Course Type	Course Code	Name of the Course	L	T	P	Credits
DC	NHSC507	Text Mining	3	1	0	4

Course Objective

The students

- receive theoretical foundation and hands-on experience in text mining techniques for humanities and social science enquiry.
- develop practical knowledge of various text mining methods, tools and other critical enquires.
- become acquainted with recent scholarly debates in the digital humanities

Learning Outcomes

Upon successful completion of this course, the students will be

- familiar with relevant concepts, methods and tools of text mining for humanities and social sciences.
- gained insights about various advantages and limitations of various text mining methods, modules, software and web applications.
- able to choose appropriate methods and tools for humanities and social sciences text mining projects.

Unit No.	Topics to be Covered	Lecture Hours	Learning Outcome
1.	**Introduction to the Course**: Theoretical foundations, questions and reasonings of text mining: What is text mining? and what is its role in humanities and social sciences?	4L+1T	Introducing students about text mining, its historical, conceptual and technological background
2.	**Text as Data:** Introducing text as data and its advantages, constraints and limitations	4L+1T	Students would learn the critical framework of text as a data and its various types, advantages, constraints and limitations
3.	**Selecting materials, building corpus and preprocessing techniques:** Introduce and discuss how to select materials and build corpus and the impact of preprocessing in text mining, including Text Encoding Initiative (TEI) models	5L+2T	Students would acquire knowledge in setting up criteria for selecting materials and building different types of corpora. Additionally, they would learn the impact of preprocessing such as stop words, lemmatization and stemming etc. in text mining.
4.	**Modellings in text mining:** Theoretical introduction to supervised and unsupervised methods in text mining	5L+2T	Introducing students to two important modeling in text mining and their application to various humanities and social sciences enquiries.
5.	**Supervised modeling:** Introduction to supervised modeling and Sentiment analysis. Studying and experimenting with sentiment analysis method and its variants	6L+2T	Students would obtain knowledge in the application of sentiment analysis to humanities and social sciences materials and its advantages and constraints
6.	**Unsupervised and semantic modeling:** Introduction to unsupervised and semantic modeling. Studying and experimenting with topic modeling method and its variants	6L+2T	Students would be introduced to one of the most important and widely used methods of text mining in digital humanities. They would also learn the advantages and limitations of the use of various topic modeling methods to different types of contents (long and short content).
7.	**Named Entity Recognition:** Identifying and classifying named entities within humanities materials into predefined categories such as names of persons, organizations, locations, dates, and more.	6L+2T	Students would acquire knowledge in named entity recognition method, classification of entities and the steps to extract entities.
8.	**Word embedding:** Identifying words with similar meanings are closer to each other in the vector space, capturing semantic relationships between words	6L+2T	Students would learn relational model word embedding and how it represents the relationships between words in a way that allows computers to understand and work with words more easily
TOTAL LECTURE HOURS		**42L + 14T**	

Text Books:

1. Moretti, Franco. *Graphs, Maps, Trees*. Verso, 2005.
2. Ramsay, Stephen. *Reading Machines: toward an Algorithmic Criticism*. University of Illinois Press, 2011.
3. Jockers, Matthew L. *Macroanalysis Digital Methods and Literary History.* University of Illinois Press, 2013.

Reference Books:

1. Guldi, Jo. *The Dangerous Art of Text Mining*. Cambridge University Press, 2023.
2. Moretti, Franco. *Distant Reading*. Verso, 2013.
3. Jockers, Matthew L. *Text Analysis with R for Students of Literature*. Springer, 2014.