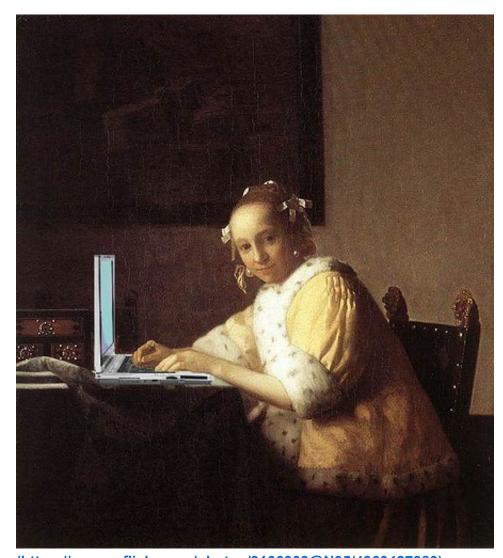
Course Syllabus

Text Mining as Historical Method

HIST 3368 (undergrad) / Hist 6322 (Grad)

Prof. Jo Guldi, PhD; Associate Professor of History & Data Science

Meets 9-10am MWF in Harold Simmons Hall 116



(https://secure.flickr.com/photos/9106303@N05/4960427380)

Computer-powered methods are changing the way that we access information about society. New

methods help us to detect change over time, to identify influential figures, and to name turning points. What happens when we apply these tools to a million congressional debates or tweets? This course - which is appropriate to both computationalists as well as those with a background in the humanities (but not code) -- will teach students how to analyze texts and as data for evidence of change over time. This course is an introduction to the cutting-edge methodologies of textual analysis that are transforming the humanities today.

About the Instructor: Professor Jo Guldi, PhD

When I was a PhD student living in Silicon Valley, something amazing happened: Google Books launched the first mass digitization experiment of its kind, releasing scans of the Harvard, Yale, and New York Public Libraries onto the web. I quickly saw that there was potential, in this textual data, for kinds of analysis that had never been attempted before. From that point forward, I helped to found the discipline of "digital history," and I have been applying myself to the study of how computational algorithms can help us learn new things about the past, the present, and ourselves.

I am also a historian of technology who is interested in questions about how we know what is true – whether new technologies or history give us the tools for discerning historical truth in a new way – and when official paper obscures more than it illuminates. My book on the importance of history as a tool for discovering truth, *The History Manifesto*, was recently named one of the most important books across all fields published in the last twenty years.

Contacting Professor Guldi

Email: jguldi@mail.smu.edu (mailto:jguldi@mail.smu.edu)

Office phone: (214) 268-3744

Office Hours: By default, every Wednesday and Friday at 2-3pm, Dr. Guldi will be in her office available to speak to students.

Office hours during these times may be reserved in advance in 15 to 30 minute slots at http://joguldi.youcanbook.me/ (http://joguldi.youcanbook.me/). Pre-booking office hours reserves Dr. Guldi's time in advance and gives you priority over other students. Pre-booking is not required, however; you are welcome to just drop in, although you will have to wait for other students if they have reserved the time. Kindly cancel your youcanbookme appointment if you make other plans so

that others may use the time.

All Office hours are held at 60 Dallas Hall (in the basement under the rotunda) unless otherwise arranged; zoom office hours are available by request.

Preferred means of contact in descending order: 1) Canvas, 2) Slack, 3) email

Here's what's expected of you to flourish in the class:

- Regular attendance of and participation in weekly meetings
- Regularly submit work and keep up with readings
- Show continuous improvement through the semester in dialogue with teacher and peer feedback

Here's what's expected of you in terms of weekly workload:

By the stipulation of university accreditation programs, a 3-credit course requires 6 hours of homework and group work each week. Your work will typically take the form of :

- 2-4 hr code
- 2-3 hr of reading (~50-100 pp)

Depending on how difficult the exercises and readings are for each student, the time required to complete assignments will vary. Any students with concerns about workload should reach out to Dr. Guldi

University guidelines stipulate that students should meet for 2 hr 50 minutes in class each week. In an online and remote class, our interactions will take the form of small group meetings, zoom breakout rooms, live lecture and discussion, pre-recorded lectures, graded discussions, and informal discussion and support via Slack. Structured interactions will be slightly tilted towards the beginning of the semester and slightly less heavy towards the end, when your focus will shift towards small groups.

