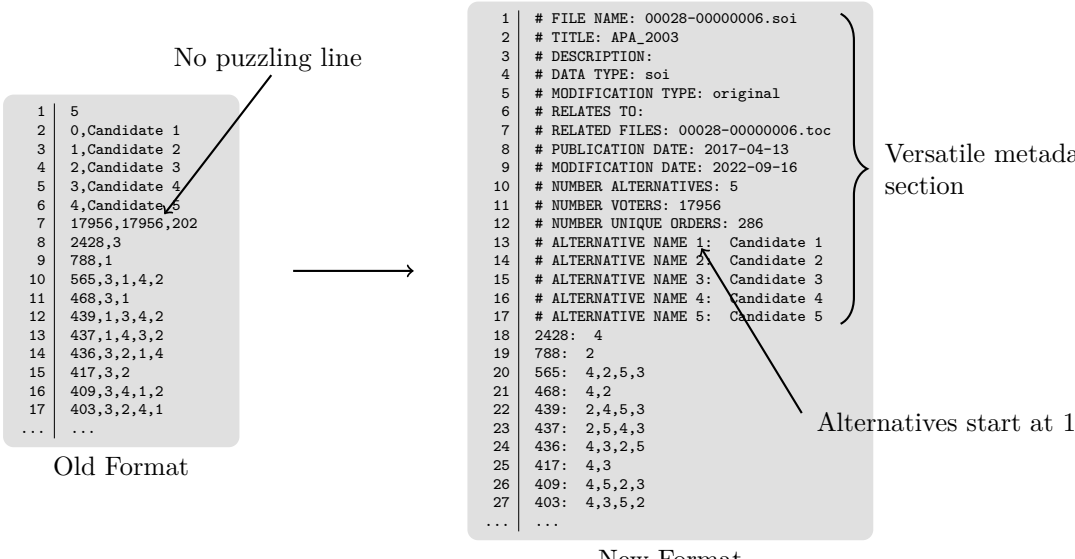
```
  1 | # FILE NAME: 00028-00000006.soi
  2 | # TITLE: APA_2003
  3 | # DESCRIPTION:
  4 | # DATA TYPE: soi
  5 | # MODIFICATION TYPE: original
  6 | # RELATES TO:
  7 | # RELATED FILES: 00028-00000006.toc
  8 | # PUBLICATION DATE: 2017-04-13
  9 | # MODIFICATION DATE: 2022-09-16
 10 | # NUMBER ALTERNATIVES: 5
 11 | # NUMBER VOTERS: 17956
 12 | # NUMBER UNIQUE ORDERS: 286
 13 | # ALTERNATIVE NAME 1:  Candidate 1
 14 | # ALTERNATIVE NAME 2:  Candidate 2
 15 | # ALTERNATIVE NAME 3:  Candidate 3
 16 | # ALTERNATIVE NAME 4:  Candidate 4
 17 | # ALTERNATIVE NAME 5:  Candidate 5
 18 | 2428:  4
 19 | 788:  2
 20 | 565:  4,2,5,3
 21 | 468:  4,2
 22 | 439:  2,4,5,3
 23 | 437:  2,5,4,3
 24 | 436:  4,3,2,5
 25 | 417:  4,3
 26 | 409:  4,5,2,3
 27 | 403:  4,3,5,2
... | ...
```

