



TEdb project

An inter-laboratory research project for
data curation of thermoelectric properties
THERMOELECTRICS SOCIETY OF JAPAN

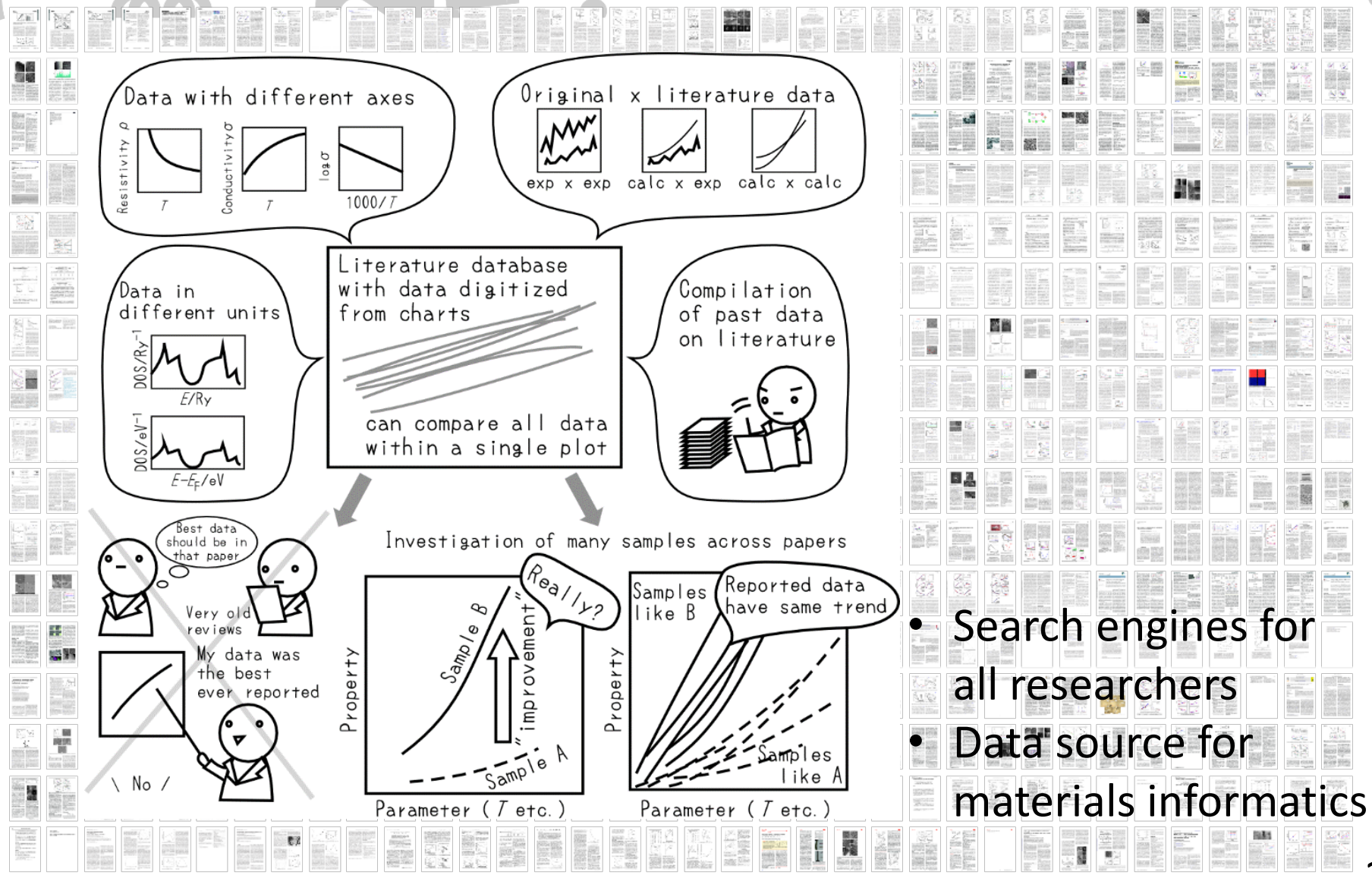
Starrydata2:

A plot mining web system for literature data collection

Yukari Katsura^{1,2*}, Masaya Kumagai³, Takushi Kodani^{1,2}, Hideyasu Ouchi^{1,2}, Sakiko Gunji², Yuki Ando², Yoji Imai³, Kaoru Kimura¹ and Koji Tsuda^{1,2,3}

