Exercises

1. Create a simple PDF file without tables (you can use Microsoft Word to create a document and save it as a PDF file) and read the text using Python. Print the results.
2. Create a simple PDF file with tables (you can use Microsoft Word to create a document and save it as a PDF file) and read the text using Python. Print the results.
3. Go through the Microsoft tutorial to create a form processing model using the Microsoft invoice samples. Do a "quick test" using the test invoice. Then, after reading about how you can incorporate your model in Power Automate, create a simple Power Automate flow that reads that test invoice and shows the data fields within it. (There may be a tutorial available from Microsoft that shows you how to do this.)
4. Assuming you installed Tesseract, use pytesseract to read the Bowers text image found in the GitHub for Week 3 (week_3\data\bowers.jpg). Then use spaCy to print out the tokens (the text, part of speech, and dependency).

For the Python assignments, you can submit Jupyter Notebooks or PDFs of your code (one for each exercise). If you submit .py files you need to also include a PDF or attachment of your results.

For the Power Automate assignment, share your completed flow with me (fneugebauer@bellevue.edu).

Don't forget to work on your project – Milestone 2 is due next week.