Modern Alchemy - Turning data into information

TL;DNR This directory contains the content necessary for facilitating a hands-on pre-conference workshop called Modern Alchemy, presented at [Code4Lib 2022](#).

Introduction

This hands-on workshop demonstrates, illustrates, and instructs partipants on the use of a command-line application called the Distant Reader Toolbox. The Toolbox takes a set of narrative text as input, outputs a data set, and provides functions designed to analyze ("read") the result. Through this process, the Toolbox takes data ("words"), and turns it into information.

Workshop outline

```
   I. Introductions - Who are you, and why are you here? What do you want to learn?
  II. Break-up into teams - Two heads are better than one.
 III. Modern alchemy - Through modeling processes, one can tell compelling stories about a given corpus.
  IV. Demonstration
      A. Get help (rdr --help)
      B. Download Homer (rdr download homer)
      C. Get info (rdr info homer)
      D. Summarize (rdr summarize homer; rdr read homer)
      E. Ngram analysis (rdr ngrams homer -c | more; rdr ngrams -s 2 -c | more)
      F. Concordance (rdr concordance homer)
      G. Keyword analysis (rdr wrd homer -c | more)
      H. Full-text search (rdr search homer -q war | more)
      I. Topic model (rdr tm homer)
      J. Collocations (rdr collocations homer)
      K. Parts-of-speech and named entities (rdr pos homer -c; rdr ent homer -c)
      L. Grammars (rdr grammars homer)
   V. Installation
      A. Anaconda - A development environment
      B. Reader Toolbox (pip install reader-toolbox)
  VI. Practice - Download and practice with Pride & Prejudice (rdr download pride)
 VII. Create your own carrel - Bring your own data, or use sample data, below.
VIII. Show & tell - What did you learn from your analysis/reading?
  IX. Summary - The Toolbox creates datasets and provides ways to model the result.
```

Sample data

Here are sets of sample study carrel data and possible questions to address against them:

- A large volume of *Code4Lib Journal* ([code4lib-journal.zip](#)) - How is MARC described and how does it compare to Linked Data?
- A single issue of *Code4Lib Journal* ([issue-053.zip](#)) - To what degree does each author discuss a specific topic, and to what degree do they all discuss the same thing?
- A large volume of *In The Library with the Lead Pipe* ([lead-pipe.zip](#)) - Using specific examples from the corpus, how would you characterize this journal?
- *Moby Dick* by Herman Melville ([moby.zip](#)) - What can you learn about the whale and whaling?
- A set of historical narratives about and surrounding plague ([plague-literature.zip](#)) - When the works mention the word "plague", what other words are used in the same breath? How is the plague associated with life, death, and love?
- The whole of Shakespeare's plays ([shakespeare.zip](#)) - Can you illustrate and demonstrate how Shakespeare's works are logically divided into commedies, tragedies, and histories? If so, then how?
- A subset of my professional literature ([whats-eric-reading.zip](#)) - What am I reading?

Links

Here is a set of links of possible interest:

- the [original Distant Reader paper/poster](#)
- the [most up-to-date documentation](#)
- an [entertaining radio interview](#) on the topic of the Distant Reader
- the [screen-play of a movie demonstrating the Toolbox](#), complete with hyperlinks to specific segments
- [this document](#), designed for printing

Eric Lease Morgan <emorgan@nd.edu>
May 20, 2022

http://dh.crc.nd.edu/tmp/modern-alchemy