Semester Project

The story of the origin of Frankenstein is almost as fanciful as the story itself.  It was written as a writing challenge between Lord Byron, John Polidori, Marry Wollstonecraft Shelly, and Percy Bysshe Shelley while the four shared a lake house in Geneva, Switzerland.  I reference the following History.com article as a source on the origin of the Frankenstein novel.  [**https://www.history.com/news/frankenstein-true-story-mary-shelley (Links to an external site.)**](https://www.history.com/news/frankenstein-true-story-mary-shelley)

There is apparently some controversy on who the actual author of Frankenstein was.  I reluctantly reference the following self-promotional article that declares the authorship of Frankenstein to be Percy Shelly, and not Mary.

[**https://www.us.mensa.org/read/bulletin/features/the-real-frankenstein-and-its-author/**](https://www.us.mensa.org/read/bulletin/features/the-real-frankenstein-and-its-author/)

 Following from the challenge article, I feel there is another unexplored option that can be used to determine authorship.  That method is data mining, or more specifically text mining.

Your semester project is to take up the challenge and use data mining to assess the authorship of Frankenstein.  I have attached two articles to start you on your way to analyzing the authorship of Frankenstein.

Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology*, *60*(1), 9-26.

[**Computational Methods in Authorship Attribution.pdf**](https://mst.instructure.com/courses/58287/files/3319555?wrap=1)[**download**](https://mst.instructure.com/courses/58287/files/3319555/download?download_frd=1)

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, *60*(3), 538-556.

[**A Survey of Modern Authorship Attribution Methods.pdf**](https://mst.instructure.com/courses/58287/files/3319545?wrap=1)[**download**](https://mst.instructure.com/courses/58287/files/3319545/download?download_frd=1)

Locate a collection of work from the four people present at the origin of the Frankenstein book, along with works from authors who were their contemporaries.  Prepare these works for text mining and convert the works to a collection of attributes.  Use three different data mining techniques, at least one of the techniques must be a technique that was not discussed in class, to determine the authorship of Frankenstein.

Follow the structure set out in the Submission Guide we used for the Programming Assignments.

[**Submission Guide - Progamming Assignment.pdf**](https://mst.instructure.com/courses/58287/files/3293664?wrap=1)[**download**](https://mst.instructure.com/courses/58287/files/3293664/download?download_frd=1)

Pay special attention to the following:

* Document how the data was collected, or where you got the data from.
* Document all the techniques used to prepare the data.
* Use three data mining techniques. 1 technique must be outside the class.
* Evaluate the success of each technique and then determine which technique is better.
* Write up your conclusion as to who the author of Frankenstein is.
* Include papers in your References about the technique that has not been covered.