Python scientific computing tutorials

Presenter: Lai Wei
lw36@email.sc.edu

UNIVERSITY OF
South Carolina

Overview of Today's Topics

Key Tools and Data Types:

- **Pymatgen:** A core Python library for materials analysis, especially for creating and manipulating materials' structural data.
- **Composition Data:** Information about the elemental composition of materials, crucial for predicting properties and behavior.
- **Structure Data:** Describes the three-dimensional arrangement of atoms within materials, essential for understanding material properties.
- **VESTA:** Visualization software for viewing and analyzing three-dimensional crystallographic data and models.
- **Matminer:** An open-source library for extracting and manipulating materials data, facilitating advanced materials data mining.

What is materials informatics?

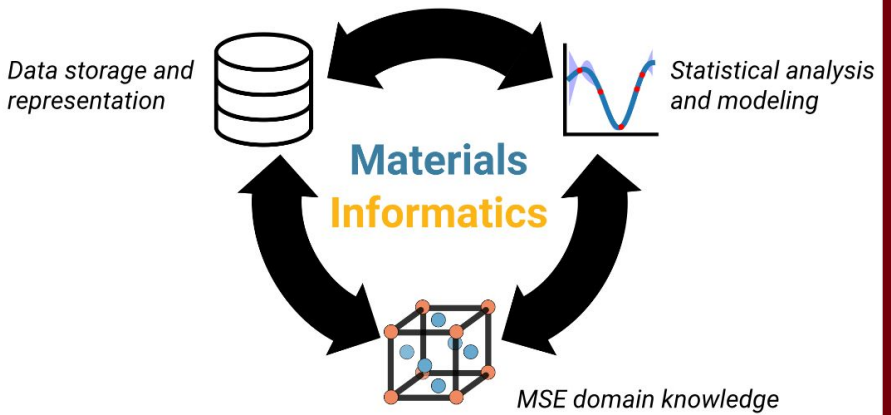
Materials informatics is an exciting field where we use computers and lots of data to explore and create new materials. It's like combining the study of materials with the power of computers to make amazing discoveries faster than ever before!

Imagine you're trying to build the perfect skateboard. You want it to be super strong, but also lightweight. In the past, scientists and engineers would have to test many different materials one by one, which could take years.

But with materials informatics, we can:

- Use big data and machine learning to study materials faster
- Predict new materials before making them in real life
- Figure out properties like strength or conductivity using simulations

Data-driven materials science → materials informatics (MI)



UNIVERSITY OF South Carolina

Why we need materials informatics?

Discovering New Materials:

- Helps scientists find new materials faster.
- Allows us to design materials with specific properties, like super strength or conductivity.

Saving Time and Money:

- Reduces the time it takes to develop new materials.
- Saves costs by minimizing trial and error in experiments.

Understanding Properties:

- Helps us understand why materials behave the way they do.
- Predicts how materials will perform in different conditions.

Environmental Impact:

- Enables development of eco-friendly materials.
- Reduces waste by improving material efficiency.

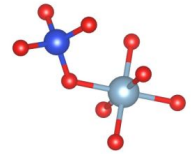
Facilities and physical lab space required $$

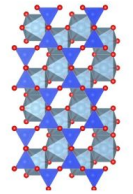
Materials need to be purchased $$

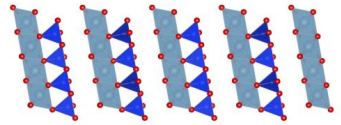
Samples need to be synthesized ⏳⏳

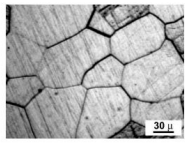
High chance of failure ⏳⏳

Diverse reasons for failure ⏳⏳









UNIVERSITY OF South Carolina

Open-Source Materials Informatics Resources

Materials Project

- Extensive database of material properties and interactive tools.
- Website: materialsproject.org

MatWeb

- Free database of material property data for metals, plastics, ceramics, and more.
- Website: matweb.com

Materials informatics resources

- Computational materials workflow libraries: Pymatgen, ASE
- Materials informatics libraries: Matminer, MEGNet
- Databases: Materials Project, OQMD, ICSD
- Visualization, data sharing, and more for materials discovery.
- Website: Materials informatics resources on GitHub

