IIT Patna Vishlesan I-Hub Foundation

Workshop Brochure for

First Workshop: Foundational NLP: Insight into Text Processing

of Workshop Series on

Advanced Workshop on Natural Language Processing (NLP) with AI/ML

IIT Patna Vishlesan I-Hub Foundation | **TIH_Vishlesan** | **tih_vishlesan** | **@TIH_Vishlesan** | **tihvishlesan** | **TIH IIT Patna**

IIT Patna Vishlesan I-Hub Foundation takes pleasure to announce the **1st** Workshop **"Foundational NLP: Insight into Text Processing"** of the Workshop-Series on **"Advanced Workshop on Natural Language Processing (NLP) with AI/ML"**. The workshop shall be held during **April 16-18, 2024**.

Address for Correspondence:

IIT Patna Vishlesan I-Hub Foundation
Indian Institute of Technology Patna,
Block-6, 4th Floor, Room No. 414
Bihta, Patna, Bihar, India, 801103

Point of Contact: **Dr. Pradeepika Verma**
Email: faculty.tih@iitp.ac.in
Mobile: **9335079535**

Note: All email correspondence must be addressed with the subject **"NLP Workshop 2024"**

Chief Patron:

Prof. T. N. Singh,
Chairman BoD, TIH IIT Patna

Program Chair:

Prof. Anup Kumar Keshri,
Project Director, TIH IIT Patna

Dr. Sai Kiran Oruganti,
CEO, TIH IIT Patna

Program Organizing Committee:

Dr. Pradeepika Verma,
Faculty Fellow, TIH IIT Patna

Mr. Rohit Prasad,
CTO, TIH IIT Patna

Ms. Rexona Khatoon,
Assistant Officer Training, TIH IIT Patna

About IIT Patna Vishlesan I-Hub Foundation:

IIT Patna Vishlesan I-Hub Foundation is one of the 25 Technology Innovation Hubs (TIHs) that have been established under the National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) implemented by The Department of Science & Technology (DST), Government of India through approval of by the Union Cabinet. The IIT Patna Vishlesan I-Hub Foundation also known as TIH IIT Patna established on the 8th of September 2020. The Hub has set up in the technology vertical of **Speech, Video & Text Analytics (SVTA).** It has wide-spread activities ranging from the fundamental basic research to translational research and development of cutting-edge technologies for creating solutions to serve the technological requirements in the areas of "Speech, Video and Text Analytics" for societal applications such as health, judiciary, education, national security, environment etc.; creating products for commercial use, nurture startups and increase the job market. Some of the important applications to be developed

include multilingual translation, chatbot, summarization system, misinformation detection system, edge computing-based health care platform, embedded with heterogeneous swarm systems for real time video analytics etc. Some of the important problems identified include the technology and product development for smart health care, swarm robotics for security, multilingual machine translation, summarization, misinformation detection, chatbot development, sentiment analysis, human activity recognition etc.

About the Natural Language Processing (NLP) with AI/ML Workshop Series:

The workshop series shall consist of the **five (5) unique workshops**. Each workshop series shall continue for **three (3) days**. The detailed workshop agenda is below:

Workshop Agenda:

Workshop Series Title: Advanced Workshop on Natural Language Processing (NLP) with AI/ML

- **Workshop 1: Foundational NLP: Insight into Text Processing**
- **Workshop 2:** Developing a Text Classifier
- **Workshop 3:** Collecting Text Data with Web Scraping and APIs and Topic Modelling
- **Workshop 4:** Vector Representation
- **Workshop 5:** Text Generation and Summarization

Workshop Duration:

Each workshop session shall be of 8 hours on each 3 days of workshop.

Delivery Format:

- The sessions will be conducted **online/live** (Microsoft Team/Google Meet Platform/Any Online Platform).
- It is the responsibility of the participant to ensure that they have reliable internet connectivity to attend all the online sessions.

