

UTKARSH ANAND

Bachelor of Technology:
Computer Science and Engineering

Phone No.: +919770041309, +917007286432

Email id: utkarsh141352cse@gmail.com

Career Objective

To achieve excellence in working as dynamic professional offering solutions to the organization where my analytical ability and analyzing quest are used. Looking forward for a challenging environment, where I can refine my skill and expand my knowledge for the dynamic growth of the organization.

Educational Qualification

Secondary & Higher Secondary

10th Board (ICSE) Metropolitan School, Gorakhpur – 89.60%

12th Board (ISC) Metropolitan School, Gorakhpur – 89.25%

Under Graduation

Jaypee University of Engineering & Technology (M.P) – Current CGPA: 8.6

Skills & Capabilities

- **Programming Languages:** JAVA, C, C++.
- **Framework & DB:** Struts2.0 with Hibernate, MySQL.
- Text Processing & Natural Language Processing.
- **Operating System:** Windows and Linux (Ubuntu).
- Proficient in MSOffice(Word,PPT,Excel),LiveMoviemaker
Adobe Photoshop, Adobe LightRoom

Areas of Interest

- Object Oriented Programming.
- Data Structures & Algorithms
- Operating System
- Software Engineering

Professional Experience

Internship in Natural Language/Text Processing

June 2017 to July 2017

CSIR-CMERI, Durgapur (W.B)

Guide: - Dr. Partha Bhattacharjee (Sr. Principal Scientist-F)

Project: Extraction of Meaningful Insights Using Tokenization in Unigram Model of Natural Language Text

Summer Trainee in J2EE-Struts with Hibernate
Hewlett Packard (HP), Lucknow

June 2016 to July 2016

Summer Trainee in Core JAVA
Hewlett Packard (HP), Lucknow

June 2015 to July 2015

Mini Projects: Online Chat Messenger, Anti-Virus and Tic-Tac-Toe Game.

Publications:-

- 1) Natural Language Information Interpretation Representation System (NLIIRS) for Information Processing (*UGC LISTED JOURNAL*) presented at SCITSE-NIT RAIPUR.
- 2) Natural Language Information Interpretation Representation System: Machine Learning Based Approach (*UGC LISTED JOURNAL*) presented at RAC2B-NIT RAIPUR.

Projects:-

1) Wise Up (Internship) [Details attached in Annexure 1]

A system developed for-Extraction of Meaningful Insights Using Tokenization in Unigram Model of Natural Language Text using Natural Language Processing built using Stanford NER and was further extended.

2) Knowledge Extraction System (may be trainable) for Natural Language Data (Final Year) [Details attached in Annexure 1]

This system will analyze critical content, draw inference and come up with answers to legitimate questions related to the content. The trainability aspect of the system will be done after the system starts understanding and analyzing a particular use case of natural language information.

3) Jaypee Groups College and Branch Predictor

A module implemented in **Python** using Tkinter that ensures users college choice and predicts the Jaypee Branch and the trade offered in the college according to analysis from previous stats.

4) Codify

An application build in Python for Tech students who could easily access their notes and tutorials on the go which are best categorized under different labels.

Position of Responsibilities

- General Secretary at JYC (Jaypee Youth Club) 2017-present
- Curator and Production Head at TedXJUET '18
- Programming Club Co-ordinator at CSI (Computer Society of India) in 2015-2016
- CAMPUS AMBASSADOR for RCPL, Noida.
- Co-ordinator, Jaypee Youth Club (Public Relations Committee)

Academic Achievements

- Credit Beholder of **The University of New South Wales, Sydney (UNSW)** International Assessments for Indian Schools in Computer Skills, Mathematics and Science consecutively for two years i.e. 2010 & 2011 in the UP state.
- Runner-up at **FMC WEEKEND, IIT BHU (2015)**.

Core Abilities

- I always try to adapt new techniques and ways of achieving goals and to perform above the expectations with innovating techniques.
- I am good at expressing my thoughts and views and believe in beholding a strong convincing power.
- I am a good team worker as well as hold the ability to lead a team for a particular goal or motive and believe team effort is better than working individually.

Annexure I:

Collaborative Research Projects Undertaken/Ongoing:

→ At CSIR Labs-Central Mechanical Engineering Research Institute:

In this work not only have I made Extraction of Meaningful Insights using Natural Language/Text Processing but have made an efficient application which is hungry for the data and will increase its efficiency exponentially.

Also, this project achieves conversion of unstructured data to structured data.

