

Mihir Shekhar

Room no. 319
New Boys' Hostel,
IIIT-Hyderabad
Gachibowli, Hyderabad - 500032
TELANGANA, INDIA

Date of Birth	March 1 st , 1988
Personal E-mail	mihirshekharcse2010@gmail.com
Institute E-mail	mihir.shekhar@research.iiit.ac.in
Contact	+91 9581 826727

I am currently pursuing **MS by Research** in CSE under **Prof. Kamalakar Karlapalem**, Center for Data Engineering, IIIT Hyderabad in Data Mining.

Research Interest

Data Mining, Text Mining, Natural Language Processing.

Work Experience

Dec 2012-present	Research Assistant at Center for Data Engineering (CDE)
July 2013-Dec 2013	Teaching Assistant for Data Warehousing and Data Mining .
Sept 2010-July 2012	Software Developer in Tata Consultancy Services , Kolkata.

Research Projects

Medical Document Analysis (Aug 2014 to Current)	This project involves creating a statistical parser for medical documents and assign medical terms like drugs, symptoms, etc. to their corresponding disease. This project is funded by Hitachi R&D Labs.
Twitter Data Analysis (Dec 2012- Aug 2013)	This project work involves retrieval of semantic knowledge from tweets. Semantics of our interest are : Event/Episode Detection, Sentiment Analysis and Concept Extraction. This project is funded by Hitachi R&D Labs. http://cdeproject.iiit.ac.in/htap/
Web Content Filtering (Jan 2010-July 2010)	This project involves creation of an automatic system for categorization of web pages into different classes based on their content. A web content filter is built on top of it, to block undesired categories dynamically.

Course Projects

- **Wikipedia Search Engine:** Created a fully functional offline search engine. This search engine was implemented over Wikipedia Corpus of size 42GB. Multilevel indices were built on page title, infobox, text and outlinks to support queries for multiple fields. Everything was done from scratch without the use of any existing tools like Lucene, Lemur and wikixmlj parser.
Technology: JAVA.
- **StackOverFlow Tag Prediction:** Created a system to predict tags for StackOverFlow data. κ -nearest neighbor approach was used.
Technology: JAVA, Lucene
- **Finding Most Influential Entities in Web:** Created a system to find and rank most influential people among a group of **Baidu** users differentiating between fake users and genuine users. **Graphchi** was used in implementing algorithms for scalability and speed. It can process a billion entities in 15 minutes.
Technology: JAVA, C++, Graphchi.
- **Data-Mining on Accidents Dataset:** This project involved data pre-processing of a huge dataset, followed by data clustering using K-Means and DBSCAN algorithms and Frequent Item-Set Generation to analyse the trends in the occurrence of traffic casualties based on several conditions.
Technology: Weka, Java.

Publications

- Mihir Shekhar, Kamalakar Karlapalem, **HotClus: A Robust Heat based and Scalable Clustering Algorithm**. *2015 SIAM International Conference on Data Mining, Under Review.*
- K Santosh, Romil Bansal, Mihir Shekhar, Vasudeva Varma, **Author Profiling: Predicting Age and Gender from Blogs**, *Notebook for PAN at CLEF 2013, Valencia, Spain.*

Education

M.S. by Research 2012-present	International Institute of Information Technology Hyderabad Computer Science and Engineering CGPA: 8.11/10
Bachelor of Technology 2006-2010	Jalpaiguri Government Engineering College, West Bengal Computer Science and Engineering CGPA: 8.55/10
Higher Secondary 2003 - 2005	National Institute of Open Schooling, Patna Percentage: 72.8%
Secondary 2003	St. Dominic Savio's High School, Patna Central Board of Secondary Education, Delhi Percentage: 85.4%

Technical Skills

Operating Systems	: Ubuntu, Windows
Programming Languages	: Java, C, R
Scripting	: JavaScript, Bash
Programming Environments	: Eclipse
Web Technologies	: JSP, HTML, XML, CSS, AJAX, jQuery
Toolkits	: Lucene, Stanford NER and POS Tagger, Metamap
Database	: MySQL, MongoDB
Document Typesetting	: L ^A T _E X(T _E XLive), Microsoft Word
Data Visualisation	: D3.js

Relevant Course Work

- Data Warehousing and Data Mining
- Information Retrieval and Extraction
- Natural Language Processing Applications
- Web Mining

Achievements

- Obtained **Second position** in Author Profiling task at CLEF 2013.
- TCS ILP Top performer 2010.
- Participated in Microsoft S2B 2009. Created a 3D space shooter game based on XNA framework.
- Got selected in final round of The Great Mind Challenge, 2008 organised by IBM.

Hobbies

- Badminton
- Chess
- Yoga
- Listening Music

Declaration

I hereby declare that the above information is true to the best of my knowledge.