Shapes of Stories User Manual

Authors: Alice Chao & Joel Krim

This program was inspired by Kurt Vonnegut's lecture on the "Shapes of Stories," where he postulated that if stories are graphed on two dimensions, time and level of fortune, there are only a few shapes stories can take (https://www.youtube.com/watch?v=oP3c1h8v2ZQ).

We decided to explore this idea further.

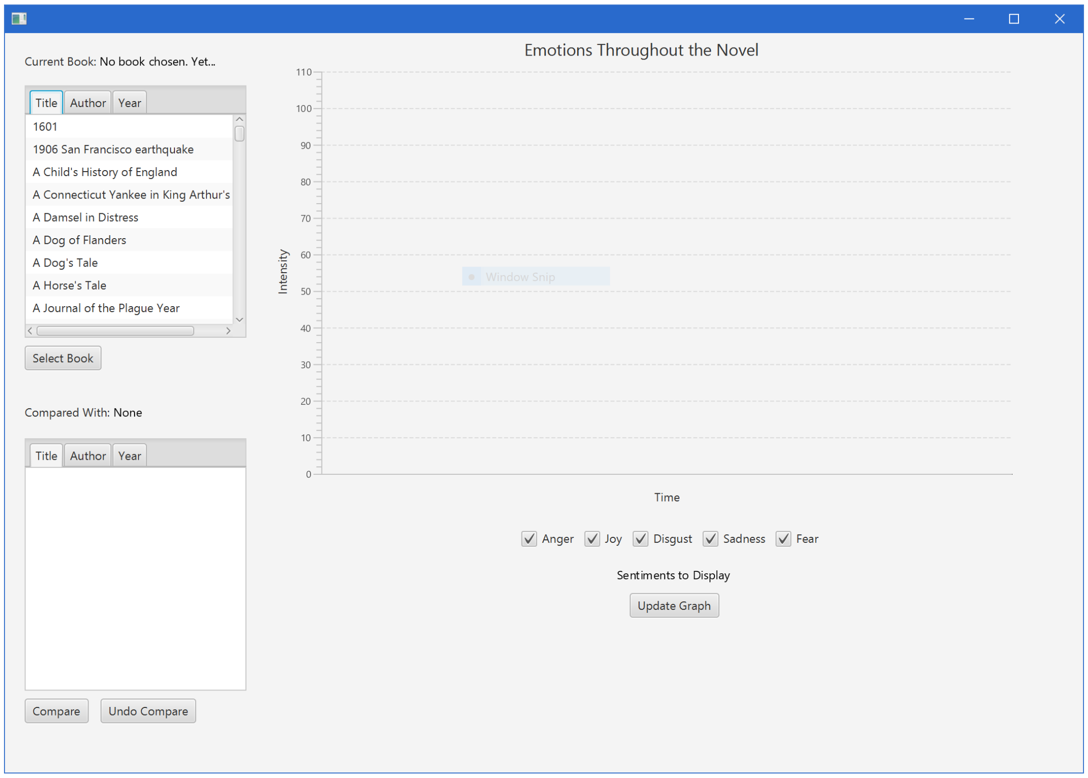
Using IBM Watson's Natural Language Understanding API, which can assess a piece of text for five emotions: anger, disgust, fear, joy, and sadness, we visualized 1500 books to discover a variety of story shapes.

REQUIREMENTS:

