|  |  |  |
| --- | --- | --- |
| IQRA University (IU) | | |
| Faculty of Engineering Sciences and Technology (FEST) | | |
| Artificial Intelligence Department (AI) | | |
| Course Code | Course Name | Credit Hr |
| AIN472 | Natural Language Processing | 3+0 |

|  |  |  |  |
| --- | --- | --- | --- |
| 1. Basic Information | | | |
| Instructor |  | Designation |  |
| Prerequisite(s) | Artificial Neural Networks | Semester | Spring 2025 |
| Email | [Dr.jabeen@iqra.edu.pk](mailto:Dr.jabeen@iqra.edu.pk) | Phone |  |
| Consulting Hours | Wednesday – Thursday 2-3:30 | Office Location |  |

|  |
| --- |
| 1. **Course Objective(s)** |
| Natural Language Processing (NLP) is the application of computational techniques to the analysis and synthesis of natural language and speech. This course is an introduction to NLP with prior programming experience in Python. |

|  |
| --- |
| 1. **Course Contents** |
| Introduction & History of NLP, Parsing algorithms, Basic Text Processing, Minimum Edit Distance, Language Modeling, Spelling Correction, Text Classification, Deterministic and stochastic grammars, CFGs, Representing meaning /Semantics, Semantic roles, Semantics and Vector models, Sentiment Analysis, Temporal representations, Corpus-based methods, N-grams and HMMs, Smoothing and backoff, POS tagging and morphology, Information retrieval, Vector space model, Precision and recall, Information extraction, Relation Extraction (dependency, constituency grammar), Language translation, Text classification, categorization, Bag of words model, Question and Answering, Text Summarization |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1. **Course Learning Outcomes** | | | | | | |
| **CLOs** | **CLO Statement** | **BT Level** | **Mapping** | | | **% Weight** |
| **GAs** | **ACM KA** | **SDGs** |
| CLO1 | **Understand** techniques for information retrieval, language translation, and text classification. | C2 | GA3 | #4  DS | 4 | 40% |
| CLO2 | **Apply** machine learning models to solve NLP tasks like text classification, sentiment analysis, and named entity recognition (NER). | C3 | GA4 | 20% |
| CLO3 | **Solve** classic and stochastic algorithms for parsing natural language. | C3 | GA10 | 40% |
| ***Note: On successful completion of course, GA 1 (Academic Education) will automatically attain.*** | | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| 1. **Course Textbook / Reference Books and Supplementary Reading Material** | | | |
| **S No** | **Book Title** | **Author(s)** | **Edition/ publication year/publisher** |
|  | Speech and Language Processing, An Introduction to Natural Language Processing | Jurafsky, Daniel, and James H. Martin | Pearson, 2018, Third Edition. Prentice Hall |
|  | Neural Network Methods for Natural Language Processing | Goldberg, Yoav | Morgan & Claypool |
|  | Foundations of Statistical Natural Language Processing, Manning and Schütze, |  | MIT Press. Cambridge, MA: May 1999 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1. **CLO Outcome Based Assessment (OBA)** | | | | | | |
| **Assessment Tool** | | **CLO Mapped** | **CLO Marks** | **% Weight** | **Total Marks** | **Assessment Date** |
| **Quizzes**  **15** | Quiz #1 | *CLO1* | 5 | 33.33% | 5 | TBD |
| Quiz #2 | *CLO2* | 5 | 33.33% | 5 | TBD |
| Quiz #3 | *CLO3* | 5 | 33.33% | 5 | TBD |
| **Total Quizzes %** | | | **100%** | **15** |  |
|  | | | | | | |
| **Assignments**  **10** | Assignment #1 | *CLO1* | 5 | 50% | 5 | TBD |
| Assignment #2 | *CLO2* | 5 | 50 % | 5 | TBD |
|  |  |  |  |  |  |
|  | **Total Assignments %** | | | **100%** | **10** |  |
|  | | | | | | |
| **Project/CCP**  **10** | Project/CCP | *CLO1-3* | **10** | 100% | **10** | TBD |
|  | **Total Project /CCP %** | | | **100%** | **10** |  |
|  | | | | | | |
| **Midterm**  **25** | Midterm Q#1 | *CLO1* | **13** | 52% | 13 | TBD |
| Midterm Q#2 | *CLO2* | **12** | 48% | 12 |  |