Versatile metadata section

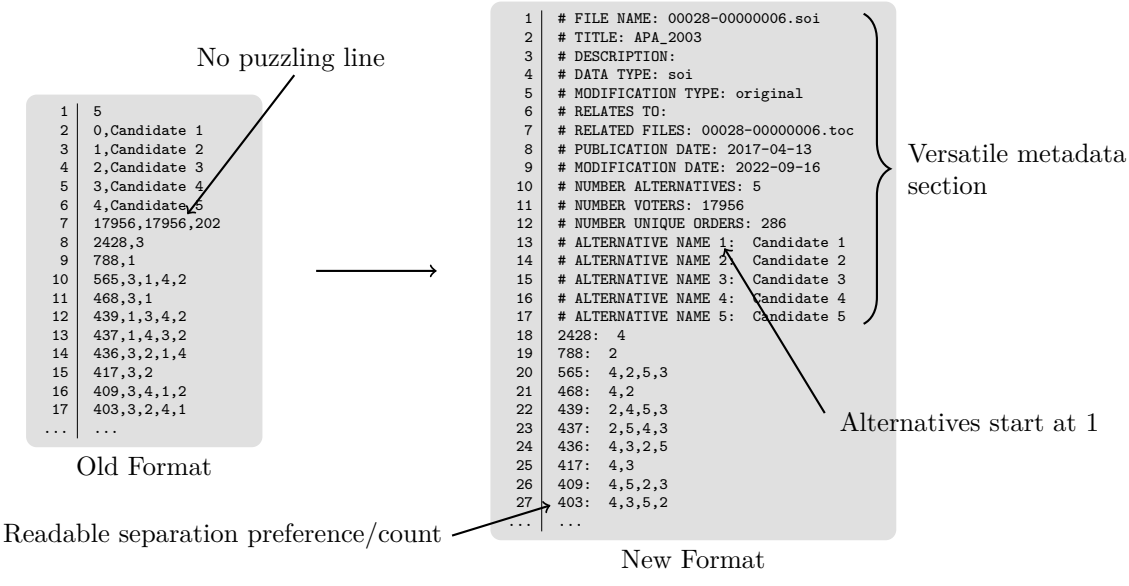
Alternatives start at 1

New Format

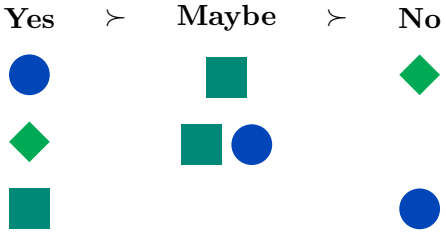
New Data Format



New Data Format



Categorical Preferences



Alternatives are placed into categories that are ordered, *e.g.*, review bids.

New Datasets

Data collected by Niclas Boehmer and Nathan Schaar in Collecting, Classifying, Analyzing, and Using Real-World Elections (AAMAS, 2023) `[link]`

↳ 16 new datasets, around 50% of the files hosted on PrefLib!

Thank you ♥

New Datasets

Data collected by Niclas Boehmer and Nathan Schaar in Collecting, Classifying, Analyzing, and Using Real-World Elections (AAMAS, 2023) `[link]`

↳ 16 new datasets, around 50% of the files hosted on PrefLib!

Breakfast Items digitalised by Dominik Peters `[link]`

↳ Same 42 respondents ranked 15 alternatives in 5 different scenarios

Thank you ♥

New Datasets

Data collected by Niclas Boehmer and Nathan Schaar in Collecting, Classifying, Analyzing, and Using Real-World Elections (AAMAS, 2023) `[link]`

↳ 16 new datasets, around 50% of the files hosted on PrefLib!

Breakfast Items digitalised by Dominik Peters `[link]`

↳ Same 42 respondents ranked 15 alternatives in 5 different scenarios

Parliamentary Elections digitalised by Martin Lackner `[link]`

↳ All Austrian parliamentary elections form 1994 to 2019 (9 elections)

Thank you ♥

New Datasets

Data collected by Niclas Boehmer and Nathan Schaar in Collecting, Classifying, Analyzing, and Using Real-World Elections (AAMAS, 2023) [link]

↳ 16 new datasets, around 50% of the files hosted on PrefLib!

Breakfast Items digitalised by Dominik Peters [link]

↳ Same 42 respondents ranked 15 alternatives in 5 different scenarios

Parliamentary Elections digitalised by Martin Lackner [link]

↳ All Austrian parliamentary elections form 1994 to 2019 (9 elections)

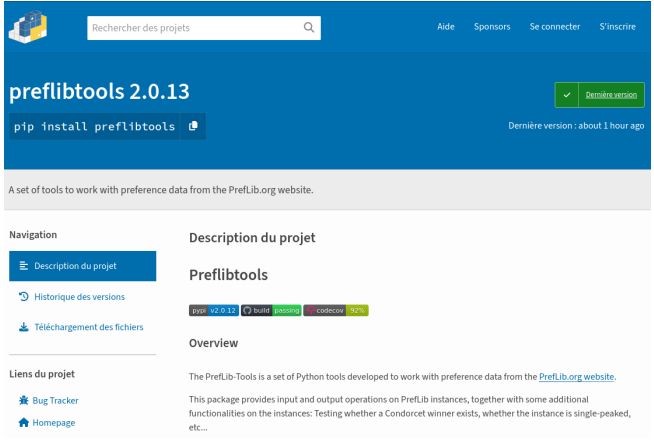
NSW Legislative Assembly digitalised by Floyd Everest [link]

↳ Strict incomplete orders for all 93 districts (years 2015, 2019, 2023)