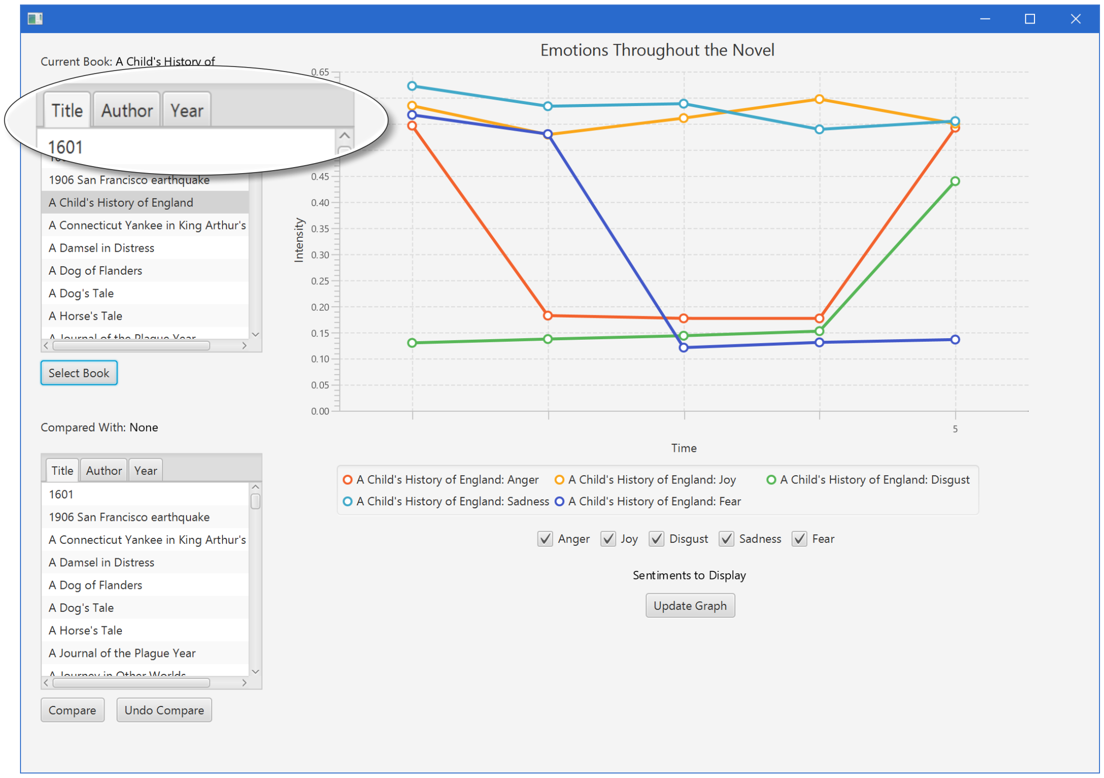
ShapesOfStories.jar	(an executable .jar file)

HOW TO USE:

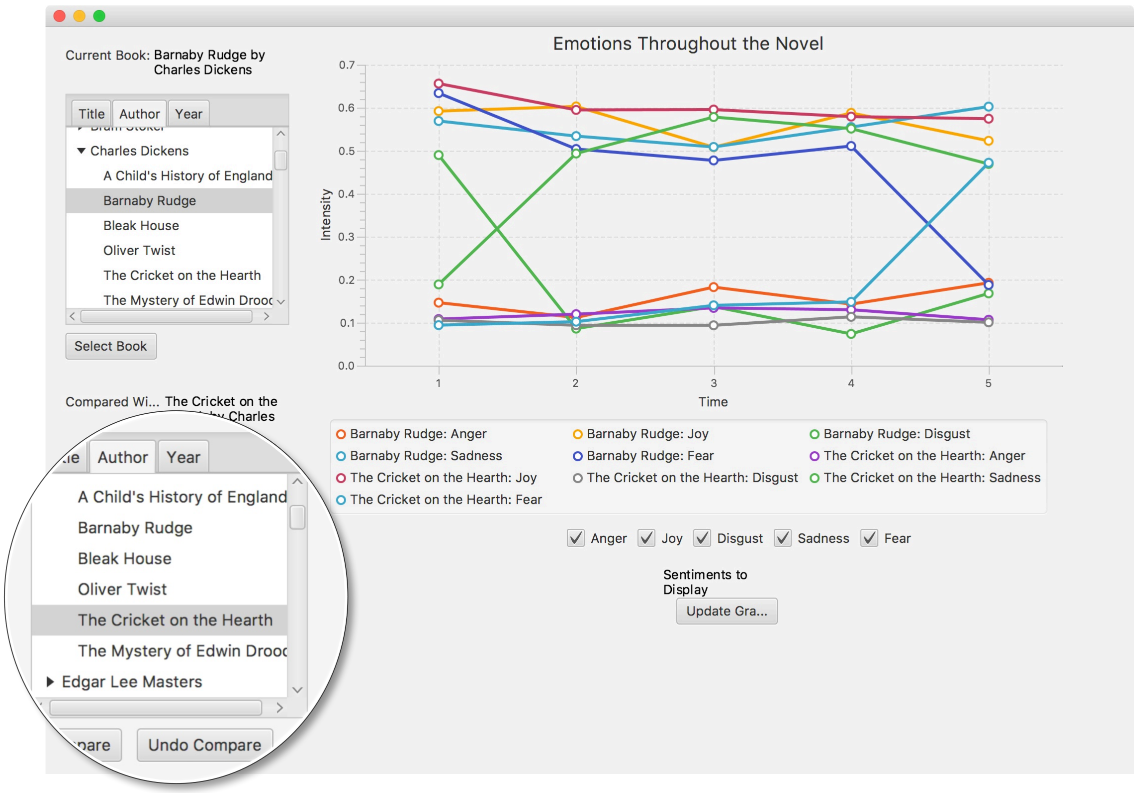
1. Run "ShapesOfStories.jar" program, an executable jar file, which should open this window:



2. Select a book by title, author, or publication year.

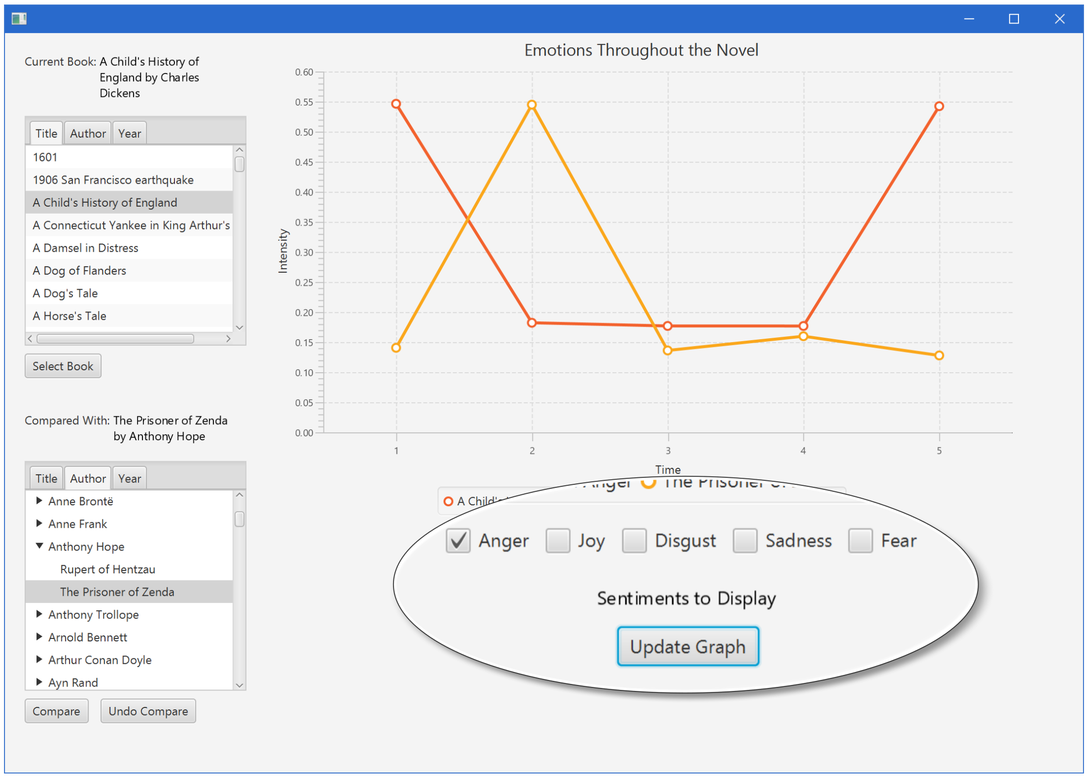

Current Book: A Child's History of
Title
Author
Year
1601
1906 San Francisco earthquake
A Child's History of England
A Connecticut Yankee in King Arthur's
A Damsel in Distress
A Dog of Flanders
A Dog's Tale
A Horse's Tale
Select Book
Compared With: None
Title
Author
Year
1601
1906 San Francisco earthquake
A Connecticut Yankee in King Arthur's
A Damsel in Distress
A Dog of Flanders
A Dog's Tale
A Horse's Tale
A Journal of the Plague Year
Compare
Undo Compare
Emotions Throughout the Novel
Intensity
Time
A Child's History of England: Anger
A Child's History of England: Joy
A Child's History of England: Disgust
A Child's History of England: Sadness
A Child's History of England: Fear
Anger
Joy
Disgust
Sadness
Fear
Sentiments to Display
Update Graph

3. Select a second book to compare books



Current Book: Barnaby Rudge by Charles Dickens
Title
Author
Year
Charles Dickens
A Child's History of England
Barnaby Rudge
Bleak House
Oliver Twist
The Cricket on the Hearth
The Mystery of Edwin Drood
Select Book
Compared Wi... The Cricket on the ... Charles
Edgar Lee Masters
Undo Compare
Emotions Throughout the Novel
Intensity
Time
Barnaby Rudge: Anger
Barnaby Rudge: Joy
Barnaby Rudge: Disgust
Barnaby Rudge: Sadness
Barnaby Rudge: Fear
The Cricket on the Hearth: Anger
The Cricket on the Hearth: Joy
The Cricket on the Hearth: Disgust
The Cricket on the Hearth: Sadness
The Cricket on the Hearth: Fear
Anger
Joy
Disgust
Sadness
Fear
Sentiments to Display
Update Gra...

4. Select or deselect attributes to view specific trends



WHAT'S UNDER THE HOOD:

The Shape of Stories program comes preloaded with a book repository that contains nearly 500 titles. The rationale for preloading the program was to improve runtime as well as to avoid running the book analysis on demand, as this would require an active IBM Watson Natural Language Understanding service account to be tied to the program.

Acquiring these stories and the emotion data involved three relatively autonomous components:

1. Book repository:
   URLGetter.java scrapes loyalbooks.com for valid book text URLs using the JSoup API to generate a "Book URLs.csv" file. BookMetaDataGetter.java and BookMetaDataHolder.java use this .csv file to comb the book's Wikipedia page for the book title, author, and publication year to filter out the books that do not have this information.

2. Emotion analysis:
   The emotion data analysis is performed using the IBM Watson Natural Language Understanding API, which can assess and score a body of text for five emotions: joy, sadness, anger, fear, and disgust. Before Watson processes the text, however, the book is first divided into segments by the BookSplitter.java class, which reads in one book at a time from the book repository URLs. These segments provide the necessary data points for the visualization. We determined 5 segments would be a reasonable number to effectively demonstrate a story arc. These segments are then processed in WatsonAnalyzer.java, which contains the customized Watson API calls, and the data is cleaned and parsed in WatsonParser.java using the Google Gson API. In order to be able to process several books at a time, BookEmotionData.java synchronizes all of these analytical tasks for a single book as an autonomous unit that requires only a book text URL.

   BookDataPrinter.java connects the book repository and book emotion analysis pieces together by running through the full "Book URLs.csv". For each book with valid information (e.g. title, author, publication year), the book's emotion data is obtained and all of this information is written out to a "Book-Data.csv" file for the program.

3. Graphical user interface (GUI):
   With the program data stored in the "Book-Data.csv" file, BookObjDataReader.java and BookObj.java were needed to read in the files. Storing the information this way prevented the need to run the analyses on demand, which would involve lengthy wait times—of the 1600 available online books, a little less than 500 had viable title, author, and publication dates, but this still took an hour to run through Watson.

   JavaFX was used to create the GUI that displays the book repository and visualize the emotion data, which can be run in an IDE via ApplicationMain.java. NLPController.java controls the program functionality in the way the data is organized and displayed on the interface, and NLP_GUI2.fxml and application.css provide the design shell.