¹Univ. of Tokyo ²MaDIS-MI²I, NIMS ³RIKEN AIP

Key for the next MI: Experimental data from thousands of past papers



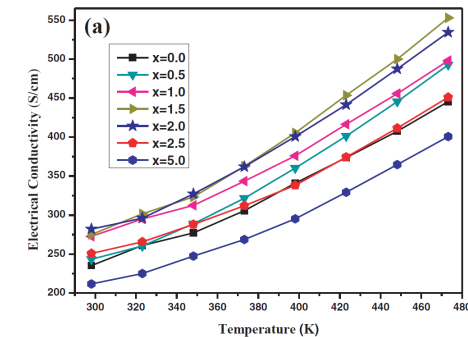
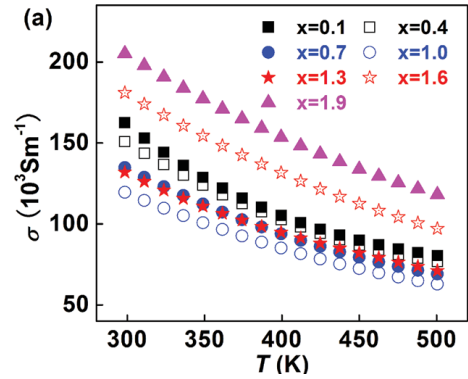
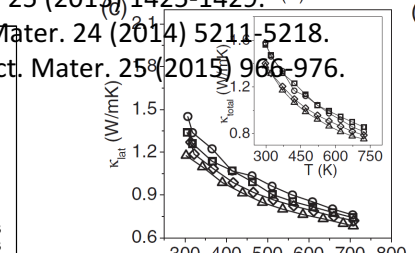
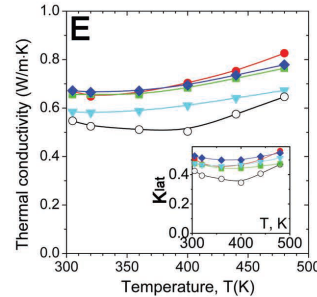
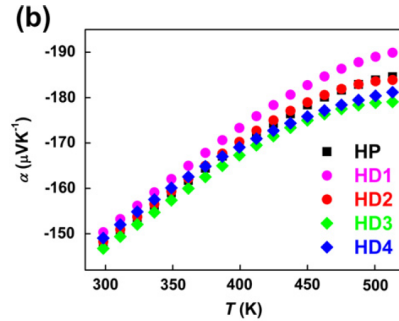
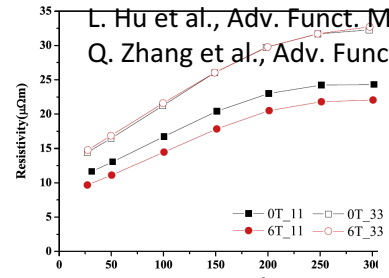
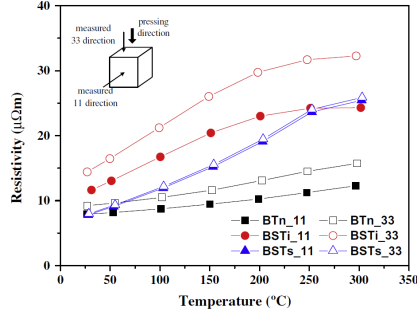
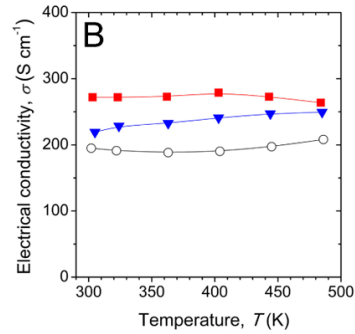
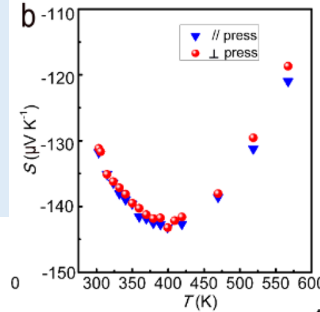
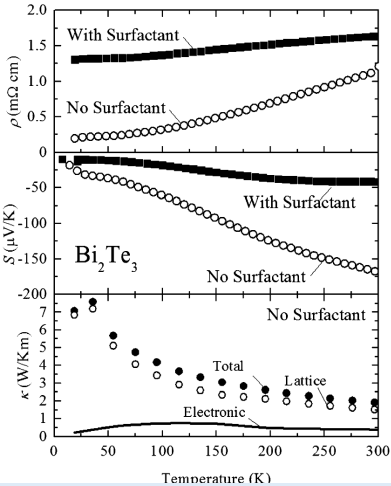
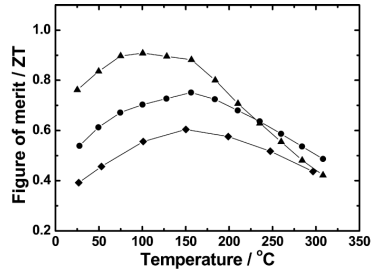
Recovery of experimental data from various plots

Tasks:

1. Plot digitization from images
2. Sample description from text (chemical composition etc.)

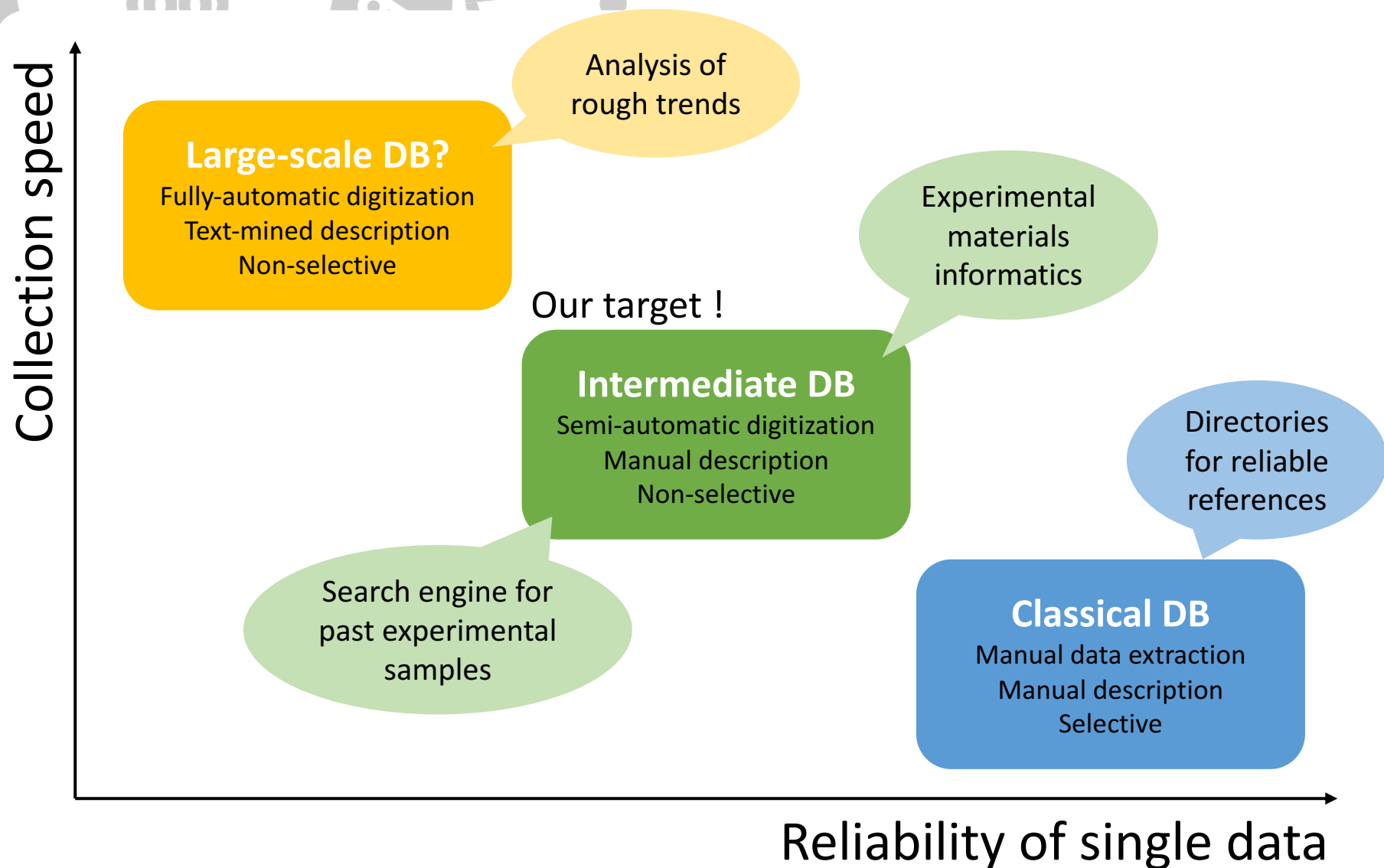
C. Kim et al., ACS Appl. Mater. Interfaces 4 (2012) 2949-2954.
 V. Stavila et al., ACS Appl. Mater. Interfaces 5 (2013) 6678-6686.
 L. Yang et al., ACS Appl. Mater. Interfaces 7 (2015) 23694-23699.
 Y. Min et al., ACS Nano 9 (2015) 6843-6853.
 D. Kim et al., Acta Materialia 59 (2011) 405-411.
 D. Kim et al., Acta Materialia 59 (2011) 4957-4963.
 L. Hu et al., Acta Materialia 60 (2012) 4431-4437.
 K. Biswas et al., Adv. Energy Mater. 2 (2012) 634-638.
 Y. Min et al., Adv. Mater. 25 (2013) 1425-1429.