Thank you ♥

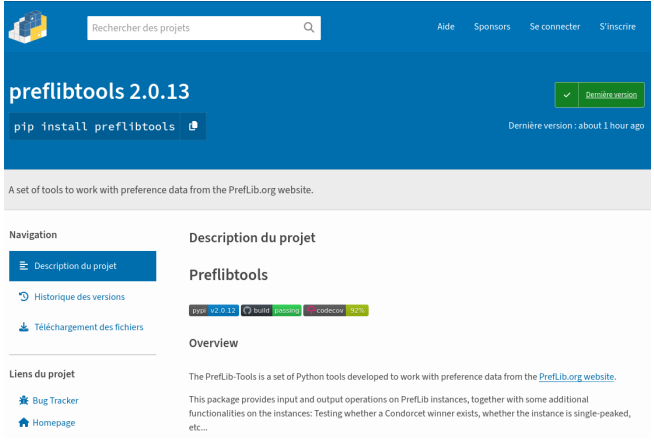
The PrefLibTools

`github.com/PrefLib/preflibtools`



The PrefLibTools

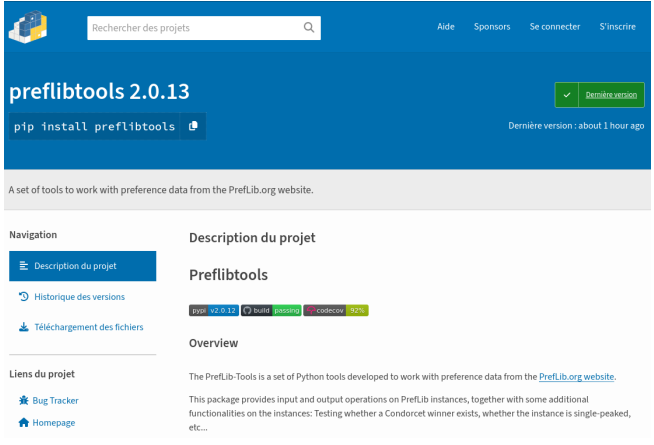
`github.com/PrefLib/preflibtools`



↪ Hosted on PyPI!

The PrefLibTools

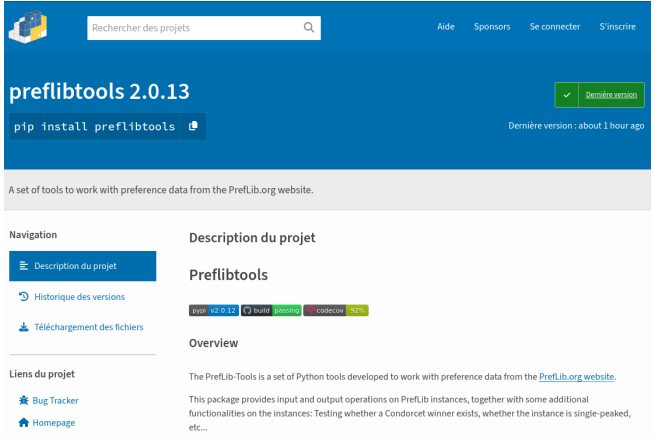
`github.com/PrefLib/preflibtools`



- Hosted on PyPI!
- Parser for all formats of preferences

The PrefLibTools

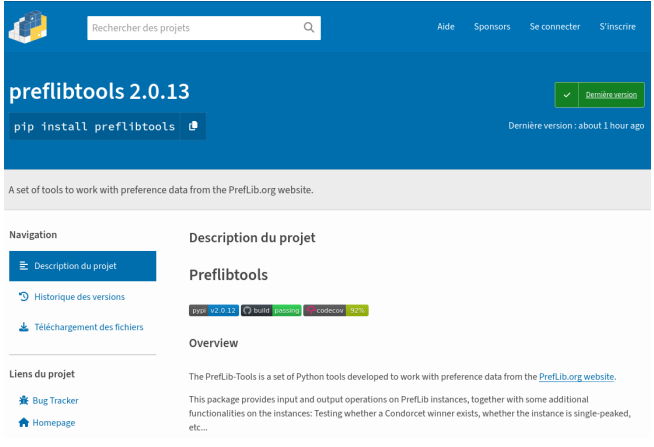
`github.com/PrefLib/preflibtools`



- Hosted on PyPI!
- Parser for all formats of preferences
- Basic properties checks

