****

Course Title: Natural Language Processing

Course Code: IS-679

Department: Information Systems Department

**HSM Vision**

HSM envisions its success in the sustainable contribution that it will make to the industry, academia and research in public and private sector. HSM will lead by providing professionally competent and ethically conscious human resources engaged in the global and local context to foster socio-economic growth and sustainability for the society. HSM envisages having faculty with high research potential and a deep desire for cutting edge research including collaboration with national and international partners.

**HSM Mission**

Being a research-oriented and student-centric business school, we emphasize research publications in impact journals as well as state-of -the-art learning methodologies.  We will prepare our students to become the future ethical business leaders and the guiding post for the society, while equipping them with the knowledge and skills required by world-class professionals.  We will be the leading choice for organizations seeking highly talented human resource. HSM will foster internationalization with key stakeholders and actively work to exchange best practices with business schools across Pakistan through collaborations, workshops, conferences and other means.

**Program Objectives**

The School of Business and Economics at UMT is foreseeing the challenges ahead both at national and international level and the utility of data science.  In Pakistan the multi dimensional economy integrated with globalization needs a boost assisted by professionally trained and skilled Data Scientists, whom may incorporate and harmonize the unlimited bucket of resources, pouring in from springs of industry, agriculture, business, human resources etc. in a manner to achieve efficiency to its apex.  
In the competitive economy the companies need to adapt data science to gain a competitive advantage in productivity, profitability and sustainable business processes to offer better products and services to their customers. To attain this goal trained and skilled workforce in this area is the need of the hour; who are equipped to manage, understand and model the data, interpret the outcome and communicate the results for business use. Professionals holding a degree in Data Science will be well positioned to help their organizations gain a competitive advantage in a data-driven world.

**Course Objectives**

The field of natural language processing is concerned with practical and theoretical issues that arise in getting computers to perform various tasks with human languages. The goal of natural language processing is to allow that kind of interaction so that non-programmers can obtain useful information from computing systems.

Whether you are interested in the intersection between the humanities and computer science, or you want a job at Google, this introductory course will help you on your way.

**Learning Objectives**

The course is an introduction to Natural Language Processing. The main objective of the course is to learn how to develop practical computer systems capable of performing intelligent tasks on natural language: analyze, understand and generate written text. This task requires learning material from several fields: linguistics, machine learning and statistical analysis, and core natural language techniques There are following objectives:

* Understand approaches to syntax and semantics in NLP.
* Understand approaches to discourse, generation, dialogue and summarization within NLP.
* Understand current methods for statistical approaches to machine translation.
* Understand machine learning techniques used in NLP, including hidden Markov models and probabilistic context-free grammars, clustering and unsupervised methods, log-linear and discriminative models, and the EM algorithm as applied within NLP.

**Learning Outcomes**

In this introductory course you will learn about techniques for filtering junk email, automatically discovering the different meanings of the word "run", efficiently encoding spelling rules, tagging words according to their part of speech, parsing English sentences, extracting from the Web names of companies employing UMass graduates, automatically translating from one language to another, and modeling language semantics. Our work will be a combination of learning new algorithms, discussing linguistics, and programming useful systems that operate on real data. The reason that a learning system is necessary is because the veracity of data is not always what one would desire. Hence NLP provide an extension to fulfil the desires of Sustainable environment of digital world.

**Teaching Methodology (List methodologies used –example are given below)**

Lecture

Interactive Classes

Case based teaching

Class activities

Presentation

Guests Speakers

Video Documentary

**STUDENTS ARE REQUIRED TO READ AND UNDERSTAND ALL ITEMS OUTLINED IN THE PARTICIPANT HANDBOOK**

**Class Policy:-**

* Be On Time

You need to be at class at the assigned time. After 10 minutes past the assigned time, you will be marked absent.

* Mobile Policy

**TURN OFF YOUR MOBILE PHONE!** It is unprofessional to be texting or otherwise.

* Email Policy

**READ YOUR EMAILS!** You are responsible if you miss a deadline because you did not read your email.

Participants should regularly check their university emails accounts regularly and respond accordingly.

