

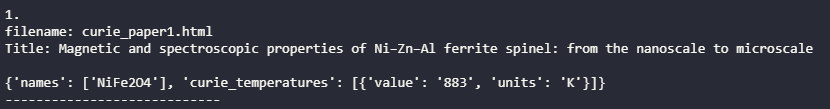
Above pic is Example of Database Output

Columns:

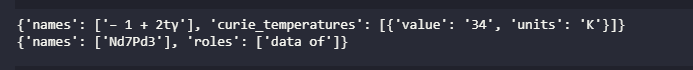
1. Names -- Chemical Compound(s) Extracted
2. Type -- Neel or Curie
3. Extracted Value
4. Extracted Unit
5. Normalize Value (in K) -- just convert extracted to Kelvin
6. Normalize Unit (K)
7. File name -- file name of the paper ← this is added not in pic above

Types of extracted examples:

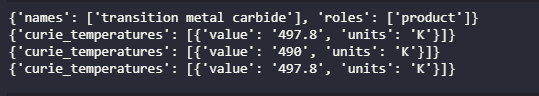
#Type 1 (most common): name and temperature are in the same dictionary



#Type 2 (rare): when Type1 name doesn’t make sense and records contains names with roles

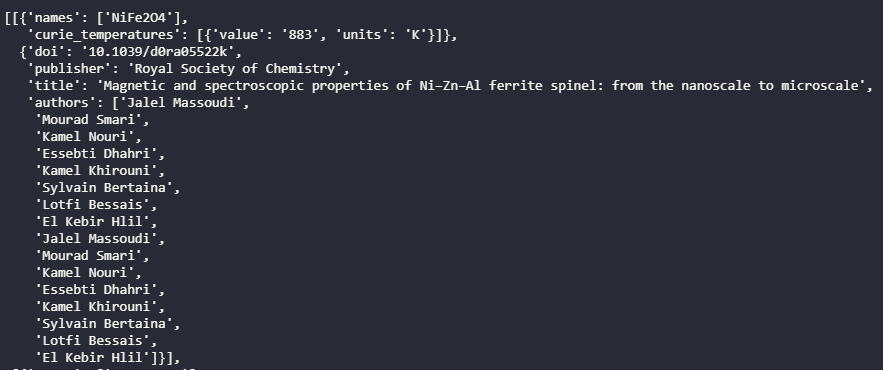


#Type3 (rare): when name (in dictionary with roles) and temperature are separately extracted



Where to work from: [link](https://drive.google.com/drive/u/1/folders/1GDfYwFLJbjcxiVu6V0FvSgRoK1Ox7Fup)

\*\*\*\*put *get\_papers\_rsc.ipynb* and *rsc\_neel\_curie\_extract.ipynb* in **CDE modify** [**link**](https://drive.google.com/drive/u/1/folders/1n3fZUw4mXWiV3y_mvC_9q5oHge6Gy9Rm)before using it

* Download 1000 curie papers and 264 neel papers -- run *get\_papers\_rsc.ipynb*
* Get **all\_extracted\_json** from *rsc\_neel\_curie\_extract.ipynb*
* **all\_extracted\_json** is a list with only papers that have useful info
* **all\_extracted\_json** [0] = first paper’s **all extracted datas (in dictionary form)(can be multiple)** + **citation (also in dictionary form)**

Note:

* There may be **multiple names** extracted with **only one** value of **temperature** extracted keep them like this 
* There may be **one name** extracted with **multiple** values of **temperature** extracted please keep all the numbers separated in different rows