Pymatgen

- **What is pymatgen?**
- Pymatgen (Python Materials Genomics) is the code that powers all of the scientific analysis behind the Materials Project. It includes a robust and efficient libraries for the handling of crystallographic structures and molecules, in addition to various mathematical and scientific tools for the handling and generation of materials data. For more details, see the pymatgen website.
- Imagine pymatgen as a super-smart digital toolbox for materials scientists:
  - **Structure Builder**: Like a virtual Lego set, it helps you build and manipulate atomic structures.
  - **Property Calculator**: Acts as a crystal ball to predict material properties without actual experiments.
  - **Data Analyzer**: Works like a detective, finding patterns and insights in large sets of materials data.
  - **Simulation Helper**: Prepares and processes data for complex computer simulations of materials.

Pymatgen

- **How do I install pymatgen?**
- pymatgen can be installed via pip:

```
pip install pymatgen
```

or conda:

```
conda install pymatgen
```

- Verify we have pymatgen installed

```
import pymatgen.core
```

We can show the specific version of pymatgen installed:

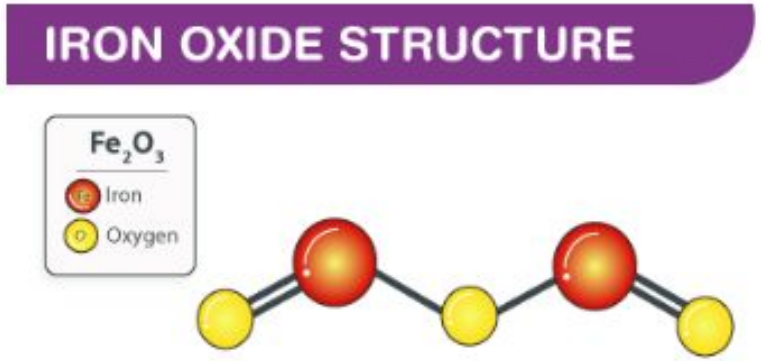
```
print(pymatgen.core.__version__)
```

Pymatgen: Composition data

- **Composition data refers to the chemical makeup of a material. It's a fundamental concept that describes what elements are present in a material and in what proportions.**

Example of Composition Data

- **Iron Oxide (Fe2O3)**:
  - Contains **2 parts Iron (Fe)**.
  - Contains **3 parts Oxygen (O)**.
  - Commonly known as rust, used in various industrial applications.



Key Operations with Composition Data (Using Pymatgen)

1. **Creating a Composition**: Instantiate composition from a formula.
   - `comp1 = Composition("Fe2O3")`
2. **Accessing Elements**: Retrieve the types of elements and their proportions.
   - `Elements in comp1: {comp1.elements}`
3. **Analyzing Properties**: Compute molecular weight, check for metallicity, etc.
   - `Weight of comp1: {comp1.weight:.2f} g/mol`

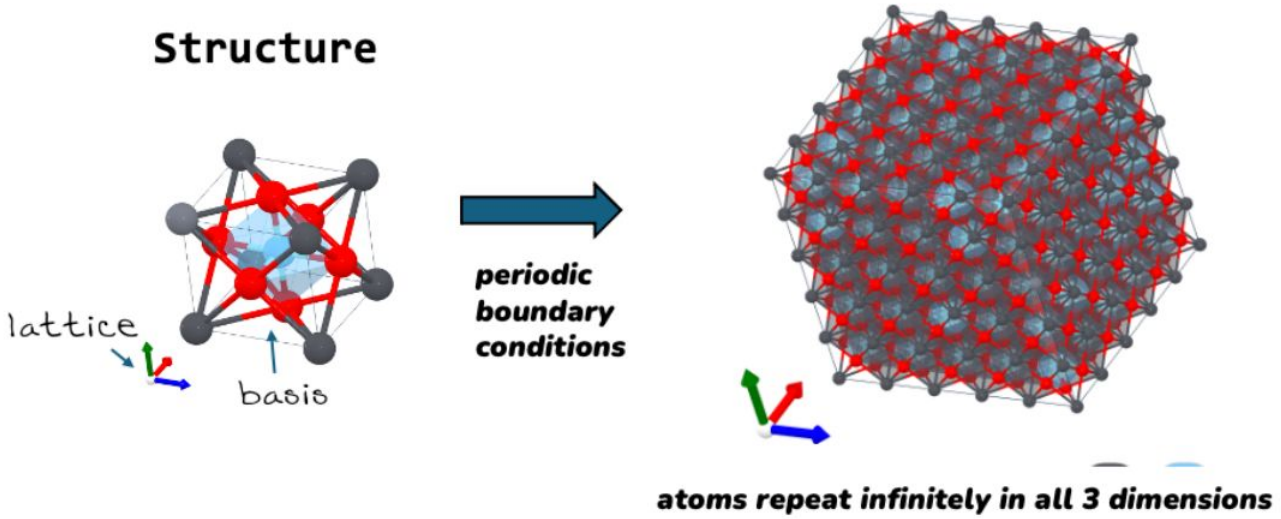
Pymatgen: Structure data

What is Structure Data?

- Describes the arrangement of atoms within a material.

What is CIF Data?

- CIF stands for Crystallographic Information File.
- **It's a type of structure data specifically formatted to include details about the positions, types of atoms, and their bonding within a lattice.**



Why is this Important?

- **Predicting Material Properties:** Fundamental for engineering materials with desired mechanical, electronic, and optical behaviors.
- **Research and Development:** Helps scientists and engineers reduce trial and error, saving time and resources.

A **supercell** is a larger version of a crystal unit cell. Imagine a single building block **(unit cell)** of a crystal.
A supercell is made by repeating this block infinite times in different directions.

UNIVERSITY OF South Carolina

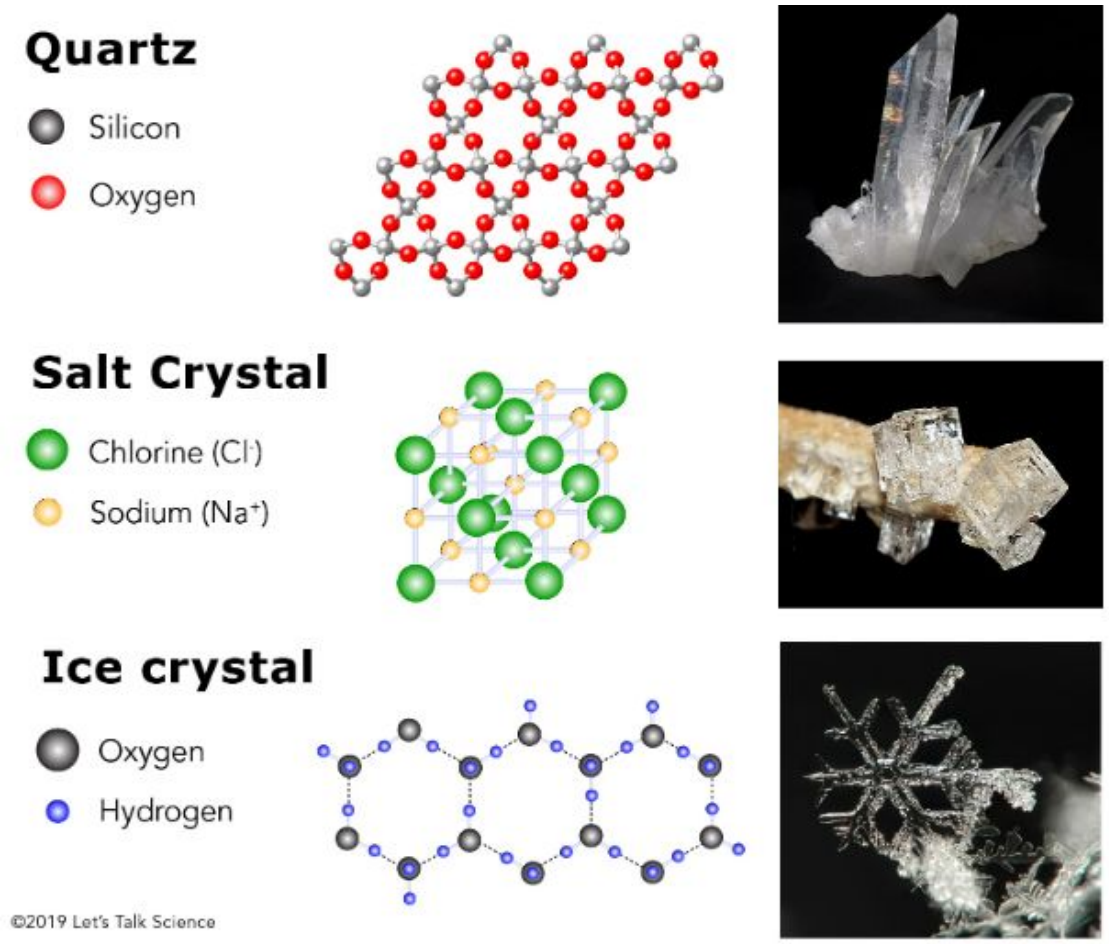
VESTA: Visualize Crystal Structures from cif files

