cheMastery Chemical Language Test

At cheMastery your day-to-day work will revolve around extracting and categorizing features from existing chemical experimental texts, matching these against and translating these into a standardised, machine-readable format. These experimentals have been written by humans and are typically messy and virtually always incomplete.

For this exercise use the 10 real-life examples of experimental descriptions provided in a separate plain text file. We have left the experimentals as they are found in the scientific literature and you are not expected to understand these fully on your own, at the job you will have constant support from the founder team, both chemists by training.

From the experimentals devise a way to extract one technique (addition) and check which of the required parameters listed below are given for it and highlight which ones are missing.

Addition:

* Type (In Portions or Continuous)
* Chemicals (what is added to what)
* How much of each constituent (no need to implement any form of unit recognition, a simple string will do for the purpose of this exercise)

Please create a git repository and write your code in Python 3.x. You are free to use any packages you deem fit for the task. The python package ChemDataExtractor (https://github.com/mcs07/ChemDataExtractor/) will likely be helpful to you, but there is no obligation to use it.

The output of your code should be a JSON object or JSON formatted string containing the original text as well as all found instances of additions with parameters found and missing parameters marked as such. Please do not spend more than a few hours on the task; you are not expected to produce an all-encompassing solution to the posed problem.

Your work will be rated on readability, maintainability of your code, quality and depth of unit testing as well as cleanliness of git commits. You will also be discussing your exercise and the approach you took during your next interview stage.