



Mehran University of Engineering and Technology, Jamshoro
Department of Software Engineering

ORIGINAL SUBMITTED SYLLABUS

Title of Subject	:	Text Mining and Language Processing
Code	:	SE820
Discipline	:	Software Engineering
Effective	:	14PhD-IICT Batch and onwards
Pre-requisite	:	Data Mining concepts & Techniques
Assessment	:	10% Sessional 30% Mid Semester 60% Final Semester Examination
Credit Hours	:	03 + 0
Minimum Contact Hours:	42	Marks : 100

Specific Objectives of course:

- The course will enable students to to apply algorithms, resources and techniques for implementing and evaluating text analytics systems. Further, this course will demonstrate familiarity with text analytics application. The main goal of this course is to increase student awareness of the power of large amount of text data and computational methods to find patterns in large text corpora. This course will introduce the concepts and methods of text mining technologies rooted from machine learning, natural language processing, and statistics.

Course outline:

- **Introduction to Natural Language Processing (NLP)**
Overview of the field; defining problems --- machine translation, question answering, information retrieval; history of approaches to NLP --- early knowledge-based approaches, heuristic techniques, statistical approaches, and approaches combining statistics and logic (or knowledge); examples of why language is a tricky thing to process.
- **Document Representation**
how to convert text to feature vectors, what to count and how to count (stopword, stemming, tfidf, ngram, collocation), Information extraction, and enrich document representations by using information extraction tools and external linguistic resources
- **Document Classification**
naïve Bayes, nearest-neighbor, Text classification and feature selection: SVM, The many facets of text: classifying text by topic, author, opinion, genre, and style
- **Document Clustering**
Document clustering (hierarchical clustering and k-Means)
- **Text Analytics**
Opinion mining and sentiment analysis

BOOKS RECOMMENDED

- Miner, G. (Latest Edition). Practical text mining and statistical analysis for non-structured text data applications. Academic Press.
- Weiss, S. M., Indurkha, N., Zhang, T. (Latest Edition). Fundamentals of Predictive Text Mining. Springer: New York.
- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. (Latest Edition). Introduction to Information Retrieval, Cambridge University Press.
- Aggarwal, Charu C., and ChengXiang Zhai. (Latest Edition). Mining text data. Springer Science & Business Media.

Approval:

Board of Studies:	Res.No.02.02	Dated:23-10-2017
AS&RB:	Res.No.13.5	Dated:03-04-2018
Academic Council:	Res.No.92.15	Dated:23-04-2018