How we're going to interact:

Prof-and-students, f2f: regular short lectures (15-60 min), office hours available every week, video discussions (full class), video check-ins (with small groups).

Prof-and-students, remote: grading, slack, video feedback on some assignments, some prerecorded lectures.

Guest profs-and-students, short lectures, introductions to code

Students-to-students f2f. hreakout discussion: extra out-of-class meetings within small grouns

otuuenta-to-atuuenta, 121. pieanout ulabuaalon, eatua out-ol-olaaa meetinga witiini amail groupa.

Students-to-students, remote: a shared interface for marking up and interrogating the readings (via Perusall in Canvas), graded discussions (via Canvas), informal discussion and support (via Slack).

What you will do on your own:

Finish short coding exercises where you are working from a book.

Major projects will be pursued with a small group of other students, in groups that are invited to meet regularly with Dr. Guldi.

Assignment policies:

Skipping assignments: The "Short essays" and exams are required and may not be skipped. Failure to submit the short essays or to show up for exams will result in failing the course.

Late policy: One-third of a letter grade (A to A-) will be *automatically* subtracted for each 24-hour period of lateness.

Resubmission policy: Up to two assignments per semester may be re-submitted for a higher grade.

Plagiarism policy: Plagiarism and cheating will earn a grade of a zero (0) for the assignment and being dropped from the course.

Grading scale:

Participation grade: 20% of your grade is for participation. It will be assigned by the professor. The to reflect the student's level of engagement in classroom discussions. High scores will reflect evidence of having prepared for classroom discussions by completing all readings and problem sets on time.

Low grades: The lowest grade will be dropped from both the Coding/Writing and Readings sections.

Group work:

Students are encouraged to share code. They are encouraged to study for exams together.

Short essays must be completed by individuals working alone. It is appropriate to share data for the work behind the short essays. It is absolutely appropriate to seek the counsel of the course tutor,

Duan Cabaaffar during his office hours

ryan ochaener, during his onice hours.

Attendance:

Attendance is mandatory.

More than 3 unexcused absences will result in a failing grade.

Attendance is taken at the beginning of class. If students are not present for the beginning of class, they will be marked absent.

Health Policy:

If you are sick, please do not come to class. You should write immediately to Dr. Guldi explaining the nature of the absence. If you have Covid-like symptoms, please test yourself and report your findings to the SMU administration and to Dr. Guldi. Due dates will be moved to accommodate sickness upon consultation with the professor. Students who are sick or who take a personal health day are expected to catch up to the coursework on their own.

Mental health days: Students can take up to 3 classes as uncounted unexcused absence days. Such days recognize that people have extenuating circumstances that do not fall into the range of conventional excused absences. Students must notify the instructor when they are taking an uncounted unexcused absence day. Documentation is not required. Due dates for assignments will not be adjusted to accommodate mental health days.

Making up work: Students who are absent should work with other students to borrow lecture notes on the material missed and to study it on their own time.

Course Summary:

Date	Details	Due
Mon Jan 16, 2023	Week 1 (https://smu.instructure.com/calendar? event_id=252156&include_contexts=course_106487)	12am
	Hist 3368: Text Mining as Historical Method / CS 5393-001: Applied Text Mining	to do: 8am
	Hist 3368 Wednesday Meeting 1 (https://smu.instructure.com/calendar? event_id=251375&include_contexts=course_106487)	9am to 9:50am

Wed Jan 18, 2023

In-Class Discussions: GPT-3, **Data Visualizations of Obama and Trump, Digital History**

to do: 9:30am

First Assignment: Installing Slack, Reviewing Etiquette Guide (3 min reading + 10 min

due by 11:59pm

installation + 3 min assignment)

(https://smu.instructure.com/courses/106487/assignments/794999)