* Class Attendance Policy

A minimum of 80% attendance is required for a participant to be eligible to sit in the final examination. Being sick and going to weddings are absences and will not be counted as present. You have the opportunity to use 6 absences out of 30 classes. Participants with less than 80% of attendance in a course will be given grade ‘F’ (Fail) and will not be allowed to take end term exams. International students who will be leaving for visa during semester should not use any days off except for visa trip. Otherwise they could reach short attendance.

* Withdraw Policy

Students may withdraw from a course till the end of the 12th week of the semester. Consequently, grade W will be awarded to the student which shall have no impact on the calculation of the GPA of the student. A Student withdrawing after the 12th week shall be automatically awarded “F” grade which shall count in the GPA.

* Moodle

UMT –LMS (Moodle) is an Open Source Course Management System (CMS), also known as a learning Management System (LMS). Participants should regularly visit the course website on MOODLE Course Management system, and fully benefit from its capabilities. If you are facing any problem using moodle, visit <http://oit.umt.edu.pk/moodle>. For further query send your queries to [moodle@umt.edu.pk](mailto:moodle@umt.edu.pk)

* Harassment Policy

Sexual or any other harassment is prohibited and is constituted as punishable offence. Sexual or any other harassment of any participant will not be tolerated. All actions categorized as sexual or any other harassment when done physically or verbally would also be considered as sexual harassment when done using electronic media such as computers, mobiles, internet, emails etc.

* Use of Unfair Means/Honesty Policy

Any participant found using unfair means or assisting another participant during a class test/quiz, assignments or examination would be liable to disciplinary action.

* Plagiarism Policy  
    
  All students are required to attach a “Turnitin” report on every assignment, big or small. Any student who attempts to bypass “Turnitin” will receive “F” grade which will count towards the CGPA. The participants submit the plagiarism report to the resource person with every assignment, report, project, thesis etc. If student attempts to cheat “Turnitin”, he/she will receive a second “F” that will count towards the CGPA. There are special rules on plagiarism for final reports etc. all outlined in your handbook.
* Communication of Results

The results of quizzes, midterms and assignments are communicated to the participants during the semester and answer books are returned to them. It is the responsibility of the course instructor to keep the participants informed about his/her progress during the semester. The course instructor will inform a participant at least one week before the final examination related to his or her performance in the course.

**Course Outline**

|  |  |
| --- | --- |
| Program | MS Data Sciences/ MS Business Analytics |
| Credit Hours | 3 |
| Duration | 15 Weeks |
| Prerequisites (If any) | N/A |
| Resource Person  Name and Email |  |
| Counseling Timing  (Room# 1N1 R#7 ) |  |
| Contact no. |  |
| Web Links:- (Face book, Linked In, Google Groups, Other platforms) |  |

**Chairman/Director Programme signature………………………………….Date……………………..**

**Dean’s signature…………………………… ………………….Date………………………………………….**

**Grade Evaluation Criteria**

Following is the criteria for the distribution of marks to evaluate final grade in a semester.

**Marks Evaluation Marks in percentage**

Quizzes 10%

Assignments 10%

Mid Term 20%

Attendance & Class Participation 5%

Term Project 15%

Case Study 10%

Presentations 5%

Final exam 35%

Total 100%

**Recommended Text Books:**

Jafsky, David, and James H. Martin (2000). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Upper Saddle River, NJ: Prentice-Hall. ISBN: 0130950696.

Manning, Christopher D. and HinrichSchütze( 1999). Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. ISBN: 0262133601.

**Reference Books:**

Natural Language Processing with Pytho Analyzing Text with the Natural Language Toolkit Steven Bird,

Introduction to Information Retrieval, by C. Manning, P. Raghavan, and H Schütze. Cambridge University Press, 2008

**Reference Videos:**

Video from Coursera by Professor Dan Jurafsky& Chris Manning on Natural Language Processing starting in March 19, 2012

Video from Coursera - Columbia University - Course: Natural Language Processing by Michael Collins

Video from Coursera- Introduction to Natural Language Processing University of Michigan by Dragomir R. Radev, Ph.D.

Natural Language Processing by Prof. Pushpak Bhattacharyya, Department of Computer science &Engineering,IIT Bombay.