The efficiency is achieved by feeding the knowledge base as much as a user can with new data plus giving an advantage of adding new entities along with the 7 class model of Stanford University which categorized Location, Person, Organization, Money, Percent, Date, Time i.e. our application enables the user for $(7C+NE*C)$ at run time where NE: -New Entities



The work done by the Stanford group has been deeply explored and studies in order to continue their work for extracting information (unigram model) efficiently and providing more meaningful insights.

The Great Depression was a severe worldwide economic depression in the decade preceding World War II. The timing of the Great Depression varied across nations, but in most countries it started in about 1929 and lasted until the late 1930s or early 1940s. It was the longest, most widespread, and deepest depression of the 20th century.

Potential tags:

- LOCATION
- TIME
- PERSON
- ORGANIZATION
- MONEY
- PERCENT
- DATE

Fig:-Stanford NER

Enhanced Work:-

Entities After Training:-

PERSON DATE LOCATION CLOTHING BREED ACTIVITY COLOR MONEY PERCENT

Set A

Albert Einstein is an American citizen since 1911 and wore red shirt. Adolf is a German having \$4000 which is 40% of German wealth since 1902.

Set B

Samuel Roger is the President of United States of America since 2008. Adolf is an American and a German respectively since 1900 and shared yellow shirt with Samuel Roger.

Set C

Samuel Roger is an American who wore red shirt. Adolf was the general of German army wearing yellow shirt.

The flow of the application developed goes as follows:-

- Unstructured text being fed in as the input by the user. The input needs to be grammatically correct according to the English language.
- The unstructured text is then annotated against a 7 Class classifier offered by the Stanford AI & NLP laboratories using `annotateText()`.
- The classifier then annotates the corpus and classifies into basic 7 entities as shown in the figure below and categories as “\O” to the ones it couldn’t decipher.
- The annotated corpus is then parsed using `parseText()`. Using the Pattern and Matcher class further helped us discriminating and indeed extracting the basic 7 Class insights.
- The string is then processed through a `textProcessing()` algorithm where I have provided the flexibility of increasing the efficiency of the application as it enables the user to add new entities belonging to a completely different class.
- For retrieval of information from the knowledge base so as to provide more information certain algorithms work as well: - `search_algo()`

Further the project was upgraded for receiving the corpus N number of times per user.

→ With National Institute of Technology (NIT) Durgapur:

I am associated in research work (final year project) assigned to me under the supervision of Mr. Partha Sarathy Banerjee, Faculty Department of CSE, JUET, Guna and Dr. Baisakhi Chakraborty, Asst Prof, Department of Computer Science & Engineering, NIT Durgapur. The work hopefully will be published in Taylor & Francis (Science Citation Index by Web of Science, Thompson Reuters) in the forth coming months. A brief work description is as follows:

The World Wide Web is an ever expanding source of data in today’s world. Millions and millions of tera-bytes of data and information is getting added every second. In this information age as the data is getting generated at an exponential rate, the fact to be noted is that most of the information is already available in the form of natural language text. Firstly the task of information extraction from mammoth data leads us to think on the quality and the form of available data. Secondly the ever increasing data poses a challenging task of extracting useful information from the available data. The third task is to develop ingenious way to answer any kind of query put up by a user from the available unstructured data. This paper proposes a trainable and integrated Natural Language Information Interpretation and Representation System (NLIIRS) that accepts any available un-annotated corpus of data in the form of natural language, and performs the following tasks: finds out the useful data, extracts relevant information in usable form (structured form / tables), and summarizes the data, structures the data in relational form. At the end the Question and Answering (Q&A) module shows

the cognitive abilities of our NLIIRS system by answering the questions in natural language relevant to the text. This is multispecialty system beyond just Question and Answering (Q&A). This is a trainable system capable of handling any unstructured data to be transformed into structured and well organized information. It also showcases the cognitive ability of the system to answer any questions raised in natural language with relatively better accuracy. It adopts the advantages of Named Entity Recognition so as to bypass the time consuming process of Parts of Speech Tagging (POS tagging) while pre-processing the available corpus (data) for information extraction.

References

1. **Dr. Shishir Kumar**
Professor & HOD
Department of Computer Science &
Engineering Jaypee University of Engineering
& Technology (M.P)
Email: shishir.kumar@juet.ac.in
2. **Mr. P.S Banerjee**
Assistant Professor
Department of Computer Science &
Engineering Jaypee University of Engineering
& Technology (M.P)
Email: partha.banerjee@juet.ac.in
3. **Dr. Baisakhi Chakraborty**
Assistant Professor,
Department of Computer Science &
Engineering, National Institute of Technology,
Durgapur, West Bengal-713209, India.
Email: baishakhi.chakraborty@it.nitdgp.ac.in