Overall Objectives of the Workshop:
- Enhancing Predictive Analysis
- Optimizing Learning for Computers

- Exploring Language and Technology
- Innovative Text Processing
- Unveiling Insights through Text Mining

First Workshop Contents:

Workshop 1: Foundational NLP: Insight into Text Processing

Day 1 Schedule in Detail:

Morning Session (9:00 am – 10:30 am)
- Exploring the Power of NLP: Recent Applications and Future Prospects

Morning Session (10:30 am – 12:00 pm)
- Basic Tools for NLP
- Python Basics
- Exercise 1.1: Python Basics

Session Break: 12:00 pm to 2:00 pm

Afternoon Session (2:00 pm – 5:00 pm)
- Various Steps in NLP
  - Exercise 1.2: Tokenization of a Simple Sentence
  - Exercise 1.3: Stop Word Removal
  - Introduction to Techniques of PoS Tagging

- ⚑ Named Entity Recognition (NER)
- ⚑ Exercise 1.4: PoS Tagging
- ⚑ Exercise 1.5: Treating Named Entities

Day 2 Schedule in Detail:

Morning Session (9:00 am – 12:00 pm)
- Text Normalization
  - ⚑ Exercise 2.3: Text Normalization
  - ⚑ Exercise 2.4: Spelling Correction of a Word and a Sentence
  - ⚑ Stemming
  - ⚑ Exercise 2.5: Using Stemming
  - ⚑ Lemmatization
  - ⚑ Exercise 2.6: Extracting the Base Word Using Lemmatization
  - ⚑ Challenges in Stemming and Lemmatization

Session Break: 12:00 pm to 2:00 pm

Afternoon Session (2:00 pm – 5:00 pm)
- Text Normalization for Low-Resource Languages
  - ⚑ Stemming
  - ⚑ Exercise 2.7: Using Stemming
  - ⚑ Lemmatization

- ⚑ Exercise 2.8: Extracting the Base Word Using Lemmatization

Day 3 Schedule in Detail:

Morning Session (9:00 am – 12:00 pm)
- Word Sense Disambiguation
- Exercise 3.1: Word Sense Disambiguation
- Word Embeddings

Session Break: 12:00 pm to 1:30 pm

Afternoon Session (1:30 pm – 4:30 pm)
- Feature Extraction Methods
- Exercise 3.2: Feature Extraction Methods

Quiz Session (4:30 pm – 5:00 pm)

Validatory Session (5:00 pm – 5:30 pm)

Workshop Instructors:

Instructor Name	Designation	Affiliation
Dr. Pramod Kumar Singh	Professor	IIIT GWALIOR
Dr. Sriparna Saha	Associate Professor	IIT Patna
Dr. Asif Ekbal	Associate Professor	IIT Patna
Dr. Snehasis Banerjee	Scientist (AI)	TCS Research, Kolkata
Dr. Aparajita Dutta	Assistant Professor, Grade II	NIT SILCHAR
Dr. Jitesh Pradhan	Assistant Professor	NIT JAMSHEDPUR
Dr. Sanjay Chatterji	Assistant Professor	IIIT KALYANI

Registration:

- **Registration Fee for <mark>One Workshop</mark> for a Participant: INR 2950** (INR 2500/- plus 18% GST)

- *If one participant is enrolling for multiple workshops. Kindly multiply the Registration Fee with how many numbers of workshops you are willing to enrol for.*

Workshop Registration Link:

https://forms.gle/Z9YoU1CwMAWg88eF6

Eligibility:

This workshop welcomes participants with diverse backgrounds and interests, including but not limited to:

- Professionals (Academicians/Data Scientists/Data Analysts/Software Developers)
- Students (Diploma/UG/PG/Ph.D)

Note: *Basic Programming Knowledge in Python, required.*

Payment Gateway:

Please submit the ***Registration Fee*** through different payments options given below:

SBI Collect:

https://www.onlinesbi.sbi/sbicollect/icollecthome.htm?corpID=5150817

Bank Transfer:

Bank Name: State Bank of India
Beneficiary Name: IIT Patna Vishlesan I-Hub Foundation
Bank Account Number: 39579970437
IFSC Code: SBIN0017164

UPI:

UPI Id: 39579970437@sbi