The PrefLibTools

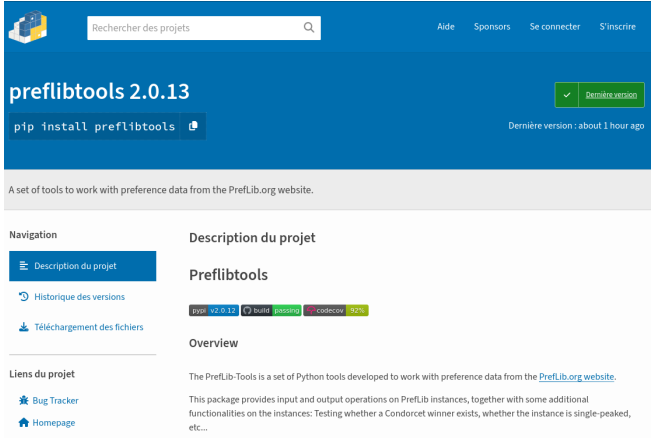
`github.com/PrefLib/preflibtools`



- Hosted on PyPI!
- Parser for all formats of preferences
- Basic properties checks
- Backward-compatible

The PrefLibTools

`github.com/PrefLib/preflibtools`



- Hosted on PyPI!
- Parser for all formats of preferences
- Basic properties checks
- Backward-compatible
  - Preference samplers (but use prefsampling instead)

A Community-Oriented Workflow

`github.com/PrefLib`

A Community-Oriented Workflow

`github.com/PrefLib`

PrefLib-Data

The data hosted on the website, together with some scripts used to maintain the integrity of the data

A Community-Oriented Workflow

`github.com/PrefLib`

PrefLib-Data

The data hosted on the website, together with some scripts used to maintain the integrity of the data

preflibtools

A small and lightweight Python package for working with the data, also hosted on PyPI

A Community-Oriented Workflow

`github.com/PrefLib`

PrefLib-Data

The data hosted on the website, together with some scripts used to maintain the integrity of the data

preflibtools

A small and lightweight Python package for working with the data, also hosted on PyPI

PrefLib-Django

Source of the PrefLib.org website (Django project)

A Community-Oriented Workflow

`github.com/PrefLib`

PrefLib-Data

The data hosted on the website, together with some scripts used to maintain the integrity of the data

preflibtools

A small and lightweight Python package for working with the data, also hosted on PyPI

PrefLib-Django

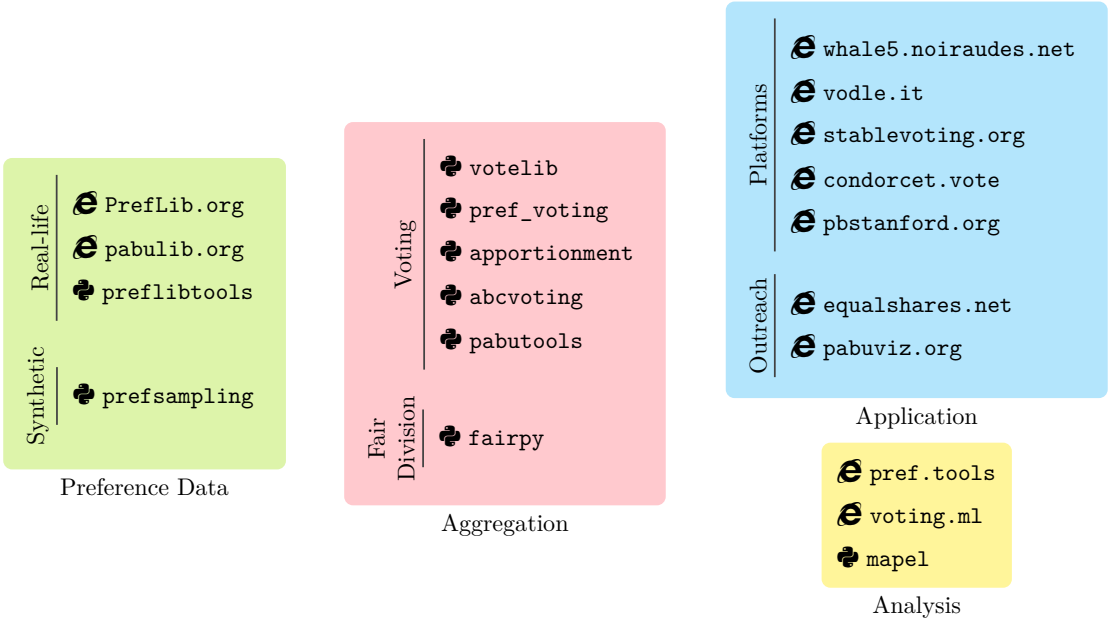
Source of the PrefLib.org website (Django project)

PrefLib-Contrib

A repository to keep track of contributions and TODO's

2. Other Tools

The ComSoC Ecosystem



New-comer — pabutools

pabutools 1.1.4

`pip install pabutools`

Dernière version

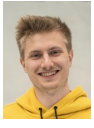
Dernière version : 6 déc. 2023

`github.com/pbvoting/pabutools`

A complete set of tools work with participatory budgeting instances.