L. Hu et al., Adv. Funct. Mater. 24 (2014) 5211-5218.
 Q. Zhang et al., Adv. Funct. Mater. 25 (2015) 966-976.



Time-consuming & boring for researchers → Developed a new web system

Target area of our experimental database



Starrydata2 web system: Online data-sharing platform for literature data

URL: <http://starrydata2.org>

Programming: Masaya Kumagai (RIKEN)

FREE to use

Database selection page

Starrydata2 Logout

Database lists

TEdbProject
Papers : 17274
Mylist : 12
Figure : 1403
Sample : 1913
Data : 6113

Other Materials database? Other Materials database? Other Materials database? Other Materials database?

Other Materials database? Other Materials database?

Catalyst materials?
Magnetic materials?
Battery materials?
Strongly-correlated systems?
Superconductors?
Dielectric materials?
....

Prototype database: “TEdbProject”

A database of experimental thermoelectric properties

Starrydata2 web system : Index page

URL: <http://starrydata2.org>

Programming: Masaya Kumagai (RIKEN)

Designed as a tool for researchers to organize information about published papers

The screenshot shows the Starrydata2 web system interface. On the left is a search and upload section. The main area displays a list of papers with various annotations. The top navigation bar includes 'Starrydata2', 'Logout', and a 'Mylists of papers' section with tabs for 'All', 'Uploaded', 'SiGe', 'Kimuralab', 'PbTe', and '2-1-1_Bi2Te3'. The 'All' tab is circled in red. Below the tabs are buttons for 'Get Reference List', 'Get Data', 'Remove Items from List', and 'Remove List'. The paper list includes entries with their titles, authors, and publication details. Annotations in red text highlight specific features: 'Bibliographic information: automatically retrieved from DOI' points to the DOI field; 'Link for plot-mined data' points to the 'Data' link; 'Link for Publisher's website for fulltext' points to the 'Link' field; and 'Suggestive filename for fulltext download' points to the 'Copy filename' link. The bottom of the page shows 'Display 25 results per page' and 'Page 1 of 2 >>'.

Starrydata2 Logout

Search for papers
Keyword
Title:

Add Papers to Lists

- Create List **Create original mylists**
- SiGe
- Kimuralab
- PbTe
- 2-1-1_Bi2Te3
- 2-1-2_Bi2Te3 画像切り出し...
- 1-1_Bi2Te3 **Addition of selected papers on mylists**
- ZT論文

Upload Papers

Input multiple DOIs (Digital Object Identifier)

List of all papers in DB

Mylists of papers

sid : 188 DOI : 10.1063/1.4938565 fignum : 0
Thermoelectric properties of p-type PbTe/Ag2Te bulk composites by extrinsic phase mixing
M. Lee, J. Rhyee, AIP Advances 5 (2015) 127223.
[Detail](#) [Data](#) [Link](#) [Copy filename](#) **Bibliographic information: automatically retrieved from DOI**

sid : 860 DOI : 10.1039/c1ee01314a fignum : 3
Reevaluation of PbTe1-xlx as high performance n-type thermoelectric material
A. LaLonde, Y. Pei, G. Snyder, Energy & Environmental Science 4 (2011) 2090.
[Detail](#) [Data](#) [Link](#) [Copy filename](#) **Link for plot-mined data**

sid : 585 DOI : 10.1021/cm803437x fignum : 0
Improvement in the Thermoelectric Figure of Merit by La/Ag Cosubstitution in PbTe
K. Ahn, C. Li, C. Uher, M. Kanatzidis, Chemistry of Materials 21 (2009) 1361-1367.
[Detail](#) [Data](#) [Link](#) [Copy filename](#) **Link for Publisher's website for fulltext**

sid : 674 DOI : 10.1088/0256-307x/22/8/077 fignum : 0
High Thermoelectric Properties of PbTe Doped with Bi2Te3and Sb2Te3
Z. Pin-Wen, I. Yoshio, I. Yukihiko, S. Yoshikazi, J. Xiao-Peng, Z. G... Chinese Physics Letters 22 (2005) 2103-2105.
[Detail](#) [Data](#) [Link](#) [Copy filename](#) **Suggestive filename for fulltext download**

sid : 676 DOI : 10.1088/0256-307x/30/1/017101 fignum : 0
Electronic Structure, Lattice Dynamics and Thermoelectric Properties of PbTe from First-P...
P. Lu, L. Qu, Chinese Physics Letters 30 (2013) 017101.
[Detail](#) [Data](#) [Link](#) [Copy filename](#)

sid : 725 DOI : 10.1002/crat.201200010 fignum : 0

Display results per page Page 1 of 2 >>

Starrydata2: Data-browsing page

Get Reference List
 Get Data
 Get DOIList
 Remove Item

sid : 6 DOI : 10.1021/am405410e figure : 3

Exploration of Zn Resonance Levels and Thermoelectric Properties in I-Doped PbTe with ZnTe Nanostructures
 P. Rawat, B. Paul, P. Banerji. ACS Applied Materials & Interfaces 6 (2014) 3995-4004.

[Detail](#)
[Data](#)
[Link](#)
[Copy filename](#)
① Click "Data"

sid : 31 DOI : 10.1016/j.actamat.2008.10.005 figure : 0

Starrydata2

Browse / Add / Edit data

Logout

sid : 6 DOI : 10.1021/am405410e figure : 3

Exploration of Zn Resonance Levels and Thermoelectric Properties in I-Doped PbTe with ZnTe Nanostructures
 P. Rawat, B. Paul, P. Banerji. ACS Applied Materials & Interfaces 6 (2014) 3995-4004.

