

HARSH V. THAKKAR

K-304, Kanak Kala-II, Opp. Seema Hall, Shyamal Cross Roads, Satellite, Ahmedabad, Gujarat, INDIA 380 015

+91-982-4547-890 • harsh9t@gmail.com • Homepage

EDUCATION

- **Dhirubhai Ambani Institute of Information and Communication Technology** Gandhinagar, Gujarat.
Doctoral student in BioNLP and Bioinformatics under the excellent guidance of Prof. Prasenjit Majumder(2013-onwards)
- **Sardar Vallabhbai National Institute of Technology** Surat, Gujarat.
Masters in Computer Engineering: Sentiment Analysis for Twitter supervised by Dr. Dhiren R. Patel (2011-2013) CGPA: **8.43/10** (Distinction)
- **L. D. College of Engineering** Ahmedabad, Gujarat.
Bachelor of Engineering in Computer Engineering (2007-2011) Percentage: **72%** (Distinction)
- **Shree Narayana Guru Vidyalaya** Ahmedabad, Gujarat.
Higher Secondary Education (2006-2007) Percentage: **80%** (Distinction)
- **Shree Narayana Guru Vidyalaya** Ahmedabad, Gujarat
Secondary Education (2005-2006) Percentage: **81%** (Distinction)

INTERNATIONAL PUBLICATIONS

- Harsh Thakkar, Ganesh Iyer, Kesha Shah and Prasenjit Majumder
CLEF/ShArE ehealth 2014 Task 3: User centered health information retrieval system accepted in CLEF 2014 eHealth Task Track, Sept 2014.
- Harsh Thakkar and Dhiren Patel
Approaches for Sentiment Analysis on Twitter: A state-of-art study **accepted** at the International Network for Social Network Analysis conference (INSNA), Xi'an, China, July 2013.
- Jitendra B. Gurubani, Harsh Thakkar and Dhiren R. Patel
Improvements over Extended LMAP+: RFID Authentication Protocol published in TRUST MANAGEMENT VI - IFIP Advances in Information and Communication Technology, Vol. 374, pp. 225-231, Surat, India, May 2012.
- Harsh Thakkar
NeurOCR : A Neural Network based approach to Optical Character Recognition systems published in proceedings of The 14th International Conference of International Academy of Physical Sciences, Surat, India, Dec 2011.

WORKING PAPERS

- Harsh Thakkar and Prasenjit Majumder
Advances and opportunities in text mining on biological documents - A state-of-art study with focus on analysing a variety of approaches in biomedical text mining.
- Harsh Thakkar and Prasenjit Majumder
Chapter 2: Essential mathematics for biologists- Chapter in a edited volume titled "Book title Trends in BioNLP", [2014-onwards]

QUALIFYING TEST SCORES

- **GRE** (General): 303 (Q-155,V-148) |**AWA**- 3.5 |Date: Sept 18, 2012
- **TOEFL** (IBT): 104 (Reading- 27, Writing- 30, Listening- 23, Speaking- 24) |Date: Sept 23, 2012
- **GATE 2011** (CS): Percentile- 92.30 |Score- 404 |Date- March 15, 2011

ACHIEVEMENTS

- **Awarded AA Grade (10/10)** for M.Tech. thesis carrying research work on Sentiment Analysis for Twitter using Hybrid Naive Bayes by Computer Department, Sardar Vallabhbai National Institute of Technology (2013)
- **Microsoft Certified Professional** Awarded by Microsoft on successfully clearing certification exam 70-536 of .NET Framework 2.0 Application Foundation(2010) [Official MS Transcript]
- **Microsoft Student Partner** Alumni Student Partner (2009-'11 batch) of Microsoft to Lalbhai Dalpatbhai College of Engineering(LDCE), Gujarat Region, INDIA.

COURSES ATTENDED

- **Forum for Information Retrieval Evaluation 2013 [FIRE]** Attended International Workshop of FIRE at India International Center, New Delhi, INDIA [official Website]. Learned key concepts on Information Retrieval process and relevant mechanisms(Dec 2013)
- **International Winter School on Trust Management** Attended IFIP WG11.11 Winter School on Trust Management in ICT at NIT Surat, INDIA [Website]. Learned key concepts on Trust Models and Reputation management Systems(Feb 2012)

SELECTED PROJECTS

- **Natural Language Processing for biomedical texts [Ph.D. Thesis](Sept 2013-present)**