Simon Rey

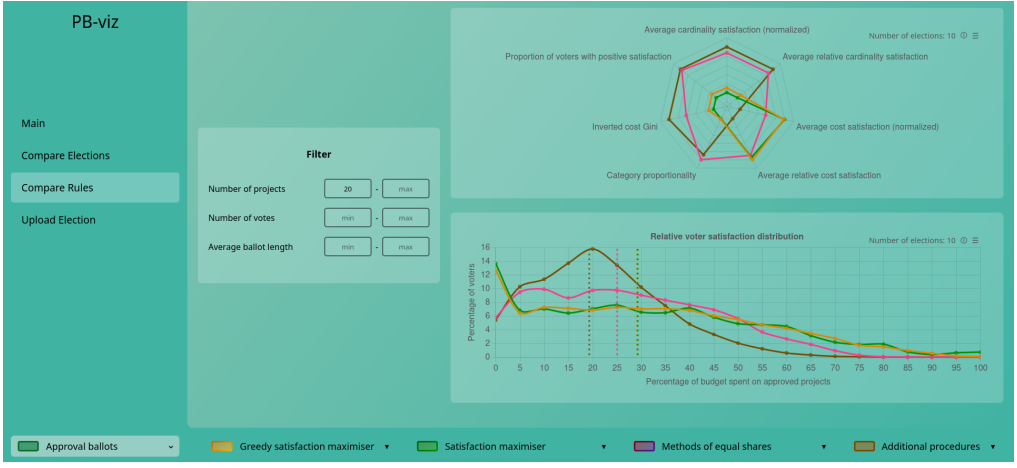
Grzegorz Pierczyński



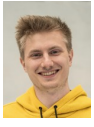
Markus Utke

Piotr Skowron

New-comer — pabuviz.org



Markus Utke



pabuviz.org

Simon Rey

New-comer — PrefSampling

prefsampling 0.0.2

Dernière version

```
pip install prefsampling
```

Dernière version : 18 déc. 2023

`github.com/Simon-Rey/prefsampling`

A lightweight Python library that provides preference samplers.

Simon Rey

Stanisław Szufa

3. Success Story?

Keep the Tools Alive!

- spliddit.org
- pnyx.dss.in.tum.de
- robovote.org

Keep the Tools Alive!

✓ Enable better research for all

✓ Increase societal impact

☠ `spliddit.org`

☠ `pnyx.dss.in.tum.de`

☠ `robovote.org`

Keep the Tools Alive!

✓ Enable better research for all

✓ Increase societal impact

✓ Develop skills

✓ Increase diversity

✓ (Somewhat) provides fame

- `spliddit.org`
- `pnyx.dss.in.tum.de`
- `robovote.org`

Keep the Tools Alive!

✓ Enable better research for all

✓ Increase societal impact

✓ Develop skills

✓ Increase diversity

✓ (Somewhat) provides fame

✗ Long to develop and to maintain

✗ Mildly recognised for one's academic career

✗ Vulnerable to the disappearance of a single person

✗ Hard to come by

PrefLib, a Community-Tool

Nick Mattei

Simon Rey

Moving PrefLib to the public domain

PrefLib, a Community-Tool

Nick Mattei

Simon Rey

Moving PrefLib to the public domain

↳ Make yourself known if you are interested!

Community Discussion

✓ Enable better research for all

✓ Increase societal impact

✓ Develop skills

✓ Increase diversity

✓ (Somewhat) provides fame

✗ Long to develop and to maintain

✗ Mildly recognised for one's academic career

✗ Vulnerable to the disappearance of a single person

✗ Hard to come by

- How to ensure continuous development of new great projects?
- How to ensure long-term survival of currently existing great projects?
- How to provide high visibility of great projects?

Afterthoughts? Reach out at `s.j.rey@uva.nl`