**Course: -E-Business/E-Commerce----Course code: ------IS-472- ------------Book: -------------------------------------------------------------------**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Topics to be**  **covered in the course** | **Learning Objective**  **of this topic** | **Expected Outcomes from Students** | **Teaching Method** | **Assessment Criteria** | **Deadlines and Homework** |
| 1 | Ethics in NLP | Discussion of Course Overview ,objectives ,Importance of NLP, Job Market of NLP. Goal-oriented design for ethical machine learning and NLP | Students will be able to answer the Importance of NLP, Job Market of NLP and introduction of ethical machine learning. | Course Introductory Slide |  | Within a Week |
| 2 | Introduction to Natural language Processing | What is NLP , Stages of NLP , Application of NLP, Why NLP is Hard | Students will be able to Stages, Application of NLP | Lecture Slides | Discussion of AI Blog | Within a Week |
| 3 | Text Pre Processing | Regular Expression How Tokenization Works , Issues in tokenization, Morphology, Normalization in text Processing, stemming , Types of stemming, Lemmatization | They will be in position to understand Text Pre Processing terminologies, how they work and what the issues were. | Book Topics | Demo in regerp  Demo in python  **Assignment 1** | Within a Week |
| 4 | Segmentation | Word segmentation, Urdu word Segmentation Issues & algorithms, sentence segmentation | After this lecture, students will be able to apply the basic knowledge of segmentation. | Lecture slides | **Quiz 1**  Demo Jyputer Notebook  Guest Speaker 1 | Within a Week |
| 5 | Introduction of Language Modeling | Probability view , N Gram , bigram , trigram Models of language modeling, perplexity, Interpolation, Smoothing , Good Turing Smoothing , Kneser Ney smoothing | They will be able to learn the basics of Modeling Language. | Lecture Slide | Assignment2 | Within a Week |
| 6 | Spelling Correction | Spelling correction task, Noisy Channel Model. Defining and computing Minimum Edit Distance | Now they will understand and perform spelling corrections by themselves. | Lecture | Quiz # 2 | Within two Weeks |
| 7 | Part of Speech Tagging (POS) | Into to POS , Sequence models for POS, Hidden Markov model ,Veterbi Algorithm for HMM | After this, they can implement part of speech tagging. | Lecture Slide + Book Topics | Assignment 3  Guest Speaker 2 | Within two Weeks |
| 8 | Parsing | Syntactic Parsing and Context Free Grammar, Probabilistic Context Free Grammar , The CKY parsing algorithm I , constituency Parsing | They will be in position to understand the fundamentals of Parsing, how they work and what the issues were. | Lecture Slide + Book Topics | Quiz # 3 |  |
| 9 | Other Parsing | Lexicalized PCFGs , Dependency parsing ,semantic Parsing ,GLM for dependency | Now they will be able to use different parsing techniques | Lecture Slide + Book Topics | Quiz # 4 | Within a Week |
| 10 | Sentimental Analysis | Sentimental analysis ,base line algorithm, sentimental lexicons ,text classification with sentimental | After this lecture, they are able to perform sentimental analysis using algorithm and classified with sentimental. | Lecture Slide + Book Topics | Assignment 4 | Within a Week |
| 11 | Information Extraction | Name Entity Recognition , Sequence NER ,Maximum Entropy Sequence model , relation Extraction | Their understanding with various extraction techniques for information extraction. | Lecture Slide + Book Topics | Quiz # 5 | Within two Weeks |
| 12 | Question Answering | What is QA ,Answer type and query formation ,passage retrieval, knowledge in QA, Architecture of QA | They will be able to do lot while resolving QA query formatting and Architecture designed for QA | Lecture Slide + Book Topics |  | Within two Weeks |
| 13 | Summarization | Intro for text summarization, generating Snippets, Summarizing multiple documents | Now they will be in position to answer the more advance concepts in NLP including text summarization. | Lecture Slide + Book Topics |  | Within a Week |
| 14 | Machine Translation | Machine translation approaches , noisy channel,advance techniques of ML | After this lecture, they are able to determine Machine Translation. | Lecture Slide + Book Topics |  |  |
| 15 | Sustainability Analysis using NLP and Social Media Content |  |  | Presentation | Presentation |  |