DESENVOLVIMENTO DE BASE DE DADOS PARA MACHINE LEARNING APLICADO À SELEÇÃO DE LIGAS PARA BRASAGEM

Débora van Putten e Guilherme Dariani

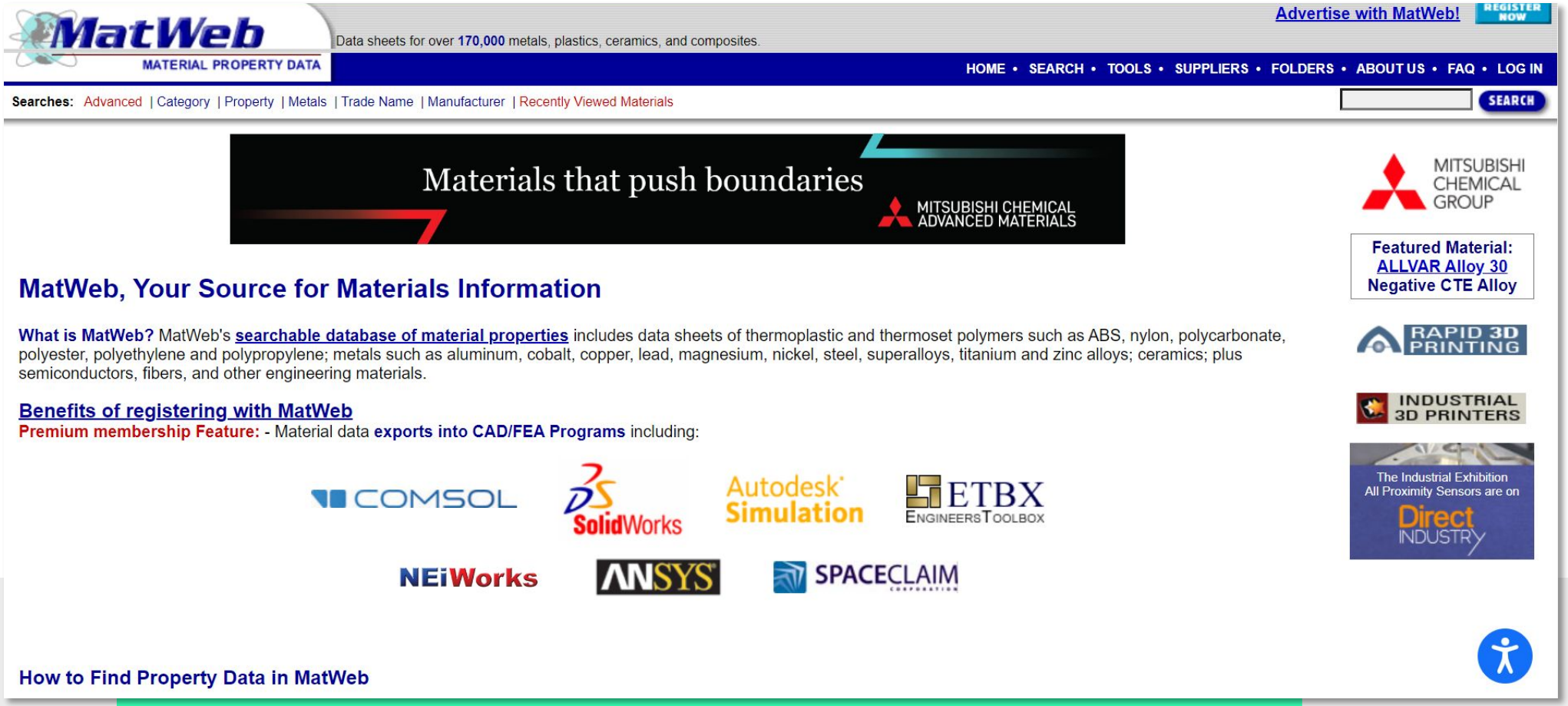
Iniciação à Pesquisa II — 2023.1

ilum escola de ciência · CNPEM Centro Nacional de Pesquisa em Energia e Materiais · MINISTÉRIO DA EDUCAÇÃO · MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO · GOVERNO FEDERAL BRASIL UNIÃO E RECONSTRUÇÃO

