

EDUCATIONAL QUALIFICATION

Ph. D.	Pursuing	IITB – Monash Research Academy	8.08
B. Tech. (CSE)	2013	Gautam Buddha Technical University, Lucknow	72.90%
Higher Secondary Examination	2005	Greenfields Public School, Dilshad Garden, Delhi (CBSE)	61.40%
Secondary Examination	2003	St. Josephs Academy, Savita Vihar, Delhi (CBSE)	75.00%

▪ Ranked among the **top three** students of undergraduate college, in all the academic years during undergraduate studies.

AWARDS

- **Awarded “Teaching Assistant of the Semester”** by Computer Science and Engineering Department for the course Software Systems Lab (CS699) working with Prof. Kavi Arya (2019).
- **Awarded “Teaching Assistant of the Semester”** by Computer Science and Engineering Department for the course Software Systems Lab (CS251) working with Prof. Kavi Arya (2017).
- **Awarded the CISCO Fellowship** among all applicants from the CSE Department in 2016.
- **Best Poster (Early Researcher Category)**, IITB-Monash Research Academy Annual Oskar Awards 2016.
- **Winner of ICON SMT Competition (2014)**, our work on Hierarchical Phrase based Statistical Machine Translation outperformed all other systems, thus making us winners of the competition.

MAIN CONFERENCE / JOURNAL PUBLICATIONS

1. Diptesh Kanojia, Pushpak Bhattacharyya, Malhar Kulkarni, Reza Haffari, **Challenge Datasets of Cognate and False Friend Pairs for Indian Languages**, Language Resources and Evaluation Conference (LREC 2020), May 2020.
2. Akash Sheoran, Diptesh Kanojia, Aditya Joshi, Pushpak Bhattacharyya, **Recommendation Chart of Domains for Cross-Domain Sentiment Analysis: Findings of A 20 Domain Study**, Language Resources and Evaluation Conference (LREC 2020), May 2020.
3. Diptesh Kanojia, Abhijeet Dubey, Malhar Kulkarni, Pushpak Bhattacharyya, Reza Haffari, **Utilizing Word Embeddings based Features for Phylogenetic Tree Generation of Sanskrit Texts**, 6th International Sanskrit Computational Linguistics Symposium (ISCLS 2019), October 2019.
4. Diptesh Kanojia, Malhar Kulkarni, Pushpak Bhattacharyya, Sayali Ghodekar, Irawati Kulkarni, Nilesh Joshi, Eivind Kahrs, **An Introduction to the Textual History Tool**, 6th International Sanskrit Computational Linguistics Symposium (ISCLS 2019), October 2019.
5. Diptesh Kanojia, Kevin Patel, Pushpak Bhattacharyya, Malhar Kulkarni, Reza Haffari, **Utilizing Wordnets for Cognate Detection among Indian Languages**, 10th Global WordNet Conference (GWC 2019), July 2019.
6. Sandeep Mathias, Diptesh Kanojia, Kevin Patel, Samarth Agarwal, Abhijit Mishra, Pushpak Bhattacharyya, **Eyes are the Windows to the Soul: Predicting the Rating of Text Quality Using Gaze Behaviour**, The 56th Annual Meeting of The Association for Computational Linguistics (ACL 2018), July 2018.
7. Diptesh Kanojia, Preethi Jyothi, Pushpak Bhattacharyya, **Synthesizing Audio for Hindi Wordnet**, Global WordNet Conference (GWC 2018), January 2018.
8. Diptesh Kanojia, Kevin Patel, Pushpak Bhattacharyya, **Indian Language Wordnets and their Linkages with Princeton WordNet**, Language Resources and Evaluation Conference (LREC 2018), May 2018.
9. Jayashree Gajjam, Diptesh Kanojia, Malhar Kulkarni, **New Vistas to study Bhartrhari: Cognitive NLP**, World Sanskrit Conference (WSC 2018), July 2018.
10. Ritesh Panjwani, Diptesh Kanojia, and Pushpak Bhattacharyya, **pyiwn: A Python-based API to access Indian Language WordNets**, 9th Global WordNet Conference (GWC 2018), January 2018.
11. Hanumant Redkar, Rajita Shukla, Sandhya Singh, Jaya Saraswati, Lakshmi Kashyap, Diptesh Kanojia, Preethi Jyothi, Malhar Kulkarni and Pushpak Bhattacharyya, **Hindi Wordnet for Language Teaching: Experiences and Lessons Learnt**, 9th Global WordNet Conference (GWC 2018), January 2018.
12. Diptesh Kanojia, Nikhil Wani, Pushpak Bhattacharyya, **Is your Statement Purposeless? Predicting Computer Science Graduation Admission Acceptance based on Statement of Purpose**, 14th International Conference on Natural Language Processing (ICON 2017), December 2017.
13. Abhijit Mishra, Diptesh Kanojia, Kuntal Dey, Seema Nagar and Pushpak Bhattacharyya, **Scanpath Complexity: Modeling Reading Effort using Gaze Information**, Thirty-first Association for the Advancement of Artificial Intelligence Conference (AAAI 2017), San Francisco, California, USA, February 2017.
14. Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya, **Leveraging Cognitive Features for Sentiment Analysis**, The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016), Berlin, Germany, August 2016.

15. Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya, **Harnessing Cognitive Features for Sarcasm Detection**, Association for Computational Linguistics (ACL 2016), Berlin, Germany, August 2016.
16. Diptesh Kanojia, Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman, **That'll do fine!: A coarse lexical resource for English-Hindi Machine Translation, using polylingual topic models**, Language Resource and Evaluation Conference (LREC 2016), Portorož, Slovenia, May 2016.
17. Shehzaad Dhuliawala, Diptesh Kanojia, and Pushpak Bhattacharyya, **SlangNet: A WordNet like resource for English Slang**, Language Resource and Evaluation Conference (LREC 2016), Portorož, Slovenia, May 2016.
18. Abhijit Mishra, Diptesh Kanojia, and Pushpak Bhattacharyya, **Predicting Readers' Sarcasm Understandability by Modelling Gaze Behaviour**, Thirtieth Association for the Advancement of Artificial Intelligence Conference (AAAI 2016), Phoenix, Arizona, U.S.A., February 2016.
19. Diptesh Kanojia, Shehzaad Dhuliawala, and Pushpak Bhattacharyya, **A picture is worth a thousand words: Using OpenClipArt library for enriching IndoWordNet**, Global WordNet Conference (GWC 2016), Bucharest, Romania, January 2016.
20. Diptesh Kanojia, Raj Dabre, and Pushpak Bhattacharyya, **Sophisticated Lexical Databases - Simplified Usage: Mobile Applications and Browser Plugins for Wordnets**, Global WordNet Conference (GWC 2016), Bucharest, Romania, January 2016.
21. Meghna Singh, Rajita Shukla, Jaya Jha, Laxmi Kashyap, Diptesh Kanojia, and Pushpak Bhattacharyya, **Mapping it differently: A solution to the linking challenges**, Global WordNet Conference (GWC 2016), Bucharest, Romania, January 2016.
22. Diptesh Kanojia, Shehzaad Dhuliawala, Abhijit Mishra, Naman Gupta and Pushpak Bhattacharyya, **TransChat: Cross-Lingual Instant Messaging for Indian Languages**, International Conference on Natural Language Processing (ICON 2015), Goa, India, December 2015.
23. Diptesh Kanojia, Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman, **Using multilingual topic models for improved alignment in English-Hindi MT**, International Conference on Natural Language Processing (ICON 2015), Goa, India, December 2015.
24. Diptesh Kanojia, Manish Shrivastava, Raj Dabre, & Pushpak Bhattacharyya, **PaCMan: Parallel Corpora Management Workbench**, International Conference on Natural Language Processing (ICON 2014), Goa, India, December 2014.
25. Diptesh Kanojia, Pushpak Bhattacharyya, Raj Dabre, Siddhartha Gunti & Manish Shrivastava, **Do not do processing, when you can look up: Towards a Discrimination Net for WSD**, Global Wordnet Conference (GWC 2014), Tartu, Estonia, January 2014.
26. Salil Joshi, Diptesh Kanojia, and Pushpak Bhattacharyya, **More than meets the eye: Study of Human Cognition in Sense Annotation**, North American Chapter of Association of Computational Linguistics (NAACL 2013), Atlanta, U.S.A., June 2013.
27. Arindam Chatterjee, Salil Joshi, Diptesh Kanojia, Akhlesh Meena and Pushpak Bhattacharyya, **A Study of the Sense Annotation Process: Man v/s Machine**, Global WordNet Conference (GWC 2011), Matsue, Japan, January, 2012.