Till the earlier decade biomedical text analysis was carried out by biology subject experts, which is time consuming and missing certain essential facts hidden within the text. From the late 20th century the term BioNLP was coined, describing possible ways of bridging the biological domain with Natural Language Processing techniques. Biomedical text mining (also known as BioNLP) refers to text mining applied to texts and literature of the biomedical and molecular biology domain. It is a recent trend in research field involving text-mining for biology toward fine-grained information extraction (IE) on the edge of natural language processing, bioinformatics, medical informatics and computational linguistics. Improved NLP methods are used to gather, store, analyse and integrate biological and genetic information which can then be applied to gene-based drug discovery. Biological text contains enormous valuable information on various hidden patterns regarding biological entities like bacteria, enzymes, biotopes, genomes and etc., interactions with each other. During this research work i dived through a broad study of applying variety of NLP techniques for extraction of knowledge from biomedical texts, i.e. biomedical articles and research papers, journals, public datasets [PubMed], etc. I am currently working the same.
- **Sentiment Analysis for Twitter using Hybrid Naive Bayes (2013)**

Carried out research work as a part of M.Tech curriculum, proposing a new hybrid approach for exploring public opinion by mining sentiments of users in twitter. Used a combination of Machine learning (*Naive Bayes*) algorithm and a special Lexical dictionary (*SentiWordNet*) and achieved competitive result (90 % accuracy) on crawled twitter datasets.
- **Light Weight Protocol for Mutual authentication for RFID Systems (2012)**

Worked with Jitendra Gurubani at Sardar Vallabhbhai National Institute of Technology, Surat in M.Tech Semester I. The goal of this Project is to study various protocols existing in the current RFID systems. Later to propose a solution to the problem based on designing a Light weight protocol for mutually authenticating RFID tags
- **NeurOCR (2011)**

Worked with Tanmay Desai at L. D. College of Engineering in Semester VIII, for Final Project of Bachelor of Engineering in Computer Engineering. The goal of the project was to develop a Software for recognizing human written character and matching it with a known english alphabet. The platform used for developing this project is .Net Framework and Heaton research API for Neural Network calibration.
- **Imassaur (2010)**

Worked with Aanshi Shah at L. D. College of Engineering in Semester VI, developing an Image processing software. The goal of this project was to study and understand the working of various algorithms (Canny, Sobel, etc) on low-level and shape based image processing and to provide a user-friendly Graphical Interface for the same. The platform used for developing this project is .Net Framework and GDI+ graphics library.
- **Baseball Dossier (2009)**

Worked with Yash Raithatha at L. D. College of Engineering in Semester IV, developing a basic user friendly Baseball Score Spreadsheet in JAVA SWING. The goal of this project was to develop an advanced spreadsheet targeting the managing council of the baseball sport for easy, reliable and rapid access and storage of information of numerous teams playing matches in a tournament or worldcup. The platform used for developing this project is Java.

EXTRA CURRICULAR ACTIVITIES

- **Treasurer, ACM Student Chapter, DA-IICT, Gandhinagar (2014-onwards)**

I currently function as the treasurer of the ACM DA-IICT. This student chapter has be established with a vision to promote research to the younger talent and bring awareness in the undergraduate and graduate students.
- **Auditor, Hostel-10, SVNIT, Surat (2011-13)**

Served the dignitary post of Auditor for succeeding two years in Hostel-10 of Sardar Vallabhbhai National Institute of Technology. Carried out final budget audits and various finance planning tasks. Learned various strategic skills for managing fund flow (in/out) in an organization during this span of my service.
- **Content Creator, Engineering Buddy (2012)**

Engineering Buddy is Open Social Networking Innovation Platform For Engineer Graduates and Under-Graduates. Here we reciprocate innovative ideas, mostly refreshingly creative and dominantly humorous.

- **Organiser, DreamSpark Yatra, Microsoft Student Patners, Gujarat (2010, 2011)**
Successfully organised Dream Spark Yatra (DSY), an initiative aimed to create a local ecosystem of technology enthusiasts, for consecutive two years in a row. Earned opportunity to meet the Tech. heads and Professionals form Microsoft India, organising and attending technical seminar across colleges in Gujarat.

HOBBIES

I indulge and enjoy

- **Writing Poems** [Blog here]
- **Cooking**
- **Reading Fiction**
- **Singing**

PERSONAL INFORMATION

- Nationality : **Indian**
- Birth Date : **9 March 1990**
- Languages known: **English, Hindi, Gujarati, Spanish** (Amateur)

REFERNCES

- **Dr. Prasenjit Majumder**
Assistant Professor,
DAIICT, Gandhinagar - 382007,
Room No: 4209,
Email: p_majumder@daiict.ac.in
Phone: (+91) 079-3051-0605

- **Dr. Dhiren R. Patel**
Professor,
Department of Computer Engineering,
Sardar Vallabhbhai National Institute of Technology,
Gujarat, Surat - 395007, India
Email: drp@coed.svnit.ac.in
Phone: (+91) 0261-2201-638

- **Dr. Audun Jøsang**
Professor,
Department of Informatics,
University of Oslo,
PO Box 1080 Blindern,
0316 Oslo, Norway
Email: josang@mn.uio.no
Phone: +47-98431433