



# Jansons Institute of Technology

Karumathampatti, Coimbatore – 641 659  
Approved by AICTE and Affiliated to Anna University  
An ISO 9001:2015 certified institution

Department of Computer Science and Engineering

Industry Readiness Course - Syllabus

## Text Analysis Using NTLK

2019 - 2020 EVEN Semester

### Course Objectives:

- To interpret about the basics of text analysis.
- To use various text classifications techniques.
- To apply the fundamental of information retrieval and NTLK from NLP.

### Course Outcomes:

- To implement the text analytics framework.
- To analyze various sources of text data.
- To interpret the results, gain insights, and recommend possible actions from analytics performed on text data using NTLK.

### Syllabus:

<b>Module 1</b>	<b>Working with Text in Python and NLTK</b>	<b>10</b>
Understanding Text- Introduction to Text Mining -Handling Text in Python - Data Formats - Cleaning Data Sets – Duplicate Detection - Tagging Text - Indexing and Search - Evaluating Algorithms - Working with Text Data - Regular Expressions - Demonstration: Regex with Pandas and Named Groups - Internationalization and Issues with Non-ASCII Characters - Basic Natural Language Processing - Basic NLP tasks with NLTK - Advanced NLP tasks with NLTK		
<b>Module 2</b>	<b>Text Classifications</b>	<b>10</b>
Classification of Text - Text Classification techniques- Identifying Features from Text - Naive Bayes classifiers - Naive Bayes Variations - Support Vector Machines - Learning Text Classifiers in Python- Case Study : Sentiment Analysis		
<b>Module 3</b>	<b>Natural Language Toolkit</b>	<b>10</b>
Python NLTK Library - Applications of NLP - Text Pre-processing - Stemming, Lemmatization, Part of Speech Tagging Getting Started With Python's NLTK – Tokenizing - Filtering Stop Words – Stemming - Tagging Parts of Speech – Lemmatizing – Chunking – Chinking - Using Named Entity Recognition (NER) - Getting Text to Analyze - Using a Concordance -Making a Dispersion Plot - Making a Frequency Distribution - Finding Collocations.		

### Course Designed By

Dr E S Shamila,  
Prof / CSE

Industry Readiness Coordinator

Dr A Velayudham,  
Prof & Head /CSE

**Dr. A. VELAYUDHAM, M.E., Ph.D.,**  
Head of Department,  
Dept. of Computer Science and Engineering,  
Jansons Institute of Technology,  
Karumathampatti, Coimbatore - 641 659.

**PRINCIPAL**  
**JANSONS INSTITUTE OF TECHNOLOGY**  
**KARUMATHAMPATTI**  
**COIMBATORE - 641 659**