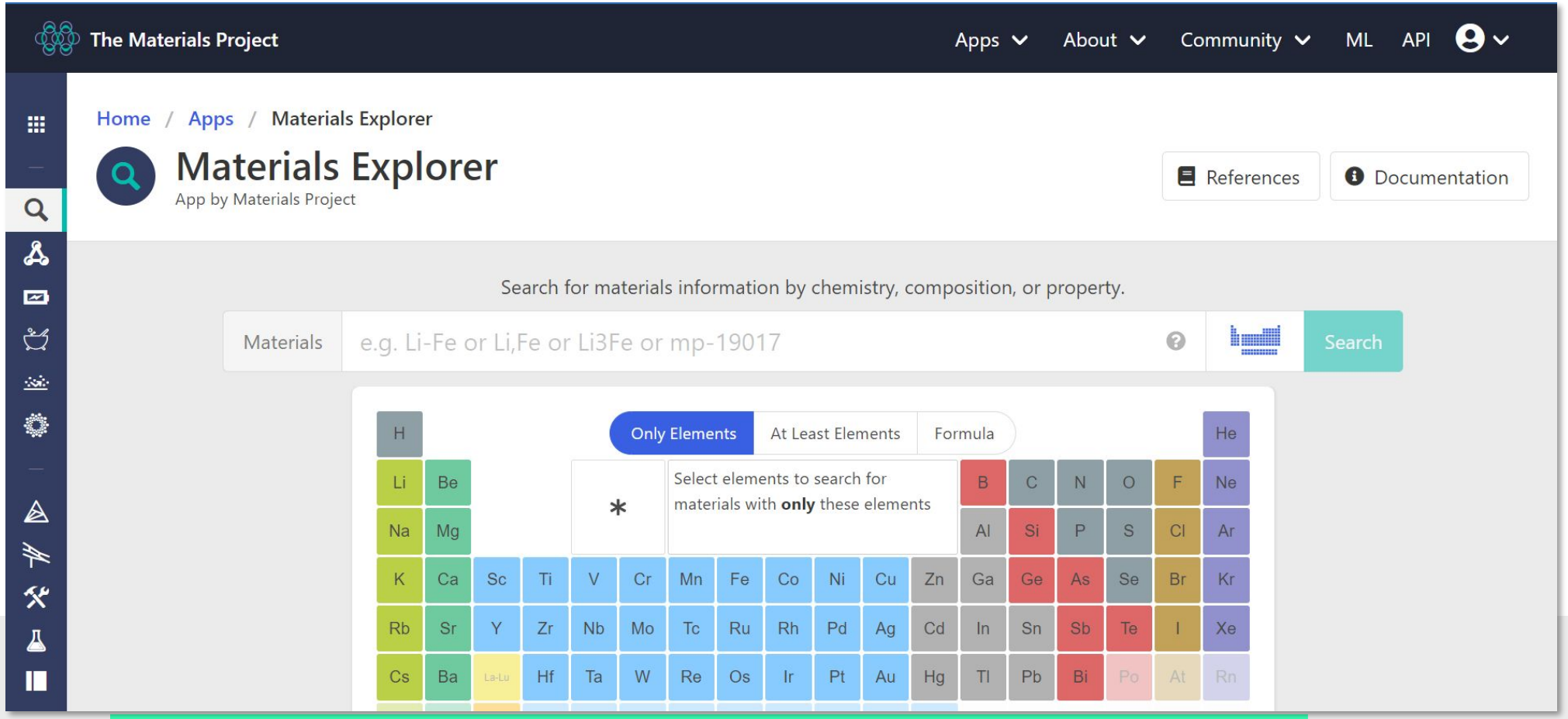
Resumo

- Bases de dados;
- MatMiner:
  - Datasets:
    - Quais planejamos usar e motivos;
    - Quais não planejamos usar e motivos;
- Problemas.

