Student Writing as Digital Humanities Method

- Mackenzie Brooks, Asst Prof & Digital Humanities Librarian // @mackymoo
- Abdurrafey Khan, DH Fellow, French major // @AbdurKhanye
- Brandon Walsh, Mellon Digital Humanities Fellow // @walshbr

Bucknell Digital Scholarship Conference
October 29, 2016

DH @ W&L | Digital Humanities at Washington and Lee University

ABOUT PROJECTS INITIATIVES BLOG

digitalhumanities.wlu.edu

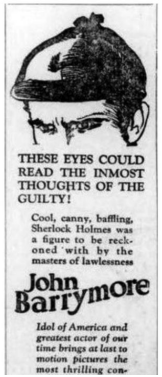
Problems with Data

So you have a text. You want to do something with it. It might be tempting to dive in and start using one of the tools in this book, but you should take a moment to examine the materials you are working with. Not all text is created equal, and your results can have real problems if you don't take care to examine the quality of the materials before you work with them.

The basic principle to remember is **garbage in, garbage out (or GIGO)**: you won't get good results unless you have good data to begin with.

OCR

Take this image, drawn from a 1922 printing of *The Duluth Herald*, of a newspaper ad for the American film version of Sherlock Holmes.

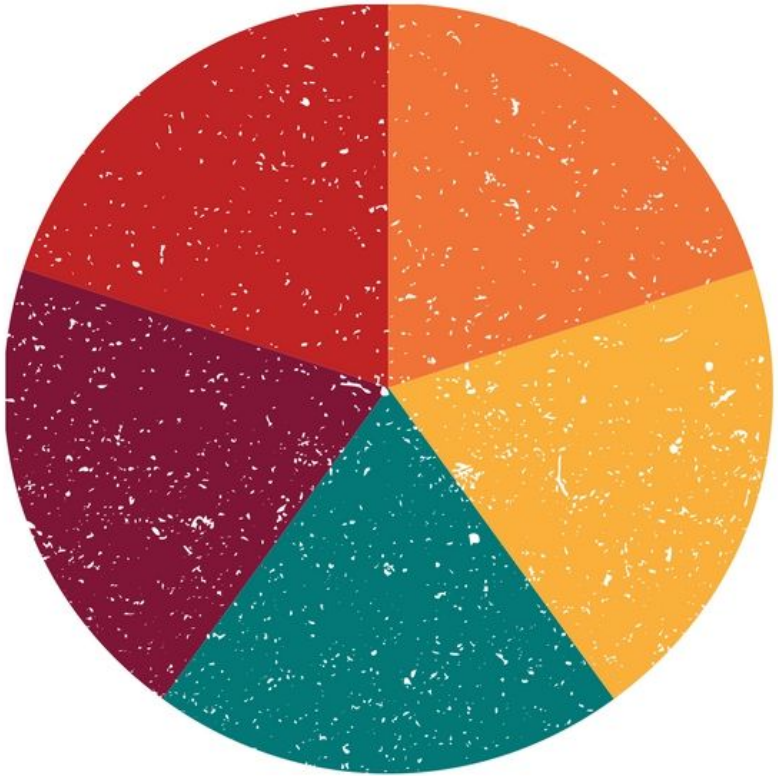


By default, the computer has no idea that there is text inside of this image. For a computer, an image is just an image, and you can only do image-y things to it. The computer could rotate it, crop it, zoom in, or paint over parts of it, but your machine cannot read the text there - unless you tell it how to do so. In fact, the computer doesn't even really know that there *is* text there. As far as it's concerned, an abstract painting and an image like this contain the same amount of textual information. The computer requires a little extra help to pull out the text information from the image.

The process of using software to extract the text from an image of a text is called **optical character recognition** or OCR. We occasionally use OCR as a noun, as in "the OCR for that document is pretty poor" or as a verb, as in "we

goo.gl/qqDEZV

DH 102: Data in the Humanities



Unit 1 / Text / Assignments

Core Assignments

Methodology Review

To familiarize yourself with the various methodologies in the DH world, please conduct a review of one DH project. You may select from the list below or find a project of your choosing (subject to professor approval), as long as it uses textual data as its source.