NO MEETING

(https://smu.instructure.com/calendar? event_id=251381&include_contexts=course_106487) 9am to 9:50am

Gillespie, The Relevance of

Algorithms due by 9am

(https://smu.instructure.com/courses/106487/assignments/799004)

Fri Jan 20, 2023

Guldi - Dangerous Art of Text

Mining - Introduction due by 9am

(https://smu.instructure.com/courses/106487/assignments/799000)

Underwood

(https://smu.instructure.com/courses/106487/assignments/799143) due by 9am



Slack Discussion: The Uses

of Text Mining

due by 11:59pm

(https://smu.instructure.com/courses/106487/assignments/795070)



Week 2

(https://smu.instructure.com/calendar? event id=252154&include contexts=course 106487) 12am

im Instructions for logging onto

M2 for the first time

(https://smu.instructure.com/calendar? event id=255727&include contexts=course 106487) 9am to 11:59pm

Mon Jan 23, 2023

Hist 3368 Monday meeting

(https://smu.instructure.com/calendar?

9am to 9:50am

event id=251361&include contexts=course 106487) Intro to Jupyter Notebooks (https://smu.instructure.com/calendar? 9:20am to 11:59pm event id=255728&include contexts=course 106487) First Problem Set: Intro to Jupyter + Python (30 min) due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795000) **Hist 3368 Wednesday Meeting** 2 9am to 9:50am (https://smu.instructure.com/calendar? event id=251376&include contexts=course 106487) Arnold and Tilton, New Data? The Role of Statistics in DH due by 9am (https://smu.instructure.com/courses/106487/assignments/799030) **Blaxill-Intro** (https://smu.instructure.com/courses/106487/assignments/799008) Wed Jan 25, 2023 Sacasas, Questions **Concerning Technology** due by 9am (https://smu.instructure.com/courses/106487/assignments/799007) **Collocates for** "Environmentalist" from Word to do: 9:45am **Embeddings Model of Congress,** 1970-2010 Hist 3368 Friday Meeting 2 (https://smu.instructure.com/calendar? 9am to 9:50am event id=251382&include contexts=course 106487) **Updating the Code -- a Practice to Use Each Time You** 9am **Begin** (https://smu.instructure.com/calendar? event_id=251389&include_contexts=course_106487)

Fri Jan 27, 2023	In class: Look at various word embeddings (https://smu.instructure.com/calendar? event_id=251477&include_contexts=course_106487) 9:15am
	Mini Problem Set: Navigating Lists and Dictionaries (10 m) due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795023)
	Short Assignment: Ask a Question, Answer a Question on Slack (15 min) (https://smu.instructure.com/courses/106487/assignments/795068)
	Week 3 (https://smu.instructure.com/calendar? event_id=252157&include_contexts=course_106487) 12am
Mon Jan 30, 2023	Hist 3368 Monday Meeting 2 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251370&include_contexts=course_106487)
	Chicago Style Citations Quiz 15 mins due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/794983)
	Early Semester Feedback due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/794996)
	Hist 3368 Friday Meeting 3 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251383&include_contexts=course_106487)
Fri Feb 3, 2023	Hist 3368 Wednesday Meeting 3 - Asynchronous Meeting on KWIC Problem Set (https://smu.instructure.com/calendar?

	event_id=251377&include_contexts=course_106487)
	Problem Set: KWIC due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795234)
Sat Feb 4, 2023	SHORT PAPER #1 – "How Congress Talked About the Environment" (https://smu.instructure.com/courses/106487/assignments/795052)
Mon Feb 6, 2023	Week 4 (https://smu.instructure.com/calendar? event_id=252158&include_contexts=course_106487) 12am
	Hist 3368 Monday Meeting 4 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251371&include_contexts=course_106487)
	Mini Problem Set: For Loops (https://smu.instructure.com/courses/106487/assignments/795022)
	Short Assignment: Ask a Question, Answer a Question #2 due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795066)
Wed Feb 8, 2023	Blaxill Chapter One due by 9am (https://smu.instructure.com/courses/106487/assignments/799014)
	Guldi - Dangerous Art of Text Mining - Chapter 1 due by 9am (https://smu.instructure.com/courses/106487/assignments/799011)
	Huijnen due by 9am (https://smu.instructure.com/courses/106487/assignments/799180)
Fri Feb 10, 2023	Two Mini Problem Sets: Cleaning and Counting due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795073)