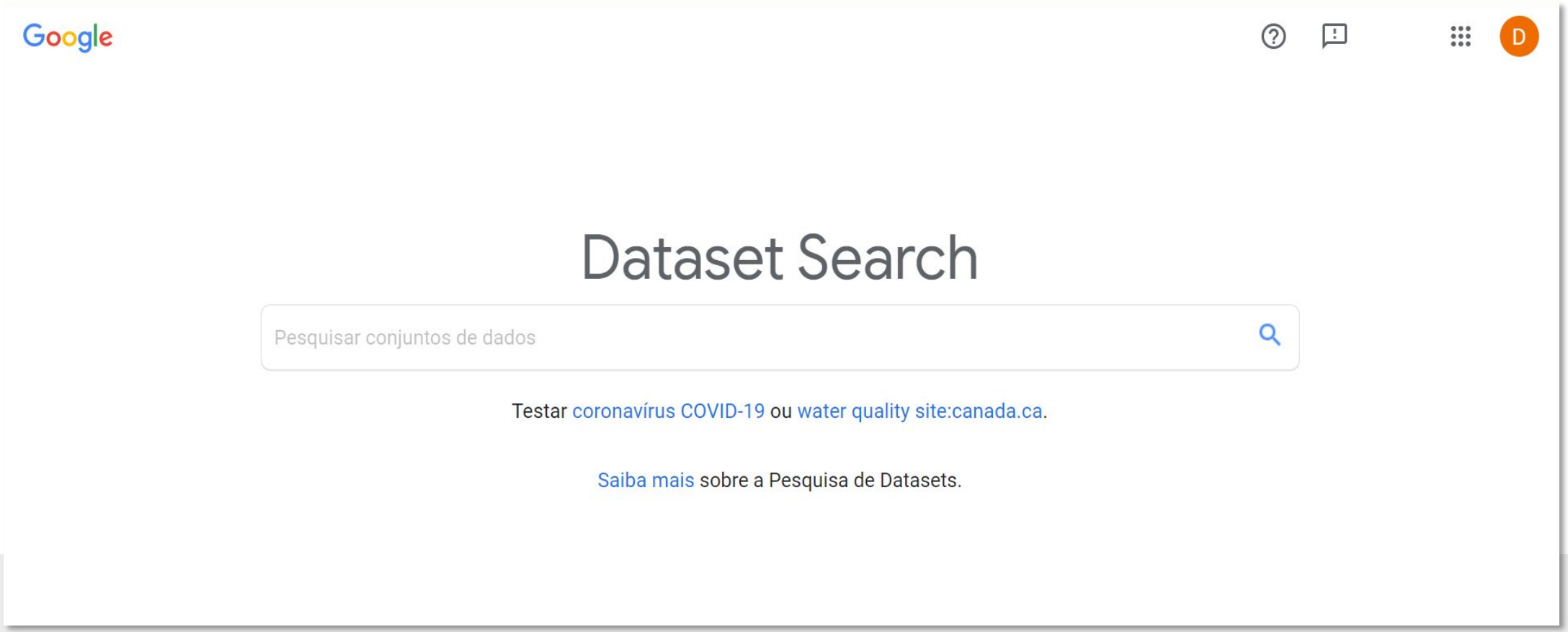
Bases de Dados



Bases de Dados



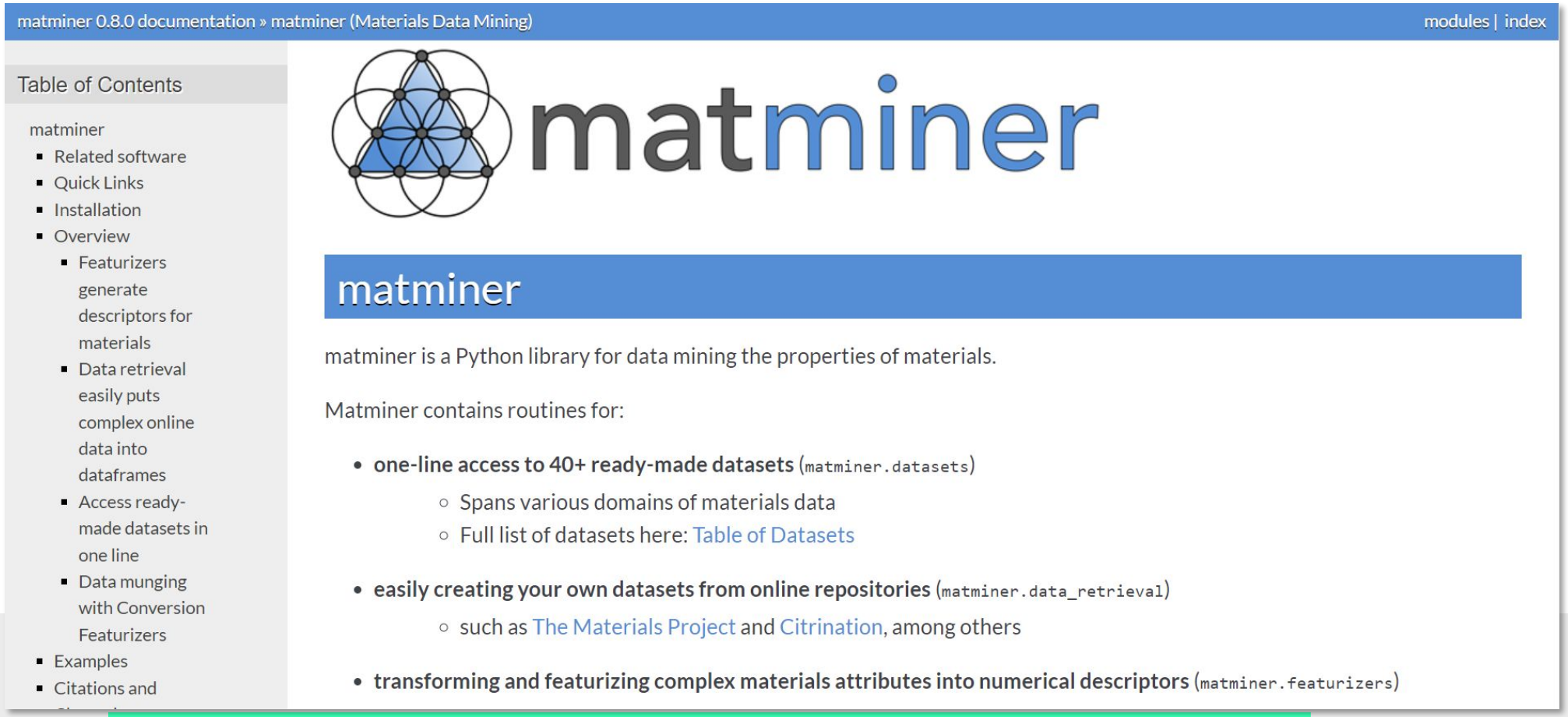
Bases de Dados



Bases de Dados

MatWeb	Materials Project	Google Dataset
Famosa	Famosa também	Muitos dados
Feedbacks ruins	API própria	Dados bem específicos
Interface complicada	Interface simples	Licença varia
Licença limitada	MatMiner entende ela	

MatMiner



matminer 0.8.0 documentation » matminer (Materials Data Mining) modules | index

Table of Contents

matminer
- Related software
- Quick Links
- Installation
- Overview
  - Featurizers generate descriptors for materials
  - Data retrieval easily puts complex online data into dataframes
  - Access ready-made datasets in one line
  - Data munging with Conversion Featurizers
- Examples
- Citations and

matminer

matminer

matminer is a Python library for data mining the properties of materials.

Matminer contains routines for:

- **one-line access to 40+ ready-made datasets** (`matminer.datasets`)
  - Spans various domains of materials data
  - Full list of datasets here: Table of Datasets
- **easily creating your own datasets from online repositories** (`matminer.data_retrieval`)
  - such as The Materials Project and Citrination, among others
- **transforming and featurizing complex materials attributes into numerical descriptors** (`matminer.featurizers`)

MatMiner

Table of Contents

Table of Datasets

Find a table of all 45 datasets available in matminer here.

Name	Description	Entries
boltztrap_mp	Effective mass and thermoelectric properties of 8924 compounds in The Materials Project database that are calculated by the BoltzTraP software package run on the GGA-PBE or GGA+U density functional theory calculation results	8924
brgoch_superhard_training	2574 materials used for training regressors that predict shear and bulk modulus.	2574
castelli_perovskites	18,928 perovskites generated with ABX combinatorics, calculating gllbsc band gap and pbe structure, and also reporting absolute band edge positions and heat of formation.	18928
citrine_thermal_conductivity	Thermal conductivity of 872 compounds measured experimentally and retrieved from Citrine database from various references	872