[Detail](#) [Link](#)
Bibliographic information of original paper

Figures			
No.	FigureNumber	Prop.X	Prop.Y
0	10(b)	Temperature	Thermal conductivity
1	3(a)	Temperature	Electrical resistivity
2	6(b)	Temperature	Seebeck coefficient

List of figures with datasets

Samples		
No.	Samplename	Composition
0	Pb1.0015Zn0.02Te1.0210.003	Pb1.0015Zn0.02Te1.0210.003
1	Pb1.00075Zn0.01Te1.0110.0015	Pb1.00075Zn0.01Te1.0110.0015
2	Pb1.00025Zn0.01Te1.0110.0005	Pb1.00025Zn0.01Te1.0110.0005
3	Pb1.00015Zn0.01Te1.0110.0003	Pb1.00015Zn0.01Te1.0110.0003
4	Pb1.0015Zn0.01Te1.0110.003	Pb1.0015Zn0.01Te1.0110.003
5	Pb1.00025Zn0.02Te1.0210.0005	Pb1.00025Zn0.02Te1.0210.0005
6	Pb1.00075Zn0.02Te1.0210.0015	Pb1.00075Zn0.02Te1.0210.0015
7	Pb1.0005Zn0.02Te1.0210.001	Pb1.0005Zn0.02Te1.0210.001
8	Pb1.00015Zn0.02Te1.0210.0003	Pb1.00015Zn0.02Te1.0210.0003
9	Pb1.0005Zn0.01Te1.0110.001	Pb1.0005Zn0.01Te1.0110.001

List of samples reported in the paper

② Figure selector

③ Sample selector

Convert Unit

Figure 3(a)

Caption unknown

SampleName Pb1.0015Zn0.02Te1.02

Composition Pb1.0015Zn0.02Te1.02

Property(x) Temperature

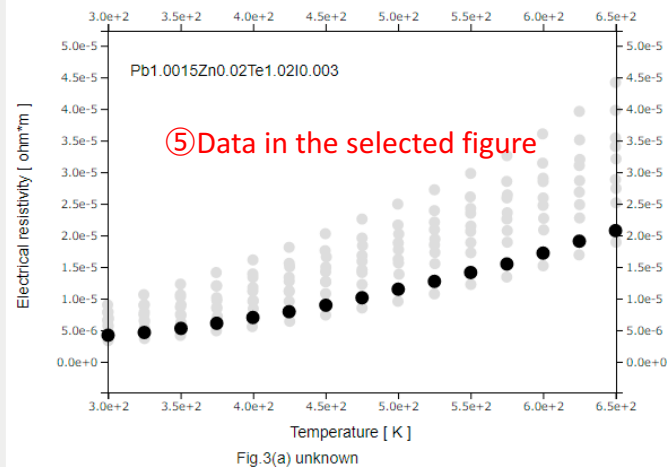
Property(y) Electrical resistivity

Unit(x) 10[^] 0 K

Unit(y) 10[^] 0 ohm*m

Add new data

④ Digital data in SI unit



Automatic unit conversion
 e.g. 10 mW/(cm*K)
 → 1 W/(m*K)

Plot-mining procedure:

(1) Get plot images

③ Capture plot images by Print Screen / FireShot / PDF reader etc.

① Access Publisher's webpage

② Access Fulltext PDF

Starrydata2

sid : 6 DOI : 10.1021/am405410e figure : 3

Exploration of Zn Resonance Level

P. Rawat, B. Paul, P. Banerji, ACS Applied

Detail **Link**

No.	FigureNumber	Figure	Pr
0	10(b)	Temp	
1	3(a)	Temp	
2	6(b)	Temp	

ACS Publications

ACS APPLIED MATERIALS & INTERFACES

Research Article

Exploration of Zn Resonance Levels and Thermoelectric Properties in I-Doped PbTe with ZnTe Nanostructures

P. K. Sanyal, S. Prasad, and P. Banerji

ACS Appl. Mater. Interfaces 2014, 6, 3995-4004

Figure 8

pubs.acs.org/doi/pdf/10.1021/am405410e

am405410e 1..10 5 / 10

temperature values of the thermopower in all of the samples (circular symbols).

above criteria, and as a result, significant enhancement has been observed in zT in these systems. The temperature dependence of the power factor ($S^2\rho$) is shown in Figure 8. The maximum power factor values and the corresponding temperatures in all of the samples are given in Table 3.

Figure 9 shows the temperature-dependent thermal diffusivity in all of the samples. The thermal diffusivity is found to increase with a rise in the PbTe content due to increased carrier concentrations at higher doping levels. The estimated

measured value of the specific heat is found to be nearly same as that of the bulk material.

values of zT measured in zT in these systems. The temperature dependence of the power factor ($S^2\rho$) is shown in Figure 8. The maximum power factor values and the corresponding temperatures in all of the samples are given in Table 3.

The thermal diffusivity is found to increase with a rise in the PbTe content due to increased carrier concentrations at higher doping levels. The estimated

選択範囲をキャプチャ Ctrl+Shift+Y
 ページ全体をキャプチャ
 表示部分をキャプチャ
 選択範囲をキャプチャ
 オプション...
 高度な機能をゲット...

Figure 8. Temperature dependence of the power factor in all of the samples.

④ Copy plot image

⑤ Go to bottom of Starrydata 2

⑦ Paste the plot image

スクリーンショットを保存

WebPlotDigitizer

提供サイト: <https://automeris.io> Ankit Rohatgi

★★★★★ (127) 学術文献 ユーザー数: 18,941人

Starrydata2

sid : 6 DOI : 10.1021/am405410e figure : 3

Exploration of Zn Resonance Levels and Thermoelectric Properties in I-Doped PbTe with ZnTe Nanostructures

P. Rawat, B. Paul, P. Banerji, ACS Applied Materials & Interfaces 2014, 6, 3995-4004

No.	FigureNumber	Figure	Prop. X	Prop. Y	SampleName	Samples	Composition
0	10(b)	Temperature	Thermal conductivity		Pb1_000520-01761_010_001	Pb1_000520-01761_010_001	Pb1_000520-01761_010_001
1	3(a)	Temperature	Electrical resistivity		Pb1_000520-01761_010_001	Pb1_000520-01761_010_001	Pb1_000520-01761_010_001
2	6(b)	Temperature	Seebeck coefficient		Pb1_000520-01761_010_001	Pb1_000520-01761_010_001	Pb1_000520-01761_010_001