	Week 5 (https://smu.instructure.com/calendar? event_id=252159&include_contexts=course_106487)
Mon Feb 13, 2023	Hist 3368 Monday Meeting 5 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251372&include_contexts=course_106487)
	Mini Problem Set: Wordcloud Barplot due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795024)
Wed Feb 15, 2023	Hist 3368 Wednesday Meeting 5 (https://smu.instructure.com/calendar? event_id=251378&include_contexts=course_106487) 9am to 9:50am
	Guldi - Dangerous Art of Text Mining - Chapter 3 due by 9am (https://smu.instructure.com/courses/106487/assignments/799020)
	Healy due by 9am (https://smu.instructure.com/courses/106487/assignments/799024)
	Munoz, Against Cleaning (https://smu.instructure.com/courses/106487/assignments/799015) due by 9am
	Problem Set: Ngrams (https://smu.instructure.com/courses/106487/assignments/795048) Que by 11:59pm
Fri Feb 17, 2023	Hist 3368 Friday Meeting 5 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251384&include_contexts=course_106487)
	Problem Set: Using Ngrams to Interpret Gender due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795051)
	Week 6 (https://smu.instructure.com/calendar? 12am

	event_id=252160&include_contexts=course_106487)
Mon Feb 20, 2023	Hist 3368 Monday Meeting 6 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251373&include_contexts=course_106487)
	Problem Set: Lemmatizing due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795044)
	Hist 3368 Wednesday Meeting (https://smu.instructure.com/calendar? event_id=251379&include_contexts=course_106487) 9am to 9:50am
Wed Feb 22, 2023	Guldi Chapter 12 due by 9am (https://smu.instructure.com/courses/106487/assignments/799459)
	Oreskes (https://smu.instructure.com/courses/106487/assignments/799473) due by 9am
Sat Feb 25, 2023	SHORT ESSAY #2 - "How Congress Talked About the Environment II" (https://smu.instructure.com/courses/106487/assignments/795233)
	Week 7 (https://smu.instructure.com/calendar? 12am event_id=252161&include_contexts=course_106487)
Mon Feb 27, 2023	Hist 3368 Monday Meeting 7 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251374&include_contexts=course_106487)
	Problem Set: WordNet (https://smu.instructure.com/courses/106487/assignments/795053) 11:59pm
	Hist 3368 Wednesday Meeting 7 (https://smu.instructure.com/calendar? event_id=251380&include_contexts=course_106487) 9am to 9:50am

Mad Mar 4, 2022	Guldi - Dangerous Art of Text Mining - Chapter 2 due by 9am (https://smu.instructure.com/courses/106487/assignments/799016)
Wed Mar 1, 2023	Guldi - Dangerous Art of Text Mining - Chapter 4 due by 9am (https://smu.instructure.com/courses/106487/assignments/799012)
	Nguyen 2 due by 9am (https://smu.instructure.com/courses/106487/assignments/799025)
	In-class student-led midterm review (https://smu.instructure.com/calendar?
	event_id=261516&include_contexts=course_106487) Hist 3368 Friday Meeting 7
Fri Mar 3, 2023	(https://smu.instructure.com/calendar? 9am to 9:50am event_id=251385&include_contexts=course_106487)
	Problem Set: Working with a Controlled Vocabulary due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795054)
Mon Mar 6, 2023	No Meeting (https://smu.instructure.com/calendar? 12am event_id=251396&include_contexts=course_106487)
Wed Mar 8, 2023	No meeting (https://smu.instructure.com/calendar? 12am event_id=251351&include_contexts=course_106487)
	Midterm Exam (https://smu.instructure.com/calendar? 12am event_id=251482&include_contexts=course_106487)
Fri Mar 10, 2023	Mid-semester feedback (https://smu.instructure.com/courses/106487/assignments/795021)