Specs:

- Due Thursday, 9/22 by 8am.
- 300-500 words.
- Submit via your Box folder.
- Be prepared to share your review in class.
- Do not quote extensively from the project website. Summarize in your own words.
- Address the following questions:
  - What is the goal of this project? Are there guiding research questions?
  - Who are authors? What are their affiliations and roles? Have they received external funding? Are students involved?
  - Tell me about the data. Where did it come from? How has it been cleaned, modified, or enhanced?
  - What established standards (technical or otherwise) does the project data use?

DH 102 Assignments

- Project/methodology review
- Project data (assessment + cleanup)
- Project narrative *
- Project visualization *
- Project documentation *
- Project reflection
- 2 blog posts *

= 2000 words per unit * post to website

Writing as DH method:

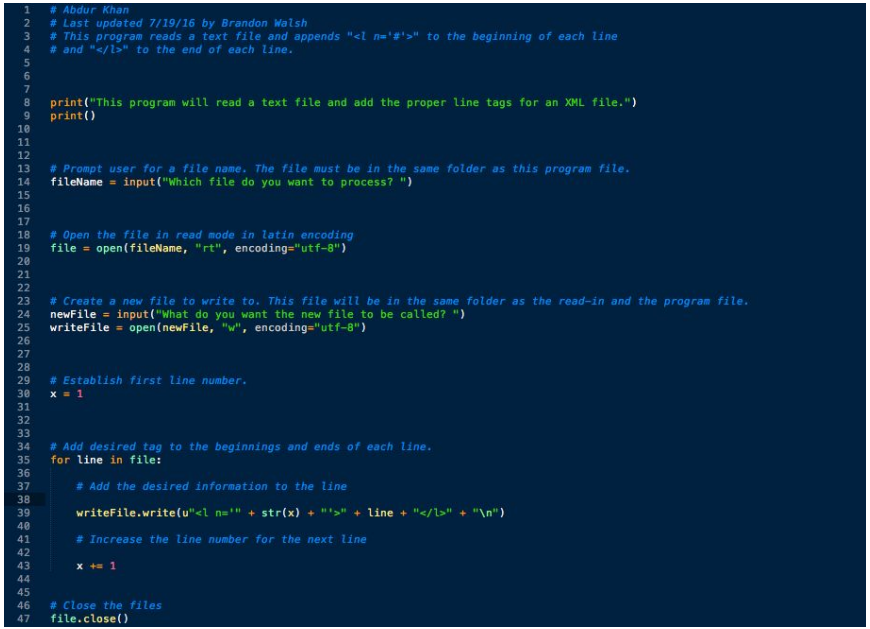
- Levels the playing field
- Exposes students to new/different genres
- Allows for that “failure” thing we all talk about

Documentation keeps me from yoking together illogically disparate ideas, writing beyond my professor's expectations, and falling down research rabbit holes, as I am wont to do as a poet, critic, and passionate, albeit overly pedantic, undergraduate student.

- Kassie Scott '18

Writing for DH

- Discussing vs. “essay”ing
- Communicating code and technical ideas through blogs
- Informal and versatile



```
# Abdur Khan
# Last updated 7/19/16 by Brandon Walsh
# This program reads a text file and appends "<l n='#'>" to the beginning of each line
# and "</l>" to the end of each line.



print("This program will read a text file and add the proper line tags for an XML file.")
print()



# Prompt user for a file name. The file must be in the same folder as this program file.
fileName = input("Which file do you want to process? ")



# Open the file in read mode in latin encoding
file = open(fileName, "rt", encoding="utf-8")



# Create a new file to write to. This file will be in the same folder as the read-in and the program file.
newFile = input("What do you want the new file to be called? ")
writeFile = open(newFile, "w", encoding="utf-8")



# Establish first line number.
x = 1



# Add desired tag to the beginnings and ends of each line.
for line in file:

    # Add the desired information to the line

    writeFile.write(u"<l n='" + str(x) + "'>" + line + "</l>" + "\n")

    # Increase the line number for the next line

    x += 1



# Close the files
file.close()
```

DH Honors Thesis?

- Different considerations for multimedia thesis
- Split between traditional text analysis and using a modern digital interface
- How to field questions - “What is DH? Why use it?”
- www.huondauvergne.org