Figure 8

WebPlotDigitizer

Align axes dialogue

SampleName

Composition

Property(x)

Property(y)

Unit(x)

Unit(y)

Save

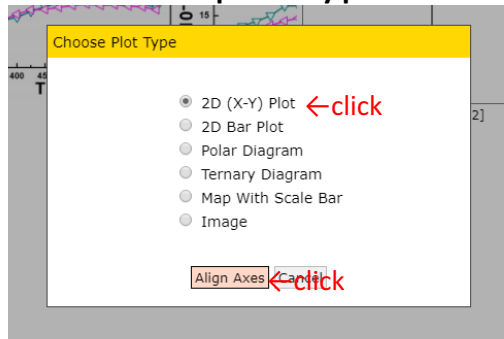
Zoom button

Align axes dialogue

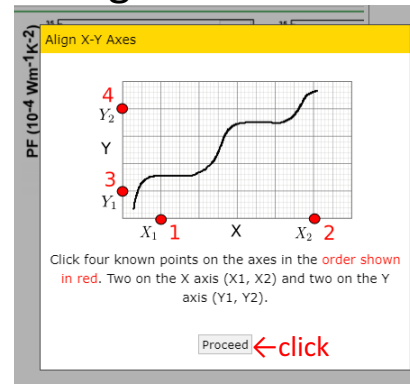
(2) Semi-automatic plot tracing by WebPlotDigitizer



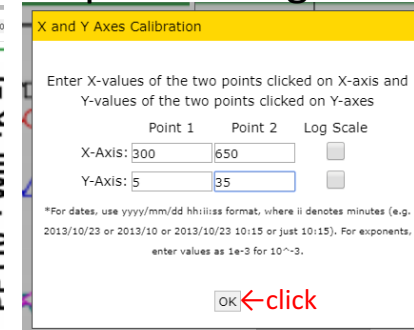
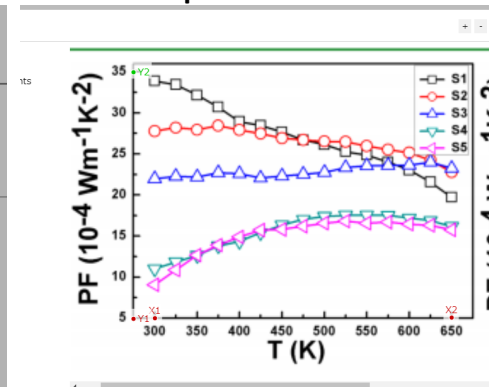
① Choose plot type