|  | **Total Midterm %** | | | **100%** | **25** |  |
|  | | | | | | |
| **Final Exam**  **40** | Final Exam Q#1 | *CLO1* | 10 | 25% | 10 | TBD |
| Final Exam Q#2 | *CLO2* | 10 | 25% | 10 |  |
| Final Exam Q#3 | *CLO3* | 20 | 50% | 20 |  |
|  | **Total Final Exam %** | | | **100%** | **40** |  |
| **100** | **Total Marks** | | | | **100** |  |
| ***Note: Please make sure every CLO must be assessed at least 3 times.*** | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Weekly Plan** | | | | |
| **Week** | **Lecture No** | **Topic Covered** | **CLO** | **Assessment Tool** |
| 1 | 1 | **Introduction to NLP** | **1** |  |
| 2 | Introduction to NLP, history, applications, and challenges. | **1** |  |
| 3 | Overview of linguistic fundamentals: Phonology, morphology, syntax, and semantics. | **1** |  |
| 2 | 4 | Regular Expression, Text Normalization | **1** |  |
| 5 | **Text Preprocessing** | **1** |  |
| 6 | Text preprocessing techniques: Tokenization, stop-word removal, stemming, lemmatization. | **1** |  |
| 3 | 7 | Edit Distance in NLP | **1** |  |
| 8 | Representation of text: Bag-of-Words (BoW), TF-IDF, and word n-grams. N-gram Language Models | **1** |  |
| 9 | N-gram Language Models | **1** |  |
| 4 | 10 | Bag of Words (BoW) Model | **1** |  |
| 11 | TF-IDF (Term Frequency-Inverse Document Frequency) | **1** |  |
| 12 | Limitations of frequency-based models | **1** |  |
| 5 | 13 | Evaluation of language models: Perplexity and cross-entropy. | **1** |  |
| 14 | Information Retrieval | **1** |  |
| 15 | Word Embeddings | **1** |  |
| 6 | 16 | Distributed word representations: Word2Vec, Skip-Gram, and CBOW. | **1** |  |
| 17 | Introduction to Distributed Representations, Word2Vec (Skip-Gram and CBOW), GloVe (Global Vectors for Word Representation) and FastText | **1** |  |
| 18 | Part-of-Speech (POS) Tagging Naïve | **1** | **A1** |
| 7 | 19 | POS Tagging Concepts and Applications, Rule-Based vs. Statistical POS Tagging. | **2** | **Q1** |
| 20 | Pre-trained POS Tagging Models (NLTK, spaCy) | **2** |  |
| 21 | **Project / CCP Discussion and Group allocation** | **1-3** |  |
| **8** | **Midterm Exam** | | | |
| 9 | 22 | Bayes and Sentiment Classification | **2** |  |
| 23 | **Named Entity Recognition (NER)** | **2** |  |
| 24 | Pre-trained NER Models, Applications of NER (news, healthcare, finance) | **2** | **A2** |
| 10 | 25 | Sentiment Classification Concepts | **2** | **Q2** |
| 26 | Use Cases in Social Media and Customer Feedback | **2** |  |
| 27 | Sentiment Analysis Models | **2** |  |
| 11 | 28 | Text classification | **3** |  |
| 29 | Algorithms: Naive Bayes, SVM, Logistic Regression | **3** |  |
| 30 | Multiclass Text Classification | **3** |  |
| 12 | 31 | Extractive vs. Abstractive Summarization | **3** |  |
| 32 | Text Rank Algorithm | **3** |  |
| 33 | Transformer-Based Summarization | **3** |  |
| 13 | 34 | Speech-to-Text Conversion | **3** |  |
| 35 | ASR (Automatic Speech Recognition) Overview | **3** |  |
| 36 | Libraries and APIs (Google Speech-to-Text, Speech Recognition) | **3** |  |
| 14 | 37 | QA System Overview | **3** |  |
| 38 | Information Retrieval-Based QA | **3** |  |
| 39 | BERT-Based QA Models | **3** | **A3** |
| 15 | 40 | Introduction to Transformers, BERT, GPT, T5 Overview, Transfer Learning and Fine-Tuning | **3** | **Q3** |
| 41 | Model Deployment Techniques, APIs and Cloud Platforms (AWS, Google Cloud) | **3** |  |
| 42 | Monitoring and Updating Models | **3** |  |
| 16 | 43 | **Project / CCP presentations and submissions** | **1-3** |  |
| 44 |  |
| 45 |  |
| **17** | **Final Exam** | | | |

|  |  |
| --- | --- |
| 1. **IU Assessment / grading Policy** | **Instructor grading for course** |
| Quizzes 10-15%  Assignments 10-15%  Projects/Presentation/CCP 0-10%  Mid Semester Examination/ 20-30%  End Semester Examination 40-50% | 15  15  05  25  40 |