Batch processing with BeautifulJASON

Dealing with large volumes of data occurs in a wide range of NMR applications, from process development to metabolomics. How to efficiently extract the information you need from these spectra is extremely important. In this post, I will explain how the BeautifulJASON python package can be used to rapidly integrate an entire directory of data and extract those integrals into a text file for further downstream processing.

In order to follow the steps in this post you will need a copy of JASON and install the BeautifulJASON python package. The latter can be installed using pip with the following command:

```
pip install beautifuljason
```

pip will take case of installing all the required dependencies.

To handle the processing we will use JASON's rules feature. This works just like the rules in an email client, when incoming data matches a rule, in this case a processing rule, it is applied to the data. So the first steps are to create a rules library and processing rule within it:

1. Create a new rule library , for example, called "Batch process and integrate"

2. Open a dataset and set up the desired processing. Then, create a processing rule from this dataset

In this example, we wish to extract a set of integrals from our data. Again, we can use the power of JASON's rules, this time creating an analysis rule:

3. Perform manual integration within the desired ranges and create an analysis rule with the "Custom Ranges" option selected

Just to check that we have done everything correctly, we should reopen the dataset and ensure that the desired processing and analysis are performed as we expect. Now, any dataset matching the rules will have the processing and analysis applied.

We now want to automate the analysis of our data stored in a given directory. To do this, we utilise a script supplied with BeautifulJASON: `jason_batch_convert_ext.py`.

4. Run the command:

```
jason_batch_convert_ext.py <input_folder>
<output_folder> --formats jjh5 --extensions jdf --rules
"Batch process and integrate" --execute
```

`<input_folder>` and `<output_folder>` should be replaced with the locations of the data and where you wish the output to be placed.

The other options are as follows:

- `--formats` specifies what format the use to write out the data, in this case JASON's .jjh5 file format
- `--extensions` specifies the extensions used for the input files. In the case of data stored in folders rather than individual files, you can use the `--patterns` option to describe the path to the data
- `--execute` runs the script. If this option is not specified, a dry run occurs and information about what would be done is returned, but no files are changed

Use the `--help` option for usage information on the script.

You should now have an output folder containing a series of JASON .jjh5 files which have all had the desired processing and analysis applied. The final step is to extract the integral information from these files.

5. Finally, run the script:

```
batch_extract_integrals.py <output_folder> output.csv
```

The first argument is the output folder from step 4, and the second argument is the name of the .csv file to create. There is the option to extract additional parameters from the .jjh5 files using the `--parameters` flag. More information can be found using the `--help` option.

The end result of this process should be a folder with processed .jjh5 files and a .csv file containing all the integrals from these data.

This is just a simple example of how BeautifulJASON can be used to help automate the processing and analysis of large volumes of NMR data. The scripts referenced in this post are included in the BeautifulJASON python package as ready to use tools and can be used as the basis of your own analysis! Ensure that the appropriate python 'Scripts' (on Windows) or 'bin' (on macOS) directory is on your path.

Happy automating!