

IFI 9000 Analytics Methods Text Representation and Applications

by **Houping Xiao**

Spring 2021



Introduction

- Over 80% of real-world data is **unstructured**



Introduction

- Turn **unstructured** into **structured** data



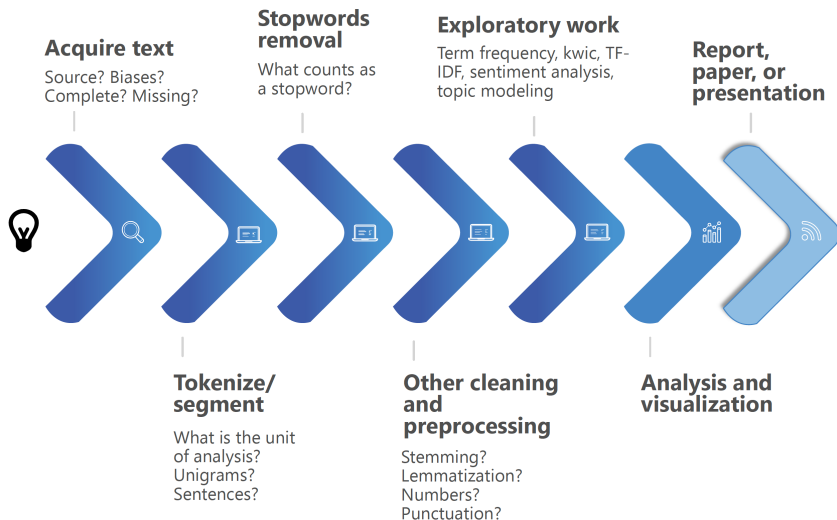
Introduction

- Turn **unstructured** into **structured** data



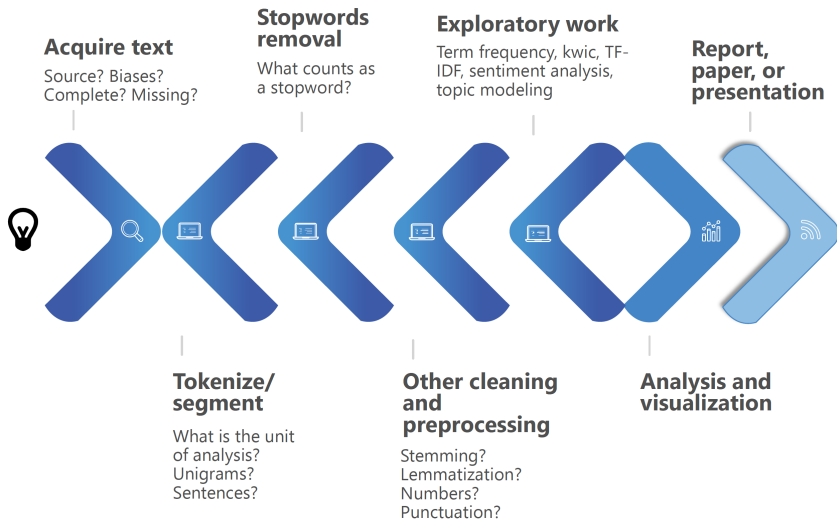
Text mining

- The well-known text mining/NLP pipeline



Text mining

- The well-known text mining/NLP pipeline



Text mining

- Your (text) analysis is only as good as your cleaning and preprocessing choices



Text mining: case study

- A petition from [change.org](https://www.change.org)

Close Georgia State University campuses - Switch to online classes until further notice



7,731 have signed. Let's get to 10,000!

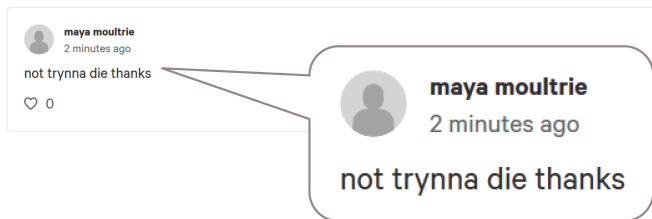


Text mining: case study

- A petition from change.org
 - Downtown campuses and other campuses of Georgia State University throughout Georgia pose a great risk of infection/transmission of the Coronavirus to both the students and the staff. Close the campuses and offer online classes for all classes and the staff until further notice to ensure the safety of all GSU members.
 - Getting infected with the virus puts families at risk. The downtown campus is located next to a hospital and surrounded by people. Keeping the school open is reckless and is bound to spread the virus unnecessarily. Classes can be offered online and should be for our safety. I do not want to get my family sick and neither do any other students.
 - The professors and students on campus are at great risk of exposure as GSU hosts numerous commuters from all over Georgia. It is time to be proactive and assume that the lack of testing does not mean a lack of cases. They are out there around us, and they are spreading the virus.

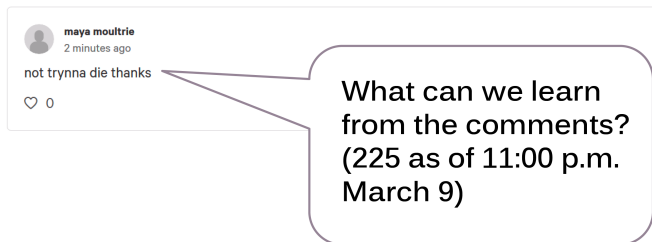
Text mining: case study

- A petition from [change.org](#)
- **Our goal:** reasons for signing
 - See why other supporters are signing, why this petition is important to them, and share your reason for signing (this will mean a lot to the starter of the petition).



Text mining: case study

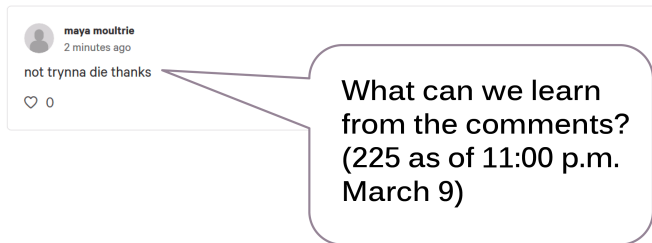
- A petition from change.org
- **Our goal:** reasons for signing
 - See why other supporters are signing, why this petition is important to them, and share your reason for signing (this will mean a lot to the starter of the petition).



The image shows a screenshot of a social media comment. On the left, a user profile for 'maya moultrie' is shown with a grey profile picture, the name 'maya moultrie', and the text '2 minutes ago'. Below the profile is the comment text 'not trynna die thanks' and a heart icon with the number '0'. A callout box on the right points to the comment text and contains the text: 'What can we learn from the comments? (225 as of 11:00 p.m. March 9)'.

Text mining: case study

- A petition from change.org
- **Our goal:** reasons for singing
 - See why other supporters are signing, why this petition is important to them, and share your reason for signing (this will mean a lot to the starter of the petition).



- On iCollege: codes and data files
 - text_comments.csv
 - text.R or text.ipynb

The End