BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES

COURSE HANDOUT
Part A: Content Design

Course Title	Natural Language Processing Applications
Course No(s)	
Credit Units	4 units
Course Author	Dr. Chetana Gavankar
Version No	1.0
Date	September 2023

Course Objectives

No	Course Objective
CO1	To provide students with the knowledge on designing and applying algorithms for real life NLP Applications
CO2	To give an overview of the major technologies used in NLP and hands-on experience of using such tools
CO3	To apply NLP techniques in state of art applications like Machine Translation, Information Extraction including Named entity recognition and Relation extraction
CO4	To give students opportunities to sharpen their programming skills for Natural Language Processing applications

Text Book(s)

T1	Speech and Language processing: An introduction to Natural Language Processing, Computational Linguistics and speech Recognition by Daniel Jurafsky and James H. Martin

Reference Book(s) & other resources

R1	Manning and Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA
R2	Neural Machine Translation by Philipp Koehn
R3	Knowledge Graphs Methodology, Tools and Selected Use Cases by Dieter Fensel , Umutcan Şimşek, Kevin Angele, Elwin Huaman , Elias Kärle , Oleksandra Panasiuk , Ioan Toma, Jürgen Umbrich, and Alexander Wahler, Springer 2019

R4	Natural Language Toolkit. Bird and Loper, and other developers. Available for free at: – http://www.nltk.org/

Modular Content Structure

1. **Overview of the course**
2. **Grammar and spellcheckers**
    - Rule based
    - Statistical
    - NN based
3. **Knowledge Graph Applications**
    - How to Use Knowledge Graphs
        - Merging Artificial Intelligence and Internet
        - Knowledge Access Layer
        - Open and Service-oriented Dialog Systems
    - Why we need Knowledge Graphs
        - Motivation and Solution
        - Touristic Use Cases
        - Energy Use Cases

      Further Verticals
4. **Question Answering and Chabot's**
    - IR-based Factoid Question answering
    - Knowledge-based Question Answering
    - NN based QA
    - Using multiple information sources: IBM's Watson
    - Evaluation of Factoid Answers
    - Properties of Human Conversation
    - GUS: Simple Frame-based Dialogue Systems
    - Dialogue System Design

4. **Machine Translation**
    - Statistical
    - Neural
    - Indic Languages

5. **Information Extraction**
    - Named Entity Recognition
    - Relation Extraction
    - Extracting Events and Time

6. **Sentiment Analysis**
    - Sentiment Analysis Methods
    - Rule based, ML based and Hybrid Systems
    - Neural Networks for Sentiment Analysis - NLP features

Part B: Contact Session Plan

Academic Term	
Course Title	
Course No	
Lead Instructor	

Course Contents

Contact session	List of Topic Title (from content structure in Part A)	Topic # (from content structure in Part A)	Text / Ref Book / External resource
1	**Introduction to course and NLP applications**		
2-3	**Grammar and spellcheckers (2 contact sessions)** • Rule based • Statistical • NN based		**T1-Appendix B + additional resources**
4-5	**Question Answering and Chabot's (2 contact sessions)** • IR-based Factoid Question answering • Knowledge-based Question Answering • NN based QA • Evaluation of QA and Chatbots • Properties of Human Conversation • GUS: Simple Frame-based Dialogue Systems • Dialogue System Design		**T1- chapter 25 and 26 + additional web resources**

6-7	**Knowledge Graph Applications (2 contact sessions)**
	• How to Use Knowledge Graphs
	  ○ Merging Artificial Intelligence and Internet
	  ○ Knowledge Access Layer
	  ○ Open and Service-oriented Dialog Systems
	• Why we need Knowledge Graphs
	  ○ Motivation and Solution
	  ○ Touristic Use Cases
	  ○ Energy Use Cases
	  ○ Further Verticals
			R3**- chapter 3 + chapter 4**

8	**Session 1 to Session 7 Review**		
9-11	**Machine Translation (3 sessions)** • Statistical • Neural • Indic Languages		**T1- chapter 10 + R2 chapter 5 + additional web resources**
12-14	**Information Extraction (3 contact sessions)** • Named Entity Recognition		**T1- chapter 18 + additional web**

	• Relation Extraction • Extracting Events and Time		**resources**
15	**Sentiment Analysis** • Sentiment Analysis Methods • Logistic Regression and Naïve Bayes Models, • Neural Networks for Sentiment Analysis		**T1- chapter 4 + additional web resources**
16	**Session 9 to Session 15 Review**		

Detailed Plan for Lab work

Lab No.	Lab Objective	Session Reference
1	Introduction to NLTK, Spacy and other open source tools	1
2	Grammars and Spellcheckers	2-3
3	Question Answering and Chabot's	4-5
4	Knowledge Graph Applications	6-7
5	Machine Translation	9,10,11
6	Named Entity Recognition	12
7	Relation Extraction	13
8	Sentiment Analysis	15

Evaluation Scheme

Evaluation Component	Name (Quiz, Lab, Project, Midterm exam, End semester exam, etc)	Type (Open book, Closed book, Online, etc.)	Weight	Duration	Day, Date, Session, Time
EC – 1	Quiz	Open book	10%		To be announced
EC – 1	Assignments (2)	Open book	30%		To be announced
EC – 2	Mid-term Exam	Open book	30%		To be announced
EC – 3	End Semester Exam	Open book	30%		To be announced

Important Information

Syllabus for Mid-Semester Test (Closed Book): Topics in Weeks 1-8 (1-18 Hours)
Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study

Notes

- Quiz and Assignments timelines will be announced on the canvas portal.
- **Deadlines for evaluation components will NOT be extended** and the student is requested not to wait for the deadline to start working on Quiz/Assignment
- Syllabus for Mid-Semester Test (Closed Book): Topics in Session Nos. 1 to 8
- Syllabus for Comprehensive Exam (Open Book): All topics (Session Nos. 1 to 16)
- **Strictly NO MAKEUPS for Quiz and Assignments** and all submissions after the announced deadlines will not be considered for evaluation.
- **All assignments will be subjected to plagiarism check, and if violated will be subject to disciplinary action apart from nullifying all the marks/grades assigned.**

Important links and information:

Canvas: Students are expected to visit the Canvas portal on a regular basis and stay up to date with the latest announcements and deadlines.

Contact sessions: Students should attend the online lectures as per the schedule provided.

Evaluation Guidelines:

1. EC-1 consists of Assignments and Quizzes. Announcements regarding the same will be made in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.