dielectric_constant	1,056 structures with dielectric properties, calculated with DFPT-PBE.	1056
double_perovskites_gap	Band gap of 1306 double perovskites (a_1-b_1-a_2-b_2-O6) calculated using Gritsenko, van Leeuwen, van Lenthe and Baerends potential (gllbsc) in GPAW.	1306

MatMiner

DATASETS QUE PODEM SER ÚTEIS:

boltztrap_mp	Effective mass and thermoelectric properties of 8924 compounds in The Materials Project database that are calculated by the BoltzTraP software package run on the GGA-PBE or GGA+U density functional theory calculation results
citrine_thermal_conductivity	Thermal conductivity of 872 compounds measured experimentally and retrieved from Citrine database from various references
ucsb_thermoelectrics	Database of ~1,100 experimental thermoelectric materials from UCSB aggregated from 108 source publications and personal communications

MatMiner

DATASETS QUE NÃO PRETENDEMOS USAR:

`flla`	3938 structures and computed formation energies from “Crystal Structure Representations for Machine Learning Models of Formation Energies.”	3938
`glass_binary`	Metallic glass formation data for binary alloys, collected from various experimental techniques such as melt-spinning or mechanical alloying	5959
`glass_binary_v2`	Identical to glass_binary dataset, but with duplicate entries merged	5483
`glass_ternary_hipt`	Metallic glass formation dataset for ternary alloys, collected from the high-throughput sputtering experiments measuring whether it is possible to form a glass using sputtering	5170
`glass_ternary_landolt`	Metallic glass formation dataset for ternary alloys, collected from the “Nonequilibrium Phase Diagrams of Ternary Amorphous Alloys,’ a volume of the Landolt– Börnstein collection	7191

Problemáticas

- Não temos a informação de qual material é serve ou não para ser um material de soldagem, e para qual material base;
- Não entendemos muito bem a formatação para a criação do banco de dados.

OBRIGADA

Débora van Putten
debora220037@ilum.cnpem.br

Guilherme Dariani
guilherme220044@ilum.cnpem.br

ilum.cnpem.br

ilum

CNPEM
Centro Nacional de Pesquisa em Energia e Materiais

MINISTÉRIO DA EDUCAÇÃO

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO