论文第六章中提示工程-AMDL

本文件中仅展示与AZM不同的提示工程。

1. 生成脚本的提示工程

1.1 生成以DOI号重命名PDF文献文件名的脚本

You are an expert in coding. Help me generate a Python script. The script needs to contain the following parts:

1. Extract the DOI of papers. On the first page of the PDF, there is a string “doi.org/10.1021/acsaem.2c01653”. The script should extract the string “10.1021/acsaem.2c01653” first. Then the “.” in the string should be replaced with “\_”.

2. File rename. The PDF files in the folder should be renamed using the DOI extracted in the first step. Note that the files are all in one folder and are all PDF files.

I have uploaded one PDF for you to analyze. Should you have any question, ask me in your response.

1.2 生成将PDF转换为图片的脚本

You are an expert at coding. Generate me a Python script to convert PDF to figures. The requests are as follows:

1. There are multiple PDF files, the script should store the figures in different folders: figures converted from one PDF in one folder. The name of the folder is the name of the source PDF file, and the name of the figures is the page number.

2. The figures should be in jpg format with 300 dpi.

Should you have any question, ask me in your response.

1.3 生成使用GPT-4挖掘文献图片中科研信息的脚本

You are an expert in coding. Help me generate a Python script. I have 1000 PDF files, which are all converted to image files. The image files are stored in separated folders, one folder for one source PDF file. The script needs to contain the following parts:

1. Image read. The script should read all images in one folder. Then, the images should be encoded to the format as the requirement of OpenAI API. The output of this step should be a data frame with three columns, one data frame for one source folder. The first column is “PaperName”. The value of this column is the folder name from which the images are read. The second column is “PageNumber”. The value of this column is the name of the image file. The third column is “Image”. The value of this column is the image file after encoding.

2. Information extraction. All images in one data frame should be input to OpenAI API together with a prompt at the same time. The model used in API is ‘gpt-4-turbo’. The output of this step should be a data frame with four columns. The first to the third column is the same as the output of the first step. The last column is “Response”. Its content is the content of the response of the API.

3. Results output. The output data frame of Step 2 should be output to a file in JSON format.

Should you have any questions or find any aspects of this prompt unclear, please include your inquiries in your response.

1.4 生成从GPT-4对文献图片挖掘结果中读取Markdown表格的脚本

You are an expert at coding. Your job is to write a Python script aiming to extract information of lithium metal oxide that is used as lithium-ion battery cathode from the given JSON file. The script should realize the following functions:

1. Extract the markdown tables in all 'Response' keys. Note that I need only the content of the markdown tables.

2. Add one column named "File name of the source PDF" to the table extracted, fill this column with the “PaperName” column of the file from which the table is extracted.

3. Merge all markdown tables to one table

I have provided four JSON files for you to analyze.

1.5 生成辅助对LCD-II中数据条目进行分类的脚本

You are an expert at coding. Revise the script to make it fulfill the functions below:

1. Revise the source of unique value from "LMO name" column to B column ("Variable Name" column).

2. Revise the option from 0 and 1 to 1 to 6. And revise the tip information from "Should ..." to:

"Where should the entry go?

1 - Structure;

2 - Property;

3 - Synthesis;

4 - Characterization;

5 - Calculation;

6 - Other."

3. Revise the destination files. The script should check if there are the following six Excel files in the path: Structure information.xlsx, Property information.xlsx, Synthesis information.xlsx, Characterization information.xlsx, Calculation information.xlsx and Other information.xlsx. If not, create them. If user inputs 1, copy the entry to Structure information.xlsx; if 2, copy the entry to Property information.xlsx; if 3, copy the entry to Synthesis information.xlsx; if 4, copy the entry to Characterization information.xlsx; if 5, copy the entry to Calculation information.xlsx; if 2, copy the entry to Other information.xlsx.

Should you have any question, ask me in your response.

*Attached file: LMOIrreleventMaterialCleaning.py*

1.6 生成从MP数据库中获取材料信息的脚本

You are an expert at coding. Generate a Python script for me. The script should have the function below:

Material information retrieve. The script should first access the API of *The Materials Project*. Then it should search materials that contains elements including lithium, oxygen and other metal elements, and exclude all non-metal elements. The result should contain mp-id, material composition and c-axis length of the material and output in a JSON file.

1.7 生成将MP数据库中材料信息的下载结果读取至Excel文件中的脚本

You are an expert at coding. Generate a Python script for me. The script should have the function below:

1. Load JSON file. The script should load the JSON file that I upload to you.

2. Information organization and output. The script should organize the information and output to an Excel file just like the Excel file I upload to you. The “mp-id” column in the JSON file should fill in the “MP ID” column in the Excel file. The “c-axis length” column in the JSON file should fill in the “c-axis length” column in the Excel file. The “composition” column in the JSON file should be separated by space before fill in the Excel file. For example, the content in “composition” column is "Li2 Co2 Sn2 O8", by separating with space, it is separated into four parts: “Li2”, “Co2”, “Sn2” and “O8”. These parts are all consists of letters and numbers. The letters should be filled in the column begin with “E” in the Excel file and the numbers should be filled in the column begin with “n” in the Excel file:

“””

|E1 | n1 | E2 | n2 | E3 | n3 | E4 | n4 | E5 | n5 | E6| n6 |

|Li | 2 | O| 8| Co| 2 | Sn | 2 | - | 0 | -| 0 |

“””

Note that Li is always filled in “E1” column and O is always filled in “E2” column. For example, the “mp-21270” and “mp-766044” in the “mp-id” column in the JSON file should be filled as:  
“””

| MP ID| c-axis length|E1 | n1 | E2 | n2 | E3 | n3 | E4 | n4 | E5 | n5 | E6| n6 |

| mp-21270| 9.750784608230509|Li | 4 | O| 24 | Fe| 4 | Ge | 8 | - | 0 | -| 0 |

| mp-766044| 5.378668|Li | 1 | O| 8 | Sr| 1 | La| 3 |Fe | 1 | -| 0 |

“””

Should you have any question, ask me in your response.

2. 信息提取提示工程

2.1 提取文本格式文献中的科研信息，Prompt A

You are a researcher in the field of lithium-ion batteries cathode who is good at extracting information from text. Your job is to extract information of lithium metal oxide that is used as lithium-ion battery cathode from the given text. This is the text which may or may not include the information wanted including:

1. The name of the lithium metal oxide;

2. The diffusion coefficient of lithium ion in the lithium metal oxide;

3. The structure parameters of lithium metal oxide (including the length of a, b and c axis);

4. The band gap of lithium metal oxide;

5. The temperature when calculating the diffusion coefficient of lithium ion;

6. The band gap of the lithium metal oxide;

7. The calculation method used to calculate the lithium metal oxide properties, for example, density function theory (DFT) or AIMD;

8. The calculation software when calculating the lithium metal oxide properties, for example, VASP, Material Studio (MS) and Quantum Espresso;

9. The cutoff energy when calculating the lithium metal oxide properties;

10. Other parameters about lithium metal oxide (such as power density and coulombic efficiency.

Please follow these steps to extract information from the given text:

1. Analyze the text whether it is about lithium metal oxide and their properties or not.

2. IF the text is about lithium metal oxide THEN proceed to step 3 ELSE go to step 6.

3. Identify the wanted information above.

4. Extract these values along with their units.

5. Organize these into a table with following columns: Lithium Metal Oxide Name, Variable Name, Numeric Value and Unit. Then, END

6. IF the text is not about lithium metal oxide and their properties, output a blank table

7. END

You can perform the following actions:

1. Identify whether the text is about the characterization of lithium metal oxide or use them as the cathode of lithium-ion batteries.