CONFERENCE DEMONSTRATION TRACK PUBLICATIONS

28. Aditya Joshi, Diptesh Kanojia, Pushpak Bhattacharyya, Mark Carman, **Sarcasm Suite: A browser-based engine for sarcasm detection and generation**, Thirty-first Association for the Advancement of Artificial Intelligence Conference: Demonstrations Track (AAAI 2017), San Francisco, California, USA, February 2017.
29. Hanumant Redkar, Sudha Bhingardive, Diptesh Kanojia & Pushpak Bhattacharyya, **World WordNet Database Structure: An Efficient Schema for Storing Information of WordNets of the World**, Association for the Advancement of Artificial Intelligence (AAAI 2015), Austin, Texas, U.S.A., January 2015.
30. Diptesh Kanojia, Arindam Chatterjee, Salil Joshi, and Pushpak Bhattacharyya, **Discrimination Net for Hindi**, International Conference for Computational Linguistics (COLING 2012), Mumbai, India, December 2012.

CONFERENCE WOKRSHOP & COMPETITION PUBLICATIONS

31. Kumar Saurav, Kumar Saunack, Diptesh Kanojia, Pushpak Bhattacharyya, **"A Passage to India": Pre-trained Word Embeddings for Indian Languages**, Joint Spoken Language Technologies for Under-resourced languages and Collaboration and Computing for Under-Resourced Languages (LREC 2020 Workshop), May 2020.
32. Diptesh Kanojia, Sravan Munukutla, Sayali Ghodekar, Pushpak Bhattacharyya, Malhar Kulkarni, **"Keep Your Dimensions on a Leash": True Cognate Detection using Siamese Deep Neural Networks**, ACM India Joint International Conference on Data Science and Management of Data (YRS @ CoDS-COMAD 2020), January 2020.
33. Yashasvi Mantha, Diptesh Kanojia, Pushpak Bhattacharyya, Malhar Kulkarni, **Harnessing Deep Cross-lingual Word Embeddings to Infer Accurate Phylogenetic Trees**, ACM India Joint International Conference on Data Science and Management of Data (YRS @ CoDS-COMAD 2020), January 2020.

Diptesh Kanojia

<https://dipteshkanojia.github.io>

Email: dipteshkanojia@gmail.com, diptesh@cse.iitb.ac.in, diptesh.kanojia@monash.edu