	Midterm Exam - Requires Respondus LockDown Browser + Webcam (https://smu.instructure.com/courses/106487/assignments/821556)	m
Mon Mar 13, 2023	Spring Break (https://smu.instructure.com/calendar? event_id=251479&include_contexts=course_106487) 12a	m
Wed Mar 15, 2023	Spring Break (https://smu.instructure.com/calendar? event_id=251480&include_contexts=course_106487) 12a	m
Fri Mar 17, 2023	Spring Break (https://smu.instructure.com/calendar? event_id=251481&include_contexts=course_106487) 12a	m
	Week 8 (https://smu.instructure.com/calendar? event_id=252162&include_contexts=course_106487) 12a	m
Mon Mar 20, 2023	Hist 3368 Monday Meeting 9 (https://smu.instructure.com/calendar? 9am to 9:50alevent_id=251388&include_contexts=course_106487)	m
	Problem Set: Working with Tabular Data due by 11:59pt (https://smu.instructure.com/courses/106487/assignments/795055)	m
Wed Mar 22, 2023	Hist 3368 Wednesday Meeting 10 (https://smu.instructure.com/calendar? event_id=251365&include_contexts=course_106487) 9am to 9:50ai	m
	Guldi - Chapter 11 due by 11:59pt (https://smu.instructure.com/courses/106487/assignments/795008)	m
	meeting (https://smu.instructure.com/calendar? event_id=251394&include_contexts=course_106487) 9am to 11:59pi	m

Fri Mar 24, 2023	Two Mini Problem Sets:
	Lemmatizing in Pandas & Plotting Change Over Time due by 11:59pm
	(https://smu.instructure.com/courses/106487/assignments/795074)
	₩eek 9
	(https://smu.instructure.com/calendar? 12am event_id=252163&include_contexts=course_106487)
Mon Mar 27, 2023	Hist 3368 Monday Meeting 10 (https://smu.instructure.com/calendar? 9am to 9:50am
	(https://smu.instructure.com/calendar? 9am to 9:50am event_id=251354&include_contexts=course_106487)
	Problem Set: Grouping Data (https://smu.instructure.com/courses/106487/assignments/795041)
	9am to 9:50am
	(https://smu.instructure.com/calendar? event_id=251366&include_contexts=course_106487)
Wed Mar 29, 2023	Guldi - Chapter 7 (https://smu.instructure.com/courses/106487/assignments/799026) due by 9am
	Guldi - Dangerous Art of Text Mining Chapter 6
	Mining - Chapter 6 due by 9am (https://smu.instructure.com/courses/106487/assignments/799021)
Fri Mar 31, 2023	Problem Set: Controlled Vocabulary Over Time due by 11:59pm
FIT Wat 31, 2023	(https://smu.instructure.com/courses/106487/assignments/798866)
	₩ Week 10 (https://emu.instructure.com/colondor?
	(https://smu.instructure.com/calendar? 12am event_id=252164&include_contexts=course_106487)
Man Ann 2, 2002	Hist 3368 Monday Meeting 12
Mon Apr 3, 2023	(<u>https://smu.instructure.com/calendar?</u> 9am to 9:50am event_id=251355&include_contexts=course_106487)

	Problem Set: Interpret Word Change Over Time due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/800106)
Wed Apr 5, 2023	Hist 3368 Wednesday Meeting 12 (https://smu.instructure.com/calendar? event_id=251367&include_contexts=course_106487) 9am to 9:50am
	Traugott, "Semantic Change" (2017) due by 9am (https://smu.instructure.com/courses/106487/assignments/800179)
Fri Apr 7, 2023	Good Friday Holiday - no meeting (https://smu.instructure.com/calendar? event_id=266346&include_contexts=course_106487) 12am
Sat Apr 8, 2023	SHORT PAPER #3 – "How Congress Talked About the Environment III" (https://smu.instructure.com/courses/106487/assignments/795263)
	Week 11 (https://smu.instructure.com/calendar? 12am event_id=252165&include_contexts=course_106487)
Mon Apr 10, 2023	Hist 3368 Monday Meeting 13 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251356&include_contexts=course_106487)
	Hist 3368 Wednesday Meeting 13 (https://smu.instructure.com/calendar? event_id=251368&include_contexts=course_106487) 9am to 9:50am
	Blatt 2 due by 9am (https://smu.instructure.com/courses/106487/assignments/799027)
Wed Apr 12, 2023	Gavin et al., "Spaces of