2. Extract related information from the given text and given sample table format.

3. Organize related information into the desired format.

4. Extract the number and unit from the related information.

5. Fill in the table with desired information.

6. Write down the condition of extracted properties at the condition column.

7. Output the table.

Below I will provide you a sample table:

“””

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

| LiCo0.333Ni0.333Mn0.333O2 | length of a axis| 2.86 | Å |

| LiCo0.333Ni0.333Mn0.333O2 | length of b axis| 2.86 | Å |

| LiCo0.333Ni0.333Mn0.333O2 | length of c axis| 14.227 | Å |

| LiCo0.333Ni0.333Mn0.333O2 | temperature | 298.15 | K |

“””

Below is the text where the information to be extracted.

“””

[text content]

“””

My other requirements are:

1.You have unlimited time to analyze the text.

2. If one variable has multiple values, you should show me all values rather than only one value. For example, you should show me:

'''

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO)| Lithium-ion diffusion coefficient| 6.42 × 10−12| cm2 s−1|

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO-C1) | Lithium-ion diffusion coefficient| 1.37 × 10−11| cm2 s−1|

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO-C2) | Lithium-ion diffusion coefficient| 8.93 × 10−12| cm2 s−1|

'''

You should not show me:

'''

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

|LLMO-C1, LLMO-C2, LLMO-C3|Lithium-ion diffusion coefficient| Varies per sample| cm² s−1 |

'''

2.2 提取文本格式文献中的科研信息，Prompt B

You are a researcher in the field of lithium-ion batteries cathode who is good at extracting information from text. Your job is to extract information of lithium metal oxide that is used as lithium-ion battery cathode from the given text. This is the text which may or may not include the information wanted:

'''

[text content]

'''

Your mission is to extract the following information from the given text:

1. The name of the lithium metal oxide;

2. The diffusion coefficient of lithium ion in the lithium metal oxide;

3. The structure parameters of lithium metal oxide (including the length of a, b and c axis);

4. The band gap of lithium metal oxide;

5. The temperature when calculating the diffusion coefficient of lithium ion;

6. The band gap of the lithium metal oxide;

7. The calculation method used to calculate the lithium metal oxide properties, for example, density function theory (DFT) or AIMD;

8. The calculation software when calculating the lithium metal oxide properties, for example, VASP, Material Studio (MS) and Quantum Espresso;

9. The cutoff energy when calculating the lithium metal oxide properties;

10. Other parameters about lithium metal oxide (such as power density and coulombic efficiency.

Please follow these steps to extract information from the given text:

1. Analyze the text, answer this question according to the content of the text: “Whether the text about the information mentioned above or not? Your answer should only be yes or no.” Keep in mind that do not output your answer, but execute Step 2.

2. IF your answer in Step 1 is “Yes”, continue the process to step 3. If your answer in Step 1 is “no”, end the process and output nothing.

3. Identify the information mentioned previously. IF you find no relevant variables, output nothing.

4. Extract name of the information, their values and their units.

5. Organize these into a table with following columns: Lithium Metal Oxide Name, Variable Name, Numeric Value and Unit.

6. END.

Below I will provide you a sample table:

'''

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

| LiCo0.333Ni0.333Mn0.333O2 | Calculation software| VASP | - |

| LiCo0.333Ni0.333Mn0.333O2 | energy barrier| 4.3 | eV |

| LiCo0.333Ni0.333Mn0.333O2 | length of c axis| 14.227 | Å |

| LiCo0.333Ni0.333Mn0.333O2 | Lithium-ion diffusion coefficient | 6.42 × 10−12 | cm2 s−1|

'''

My other requirements are:

1.You have unlimited time to analyze the text.

2. If one variable has multiple values, you should show me all values rather than only one value. For example, you should show me:

'''

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO)| Lithium-ion diffusion coefficient| 6.42 × 10−12| cm2 s−1|

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO-C1) | Lithium-ion diffusion coefficient| 1.37 × 10−11| cm2 s−1|

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO-C2) | Lithium-ion diffusion coefficient| 8.93 × 10−12| cm2 s−1|

'''

You should not show me:

'''

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

|LLMO-C1, LLMO-C2, LLMO-C3|Lithium-ion diffusion coefficient| Varies per sample| cm² s−1 |

'''

3. If you failed to extract the information wanted or there is no wanted information in the text, your output should only be:

'''

Nothing extracted.

