Course Description
This course explores the syntax, sentiment and hidden information contained in text documents. Students will learn how to frame text problems, choose NLP models, represent text to a computer, identify and execute machine learning techniques and evaluate model output. Includes information retrieval, natural language processing, text classification, summarization and stylometrics. Familiarity with programming is recommended.

Class Time
This class is offered asynchronously online. While you control when you watch videos and work on assignments, be aware of course pacing and specific deadlines.

Instructor Information
Dr. Robert P. Schumaker
Professor, Computer Science Dept.
rschumaker@uttyler.edu

Office Hours
DM through Slack (preferred), Zoom, email

If your inquiry is grade-related, please make a Zoom or physical appointment

Textbook Information
Text Mining with R (Slige and Robinson)
ISBN: 978-1-491-98165-8

Course Objective
- Identify the appropriate analytical tool based on the problem type and characteristics
- Properly conduct a collection and analysis of big-data
- Relate the tools learned to appropriate classes of problems
- Analyze a variety of business datasets for non-trivial patterns
- Interpret results and appropriately explain them to business managers

Computer Account Access
Students will need a Patriot account and password for computer access. This information can be found at `https://www.uttyler.edu/ccs`

Course Documents and Slides
This class will use Canvas for course documents, slides, quizzes and other class-related materials. Students are encouraged to check the website frequently during the course of the semester to keep up to date about course activity.

Course Grading

Course evaluation will be based on the following:

Contribution to Learning (CTL)	50
Student Reflection Paper (SRP)	50
Total Points	100

Grading Scale

A 85.0 points or more
B 70.0 to 84.999 points
C 55.0 to 69.999 points
D 40.0 to 54.999 points
F 39.999 points or less

Tentative Course Schedule and Assignments

Date	Concept	Readings
Aug 21-27	Introduction to Text Analytics	
	Text Analytics with R	Ch 1
Aug 28-Sep 3	tf·idf, HITS, PageRank and Search Engines	Ch 3
	Tokenization, Stemming and Stopwords	Ch 4
Sep 4-10	Parts of Speech	
	Sentiment Analysis	Ch 2
Sep 11-17	Summarization	
	Topic Modeling	Ch 6
Sep 18-24	Chatbots and LLMs	
	Stylometrics	
Sep 25-Oct 1	Data Collection: Computer Logs	Ch 9
	Data Collection: Financial News	Ch 5
Oct 2-7	Data Collection: Twitter	
	Fun with Twitter	Ch 7