	Meaning" (2019) due by 9am (https://smu.instructure.com/courses/106487/assignments/800180)
	Guldi - Chapter 8 (https://smu.instructure.com/courses/106487/assignments/795011) due by 9am
	Text Mining South Park full text due by 9am (https://smu.instructure.com/courses/106487/assignments/799028)
Fri Apr 14, 2023	Hist 3368 Friday Meeting (https://smu.instructure.com/calendar? 9am event_id=251506&include_contexts=course_106487)
	Problem Set: Distinctiveness (https://smu.instructure.com/courses/106487/assignments/798872)
Mon Apr 17, 2023	Week 12 (https://smu.instructure.com/calendar? 12am event_id=252166&include_contexts=course_106487)
	Hist 3368 Monday Meeting 14 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251357&include_contexts=course_106487)
Wed Apr 19, 2023	Hist 3368 Wednesday Meeting (https://smu.instructure.com/calendar? event_id=251488&include_contexts=course_106487) 9am
	Problem Set: Intro to POS (https://smu.instructure.com/courses/106487/assignments/798876) due by 11:59pm
Fri Apr 21, 2023	Professor is away - no meeting (https://smu.instructure.com/calendar? event_id=251503&include_contexts=course_106487) 9am
	Problem Set: Advanced POS (https://smu.instructure.com/courses/106487/assignments/798873) due by 11:59pm

Mon Apr 24, 2023	Week 13 (https://smu.instructure.com/calendar? 12am event_id=252167&include_contexts=course_106487)
	Hist 3368 Monday Meeting 15 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251358&include_contexts=course_106487)
	Hist 3368 Wednesday Meeting 15 (https://smu.instructure.com/calendar? event_id=251369&include_contexts=course_106487) 9am to 9:50am
Wed Apr 26, 2023	Ouellette due by 9am (https://smu.instructure.com/courses/106487/assignments/799031)
	Problem Set: Word Embeddings due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/798874)
Fri Apr 28, 2023	Hist 3368 Friday Meeting (https://smu.instructure.com/calendar? event_id=251486&include_contexts=course_106487) 12am
Sat Apr 29, 2023	SHORT PAPER #4: "Trump vs. Obama" due by 11:59pm (https://smu.instructure.com/courses/106487/assignments/795349)
Mon May 1, 2023	Student-Organized Review Session (https://smu.instructure.com/calendar? event_id=251502&include_contexts=course_106487) 12am
	Hist 3368 Monday Meeting 16 (https://smu.instructure.com/calendar? 9am to 9:50am event_id=251359&include_contexts=course_106487)
	Hist 3368 FINAL EXAM (https://smu.instructure.com/calendar? event_id=252475&include_contexts=course_106487) 8am to 11am

Sat May 6, 2023

Final Exam- Requires

Respondus LockDown Browser

due by 10:30am

(https://smu.instructure.com/courses/106487/assignments/835835)

Extra Credit

(https://smu.instructure.com/courses/106487/assignments/794997)

Interdisciplinarity revisited

(https://smu.instructure.com/courses/106487/assignments/794984)

Participation Grade

(https://smu.instructure.com/courses/106487/assignments/795028)

Problem Set: Lemmatizing

(https://smu.instructure.com/courses/106487/assignments/795045)

Quiz for "This is not a List"

(15 min)

(https://smu.instructure.com/courses/106487/assignments/795057)

Roll Call Attendance

(https://smu.instructure.com/courses/106487/assignments/795063)

UNFINISHED Readings: Basic

Textual Analysis Methods

(https://smu.instructure.com/courses/106487/assignments/795076)