'''

If you succeed in extracting the information wanted, your output should only be:

'''

Succeed in extracting information:

[The markdown table]

'''

Remember, NO other sentences or other reply is needed.

2.3 提取图片格式文献中的科研信息

You are a researcher in the field of lithium-ion batteries cathode who is good at extracting information from images. Your job is to extract information of lithium metal oxide from the text in the multiple images provided to you.

Your mission is to extract the following information from the given text:

1. The name of the lithium metal oxide;

2. The diffusion coefficient of lithium ion in the lithium metal oxide;

3. The structure parameters of lithium metal oxide (including the length of a, b and c axis);

4. The band gap of lithium metal oxide;

5. The temperature when calculating the diffusion coefficient of lithium ion;

6. The band gap of the lithium metal oxide;

7. The calculation method used to calculate the lithium metal oxide properties, for example, density function theory (DFT) or AIMD;

8. The calculation software when calculating the lithium metal oxide properties, for example, VASP, Material Studio (MS) and Quantum Espresso;

9. The cutoff energy when calculating the lithium metal oxide properties;

10. Other parameters about lithium metal oxide (such as power density and coulombic efficiency.

Please follow these steps to extract information from the given text:

1. Analyze the text, answer this question according to the content of the text: “Whether the text about the information mentioned above or not? Your answer should only be yes or no.”

2. IF your answer is “Yes”, continue the process to step 3. If your answer is “no”, end the process and output nothing.

3. Identify the information mentioned previously. IF you find no relevant variables, output nothing.

4. Extract name of the information, their values and their units.

5. Organize these into a table with following columns: Lithium Metal Oxide Name, Variable Name, Numeric Value and Unit.

6. END.

My other requirements are:

1.You have unlimited time to analyze the images.

2. If one variable has multiple values, you should show me all values rather than only one value. For example, you should show me:

‘’’

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO)| Lithium-ion diffusion coefficient| 6.42 × 10−12| cm2 s−1|

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO-C1) | Lithium-ion diffusion coefficient| 1.37 × 10−11| cm2 s−1|

|Li1.2Mn0.54Ni0.13Co0.13O2 (LLMO-C2) | Lithium-ion diffusion coefficient| 8.93 × 10−12| cm2 s−1|

‘’’

You should not show me:

‘’’

| Lithium Metal Oxide Name | Variable Name | Numeric Value | Unit |

|LLMO-C1, LLMO-C2, LLMO-C3|Lithium-ion diffusion coefficient| Varies per sample| cm² s−1 |

‘’’

3.There is no token limitation for the final result.

4.When analyzing each image, you should record the abbreviations used in the content and use them when analyzing the rest part of the PDF images. For example, the content of the PDF is:

“””

Materials synthesized with different concentrations of transition metal ions are labeled as LLMO-C1 (0.2 M), LLMO-C2 (0.3 M), and LLMO-C3 (0.4 M). Materials synthesized using different molar ratios of oxalic acid to transition metal ions are labeled as LLMO-P1 (2:1) and LLMO-P2 (3:1), as further discussed in the Supporting Information.

“””

Then you should record the labels “LLMO-C1”, “LLMO-C2”, “LLMO-C3”, “LLMO-P1” and “LLMO-P2”. And when analyzing the rest of the PDF, you should extract the information that involves these labels.

5. If you failed to extract the information wanted or there is no wanted information in the text, your output should only be:

‘’’

Nothing extracted.

‘’’

If you succeed in extracting the information wanted, your output should only be:

‘’’

Succeed in extracting information:

[The markdown table]

‘’’

Remember, NO other sentences or other reply is needed.

3. 生成DFT计算输入文件的提示工程

You are expert at DFT calculation. Generate me the necessary files to calculate the formation energy of the material using VASP. I have upload the .cif file for you to analyze.