What is VESTA?

- VESTA is a powerful software for visualizing crystal structures from CIF data files.
- It's widely used in materials science for displaying three-dimensional atomic arrangements of crystals.

Where to Download VESTA:

- Free download available at: VESTA Download



UNIVERSITY OF South Carolina

Visualize Structures using python code

Visualizing crystal structures from CIF files using code: visualize structures

How to use it?

The script accepts various command-line arguments to customize the visualization of CIF files, such as the type of visualization, resolution, and supercell configuration.

```
        parser.add_argument('-i', '--input', type=Path, help='folder to get CIFs from')
        parser.add_argument('-s', '--style', choices=['spacefilling', 'ballstick'], help='structure drawing type',
    default='spacefilling')
        parser.add_argument('-o', '--output', type=Path, help='folder to put drawings in', default='images')
        parser.add_argument('-c', '--cell', type=scale_parse, help='supercell: format either as 3 or 3,3,3',
    default='1')
        parser.add_argument('-r', '--rez', type=int, help='resolution (0-10): more takes longer but is better',
    default='10')
```

for example:

```
# Navigate to the script's directory
cd path/to/script
# Run the script with specified parameters python
visual_cif.py --input /home/weilai/SUMMER-CAMP/cifs/ --style spacefilling --output ~/images --cell 1
--rez 10
# Check the output image
"Image saved as ~/images/SrTiO3.png"
```

UNIVERSITY OF South Carolina

Pymatgen: Structure data

Creating a Structure

- **Example**: Define a cubic lattice and create a structure for Iron Oxide.

```
- lattice = Lattice.cubic(4.2)  # lattice parameter in Ångström
- structure = Structure(lattice, ["Fe", "O"], [[0, 0, 0], [0.5, 0.5, 0.5]])
```

Accessing and Modifying Structure Information

- Retrieve information about lattice, sites, and atoms.
  - `Number of sites: len(structure)`
  - `Species at first site: structure[0].species`
- Modify the structure (e.g., replace an atom, scale the lattice).
  - `structure.replace(0, 'Mg') # Replace Fe with Mg at the first site`

Pymatgen: Structure data

You can also load a Structure from a CIF file:

- Download structure file from Materials Project website: Materials Project

- `PTO_structure_from_cif = Structure.from_file("files/PbTiO3.cif")`
- `print(PTO_structure_from_cif)  # print the structure - note the oxidation states!`
- `PTO_structure_from_cif`

Pymatgen: Structure data

Loading a structure from an API (Materials Project)

- Let's load a crystal structure from an online database (The Materials Project)
- Using the API will give you instant access to >150,000 crystal structures that are available on the Materials Project web site
- Materials Project website: Materials Project

Pymatgen: Structure data

Analyzing Structure Data

Neighbor Analysis

- Identify neighboring atoms within a certain radius.
  - `neighbors = structure.get_neighbors(structure[0], 2) # 2 Å radius`

Symmetry and Properties

- Determine the space group, symmetry operations, and calculate the volume of the unit cell.
  - `Space group: structure.get_space_group_info()`
  - `Volume of unit cell: structure.volume`

Visualizing Structures

- Tools and techniques to visualize complex crystal structures for better understanding and analysis.
  - Suggested software: **VESTA**, CrystalMaker, python code, etc.