② Align X-Y Axes

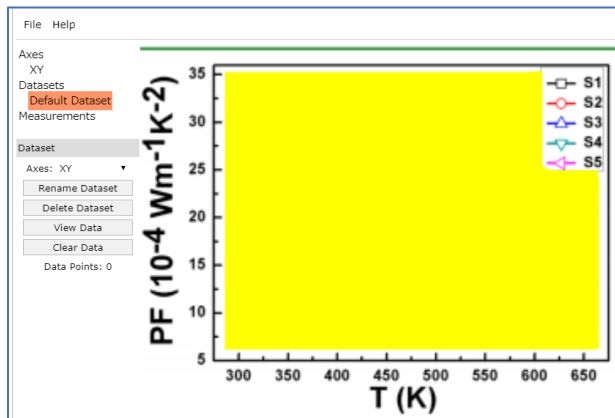


③ Click 4 points on axes ④ Input readings



⑤ Select tracing regions

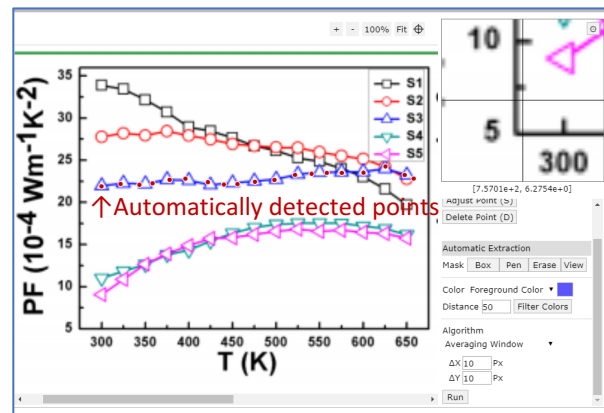
Box, Pen, Erase tools are available



Optional

⑥ Select color and Run

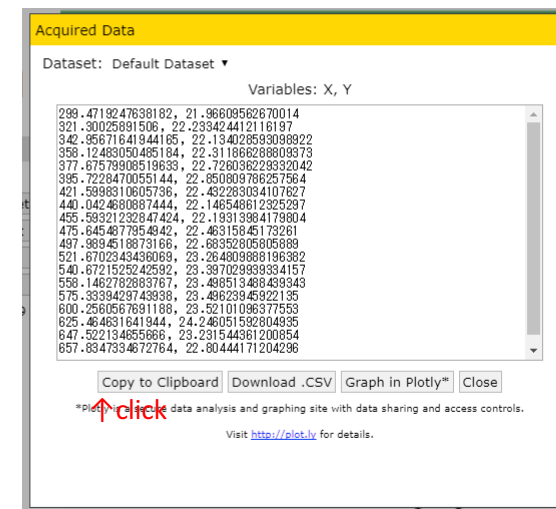
Select algorithm and tune parameters



Tracing by manual clicking is also possible

⑦ View data and

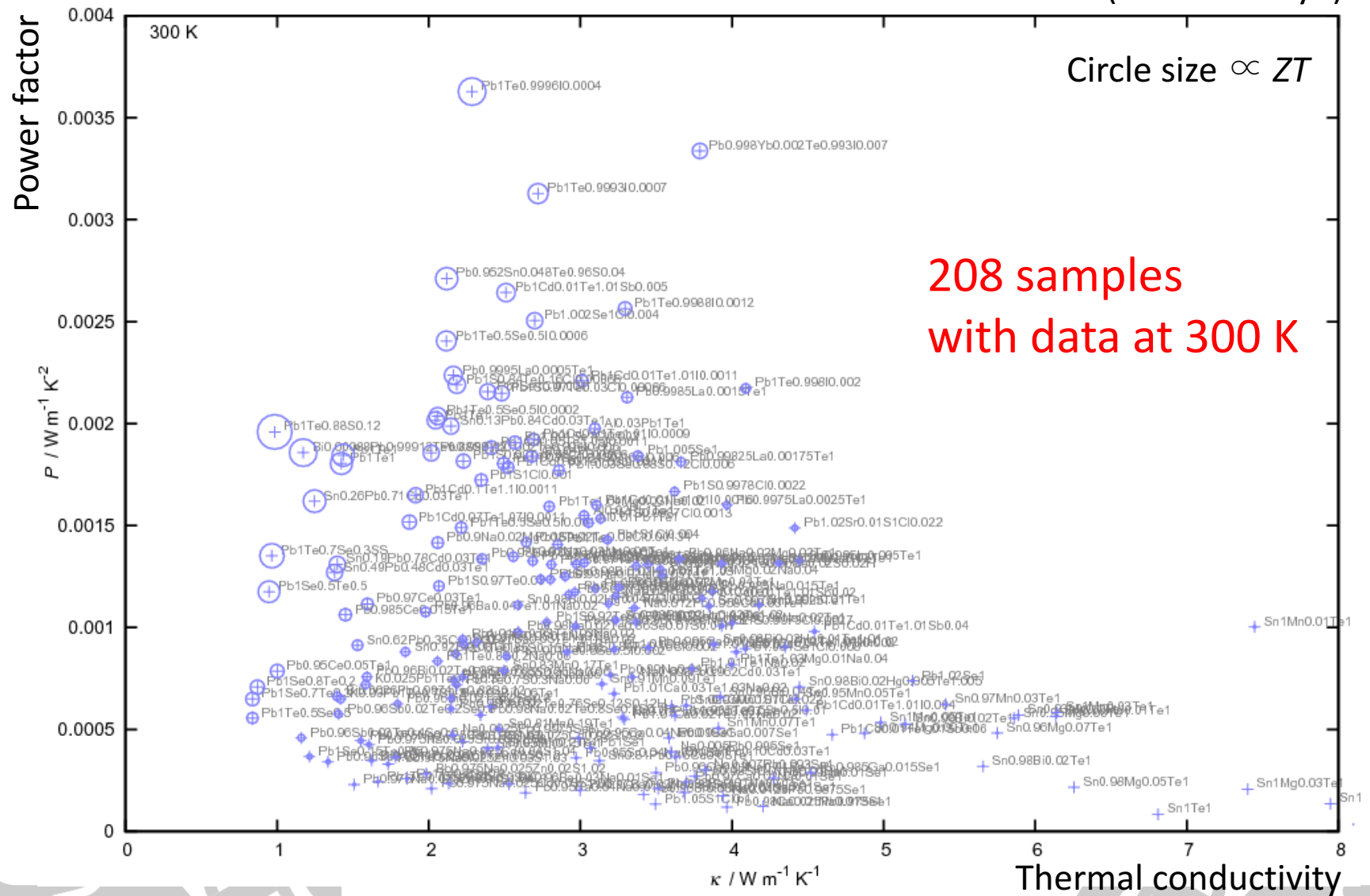
copy the digitized data



⑧ Paste the digitized data to Starrydata2 → Save

Recorded samples of PbTe-type thermoelectric materials

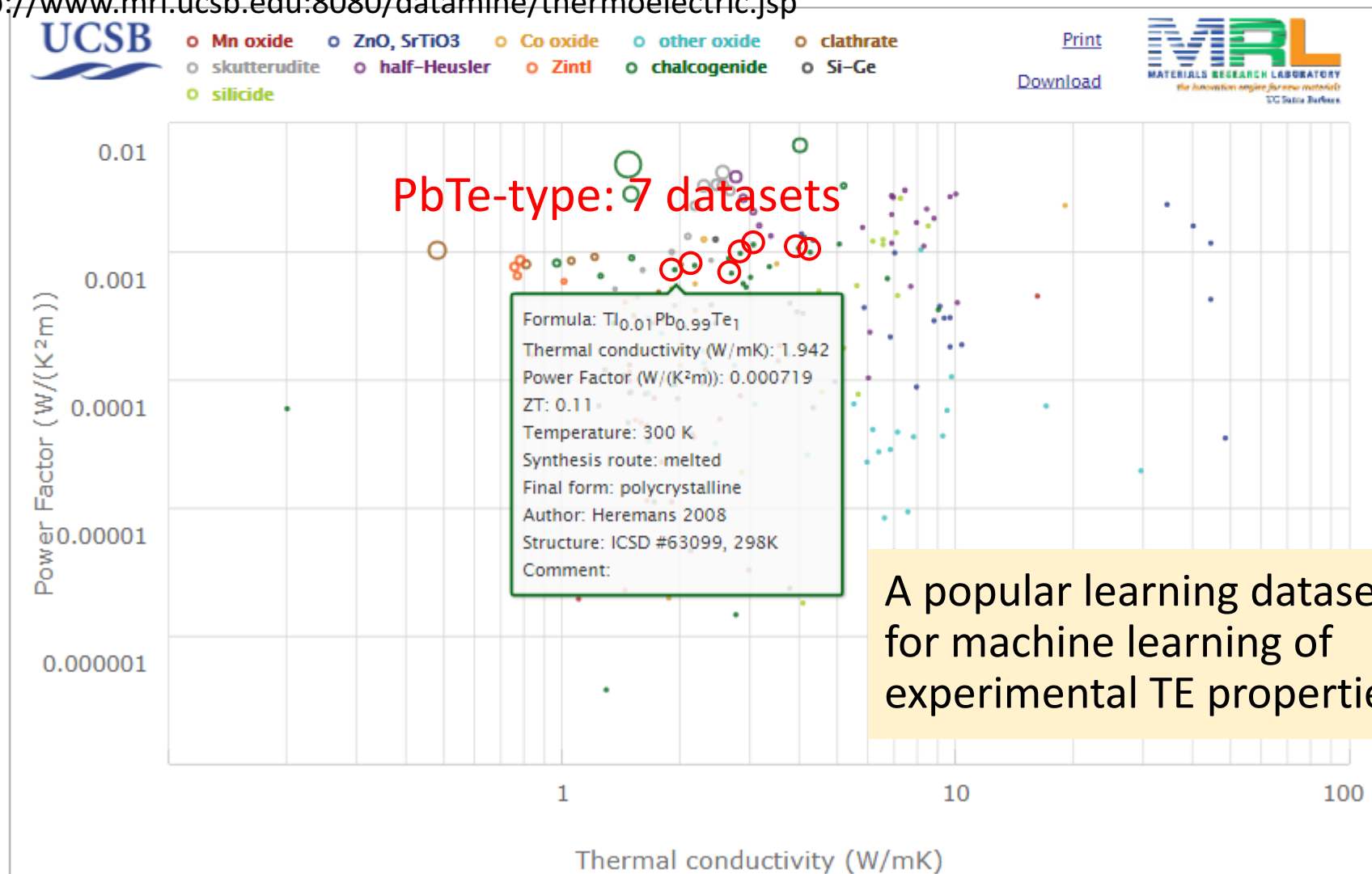
Data collection: Takushi Kodani (Univ. of Tokyo)



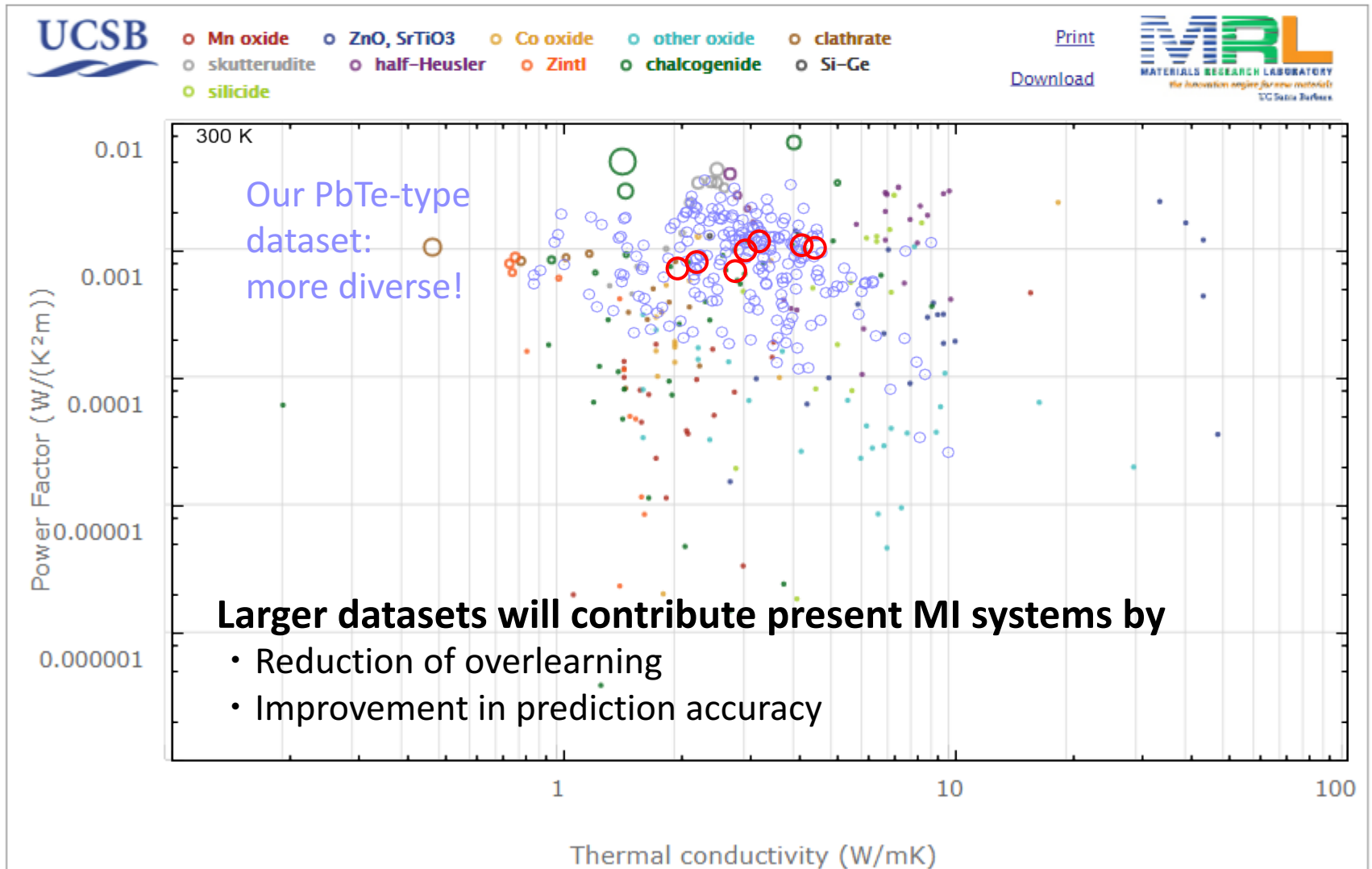
Comparison to present DB: UCSB-MRL datamining chart

M.W. Gaultois et al., Chemistry of Materials, 25 (2013) 2911.

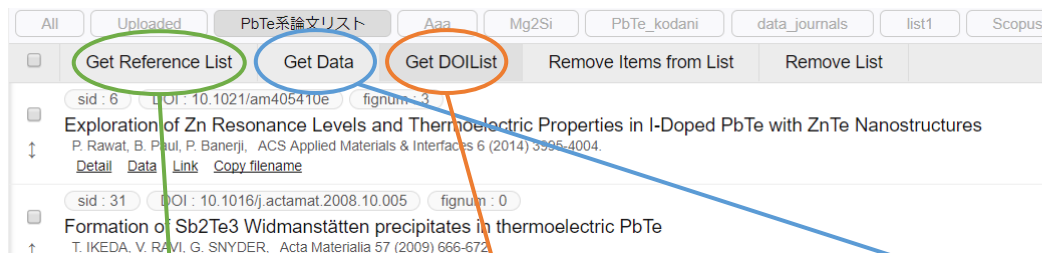
<http://www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp>



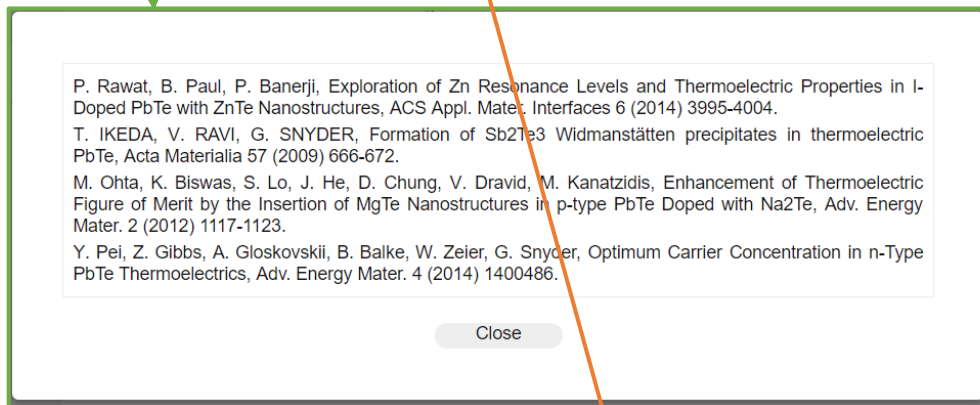
Comparison of our dataset with UCSB-MRL datamining chart



