Hint 1

To access the newspapers in the zipfile, you must first use the Zipfile library to open the zipfile then iterate through the objects (newspapers) in the zipfile using .infolist(). Try and write a simple routine to just go through the zipfile, printing out the name of the file as well as using display(). Remember that the PIL.Image library can .open() files, and that items in .infolist() in the zipfile each appear to Python just as if they were a file (these are called "file-like" objects).

Hint 2

You can spend a lot of time converting between PIL.Image files and byte arrays, but you don't have to. Why not just store the PIL.Image objects in a global data structure, maybe a list or a dictionary indexed by name? Then you can further process this data structure, by adding in information such as the text detected on the pages or the bounding boxes behind faces. Come to think of it, a list of dictionary objects, where each entry in the list would have the PIL image, the bounding boxes, and the text discovered on the page, would be a handy way to store this data.

Hint 3

A quick reminder - in Python all strings are just like lists of characters. Kind of (remember they are immutable lists - more like tuples!). But this means you can use the in keyword to find substrings really easily. So the following statement will return True if the substring is matched: if "Christopher" in my\_text

Hint 4

Creating the contact sheet can be a bit of a pain. But you can resize images without having to worry about the aspect ratio if you use the PIL.Image.thumbnail function. I used it when creating out the output images, maybe you should too! And check out the lecture on the contact sheet, you want to be careful that you don't "walk off" the end of the images when creating a row (or column).