Pymatgen: Structure data

- **Structure Data Examples: What can you do directly with a Structure?**
  1. Creating Structure objects
  2. Accessing structure information
  3. Site-specific information
  4. Neighbor analysis
  5. Structure properties
  6. Symmetry and system properties

Matminer

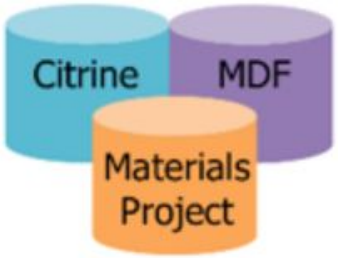
What is Matminer?

- Matminer is an open-source library for performing materials data mining.
- It facilitates the extraction of materials data from various sources, manipulation of materials data, and running data mining algorithms on materials data.

Matminer

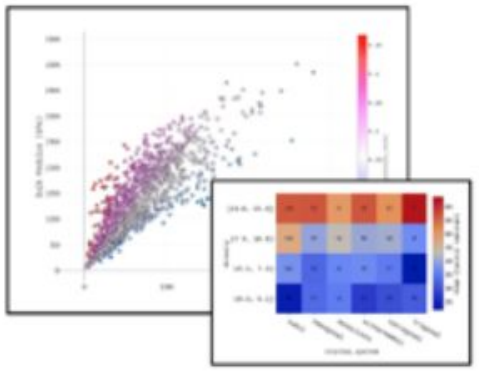
matminer

Data Retrieval → Feature Extraction → Visualization



Materials Databases

MATERIAL	FEATURES				PROPERTY
TiO_2 rutile	F_{11}	F_{12}	...	F_{1N}	gap = 3.0 eV
C diamond	F_{21}	F_{22}	...	F_{2N}	gap = 5.5 eV
...	...	...	...	...	...
PbTe rocksalt	F_{M1}	F_{M2}	...	F_{MN}	gap = 0.3 eV



scikit learn

Machine Learning

K Keras

UNIVERSITY OF South Carolina

Matminer

Installation

```
pip install matminer
```

Importing Essential Libraries

```
from matminer.data_retrieval.retrieve_MP import MPDataRetrieval
from matminer.featurizers.composition import ElementProperty
from sklearn.ensemble import RandomForestRegressor
```

Matminer

Feature Engineering

- Extracting meaningful features from material compositions:

```
# Add features based on element properties
ep_feat = ElementProperty.from_preset(preset_name="magpie")
df = ep_feat.featurize_dataframe(df, col_id="composition")
```

Machine Learning

- Predicting material properties using Random Forest:

```
# Prepare features and target
X = df.drop(['band_gap'], axis=1)
y = df['band_gap']

# Train a Random Forest model
model = RandomForestRegressor(n_estimators=100)
model.fit(X, y)

# Display model accuracy
print(f"Model Accuracy: {model.score(X, y)}")
```

Matminer

- **Matminer Examples: Data Mining in Materials Science**

  1. Loads the "elastic_tensor_2015" dataset.
  2. Displays the first few rows and column names of the dataset.
  3. Shows properties of the first material in the dataset, including its formula, number of sites, volume, and elastic properties.
  4. Finds the material with the highest bulk modulus.
  5. Calculates and prints average properties for all materials.
  6. Creates a scatter plot of bulk modulus vs shear modulus.
  7. Creates a histogram of the number of sites in the materials.

Matminer Tutorials - The Matminer tutorial series uses interactive Jupyter notebooks to walk through the basic steps of materials informatics: data retrieval, featurization, machine learning, and data evaluation/visualization.

UNIVERSITY OF South Carolina

THANKS!

Name: Lai Wei

Email: lw36@email.sc.edu

UNIVERSITY OF
South Carolina