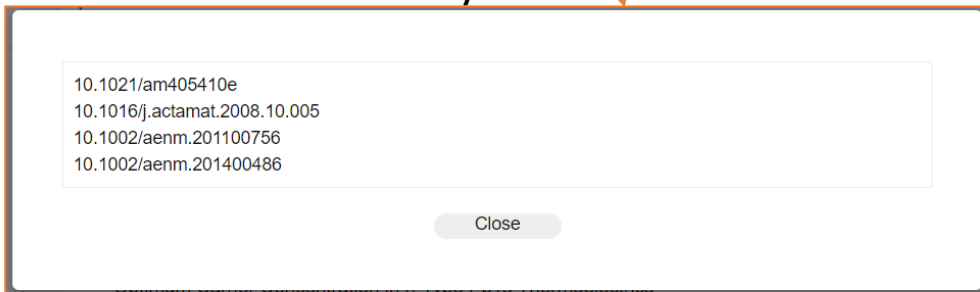
Starrydata 2 web system: Data export functions



Reference list to write papers



DOI list to share the mylist



A JSON file containing all datasets in the mylist

```
...  
{'figureid': 14, 'paperid': 2, 'propertyid_x': 0,  
'propertyid_y': 1, 'sampleid': 17, 'x':  
1003.2921810699587, 'y': -  
7.749077490774911e-05}, {'figureid': 14,  
'paperid': 2, 'propertyid_x': 0, 'propertyid_y': 1,  
'sampleid': 18, 'x': 300.4115226337448, 'y': -  
0.00011881918819188195}, ...],  
'sample': [  
{'composition': 'Ba24Ga0Ag0Ge100', 'paperid': 0,  
'sampleid': 0, 'samplename': 'x=0,y=0'},  
{'composition': 'Ba24Ga4Ag0Ge96', 'paperid': 0,  
'sampleid': 1, 'samplename': 'x=4,y=0'},  
...  
...
```

clathrate.json

149 papers / 623 figures / 890 samples

73281 data points

File size: 10 MB / Download time: ~3 min

Reading Starrydata2 rawdata (all numeric) data on Python

clathrate.json (10 MB)

```
import json
import pandas as pd
```

```
f=open('clathrate.json','r')
dict_all=json.load(f)
```

```
df_rawdata=pd.DataFrame(dict_all["rawdata"])
df_paper=pd.DataFrame(dict_all["paper"])
df_figure=pd.DataFrame(dict_all["figure"])
df_sample=pd.DataFrame(dict_all["sample"])
df_property=pd.DataFrame(dict_all["property"])
```

	figureid	paperid	propertyid_x	propertyid_y	sampleid	x	y
73272	623	145	0	3	904	820.619060	1.257546
73273	623	145	0	3	904	837.799285	1.250562
73274	623	145	0	3	904	869.048740	1.250039
73275	623	145	0	3	904	894.034353	1.236228
73276	623	145	0	3	904	915.936873	1.262647
73277	623	145	0	3	904	933.138024	1.275752
73278						9259	1.295658
73279						1913	1.315485
73280						0148	1.335391
73281						0747	1.355166

sampleid	composition	paperid	sampleid	sampleid	sampleid	sampleid	sampleid
884	Ba8Cu6Si16Ge24	140	884	884	4GPa		
885	Ba8Cu6Si16Ge24	140	885	885	5GPa		
886	Ba8Cu6Si40	141	886	886	Ba8Cu6Si40		
887	Ba8Cu6Ge20Si20	141	887	887	Ba8Cu6Ge20Si20		
888	Ba8Cu6Ge8Si32	142	888	888	x=0		
889	Ba8Cu6Ge16Si16	142	889	889			
890	Ba8Cu6Ge16Si16	142	890	890			

paperid	author	author_full	doi	issue	journal	journal_full
145	A. Saramat, G. S...	A. Saramat, G. S...	10.1063/1.2163979	2	Journal of Appli...	Journal of Appli...
146	J. Martin, S. Er...	J. Martin, S. Er...	10.1063/1.2171775	4	Journal of Appli...	Journal of Appli...
147	Chr...					
148						
149	A.					

figureid	caption	figureid	figurename	paperid
619	Total thermal conductivity of purified Ba8Ga16...	619	5	144
620	Figure of merit ZT of purified Ba8Ga16Ge30 sam...	620	6	144
621	a Seebeck coefficient S as a function of t...	621	2(a)	145
622	2. a Seebeck coefficient S as a function o...	622	2(b)	145
623	a Seebeck coefficient S as a function of t...	623	2(c)	145

propertyid	propertyid	propertyname	unit
0	0	Temperature	K
1	2	Electrical conductivity	ohm ⁽⁻¹⁾ *m ⁽⁻¹⁾
2	1	Seebeck coefficient	V*K ⁽⁻¹⁾
3	5	Power factor	W*m ⁽⁻¹⁾ *K ⁽⁻²⁾
4	3	Thermal conductivity	W*m ⁽⁻¹⁾ *K ⁽⁻¹⁾
5	8	Dimensinoless figure of merit	-
6	4	Electrical resistivity	ohm*m
7	7	Inverse temperature	K ⁽⁻¹⁾

7 7 Inverse temperature K⁽⁻¹⁾

Possible collaborations

To use the electronic journal systems, the project has to be **non-commercial**.

Current project (TEdb):

Research budgets
MEXT, JST etc.

RIKEN

NIMS

Research assistants

For greater database projects:

Donation from industries
Datasets of interest can be requested

Donation
Univ. of Tokyo
Project manager (Our team + α)

Payment
Student workers

\sim ¥1,000/paper
Research assistants

Papers
Collaborators from industries

Data

Data

Data

Data

New data is added to Starrydata2 (public to everyone)

Summary

Starrydata2 web system (FREE)

- For efficient collection and sharing of data from plot images on literature
- Includes data export system for materials informatics

Prototype database: TEdbProject (FREE)

- Sample-based database of experimental thermoelectric properties
- 200~500 samples per material family (Largest experiential dataset)

URL: <http://starrydata2.org>

for first use: <http://starrydata2.org/signup>

Manual page: <https://sites.google.com/site/yukarisearch/starrydata>

E-mail: katsura@phys.mm.t.u-tokyo.ac.jp