Contact: +91 8879058668



34. Diptesh Kanojia, Malhar Kulkarni, Sayali Ghodekar, Eivind Kahrs, Pushpak Bhattacharyya, **Strategies of Effective Digitization of Commentaries and Sub-commentaries: Towards the Construction of Textual History, Tradition of Commentaries and the Dynamics of Knowledge (TCDK @ SSSU 2020)**, February 2020.
35. Diptesh Kanojia, Malhar Kulkarni, Pushpak Bhattacharyya, Reza Haffari, **Cognate Identification to improve Phylogenetic trees for Indian Languages**, ACM India Joint International Conference on Data Science and Management of Data (YRS @ CoDS-COMAD 2019), January 2019.
36. Kevin Patel, Diptesh Kanojia, and Pushpak Bhattacharyya, **Semi-automatic WordNet Linking using Word Embeddings**, Workshop on Word Embeddings at Global WordNet Conference (GWC 2018), January 2018.
37. Diptesh Kanojia, Vishwajeet Kumar, Krithi Ramamritham, **Civique: Using Social Media to Detect Urban Emergencies**, Workshop on Social Data Analytics and Management (SoDAM 2016), at VLDB 2016, New Delhi, India, September 2016.
38. Neha Prabhugaonkar, Apurv Nagvenkar, Diptesh Kanojia, Jyoti Pawar, Pushpak Bhattacharyya, & Manish Shrivastava. **PanchBhoota: Hierarchical Phrase Based Machine Translation Systems for Five Indian Languages**, International Conference on Natural Language Processing (ICON 2014), Goa, India, Dec. 2014.

COURSEWORK

Ph. D. Courses (2016 – 2017)	<ul style="list-style-type: none">Linear Optimization (CS750), Kernel Programming (CS746), Information and Communication Technology for Socio-Economic Development (CS671), Communication Skills (CS792 + HS791), Artificial Intelligence (CS621), Software Lab (CS699), Embedded Systems (CS684), Automatic Speech Recognition (CS753), Software Architecture (CS718), Seminar on "Computational Phylogenetics" (CS801)
---	--

WORK EXPERIENCE

Research Engineer, CFIL, IITB (2013 – 2015)	<ul style="list-style-type: none">Previously employed as a Research Engineer at Centre for Indian Languages Technology, IIT Bombay.Cognitive studies for Sentiment detection, Sarcasm Understandability, and Detection, through Gaze Tracking.Multilingual Topic Modelling to improve Indian Language Machine Translation.Working on maintenance Hindi, Marathi, Sanskrit WordNets, and IndoWordNet.PanchaBhoota - Hierarchical Phrase based Statistical Machine Translation system for five language pairs.Mobile App and Browser Plugin development for WordNets on Android, iOS, and Windows Phones.Core Systems Administrator for Centre for Indian Language Technology handling / troubleshooting / maintaining both server systems, and desktop computers.
--	--

POSITIONS OF RESPONSIBILITY

Peer Reviewer	Various Conferences	July 2017 – Present
Conferences / Workshop / Journals	<ul style="list-style-type: none">Reviewed research papers submitted at ACL / GWC / EMNLP / CoNLL / AACL / COLINGReviewed research papers submitted at GLOBALEX Workshop at LREC 2018Reviewed papers submitted in the Linguistic Issues in Language Technology Journal (WordNet Issue)	
Teaching Assistant & Mentorship	IIT Bombay	Jan 2017 – Present
CS684 CS699 CS251 CS213 CS226	<ul style="list-style-type: none">Teaching Assistant for the course "Software Lab (PG)" run by Prof. Kavi Arya.<ul style="list-style-type: none">Awarded TA of the semester for CS699 – Software LabTeaching Assistant for the course "Software Systems Lab" run by Prof. Kavi Arya.<ul style="list-style-type: none">Awarded TA of the semester for CS251 – Software Systems LabTeaching Assistant for the course "Software Architecture" run by Prof. R. K. Joshi.Teaching Assistant for the course "Data Structures" conducted by Prof. R.K. Shyamsundar.Teaching Assistant for the course "Digital Logic Design" conducted by Prof. Supratik Chakraborty.	
Mentor, SoS	<ul style="list-style-type: none">Mentored a group of undergraduate students for "Summer of Science (SoS)" organized by Maths and Physics Club, IIT Bombay for projects based on Machine learning.	
Web Nominee & Design Lead	PG Academic Council, IIT Bombay	Mar 2017 – Mar 2018
Head, Design Team & Doctoral Council	<ul style="list-style-type: none">Worked with the Institute Doctoral Council as an internal group lead and conceptualized the PhD Qualifier Document, also lead the design team for various activities.Discussed the issue of tagging for PhD Students and proposed an alternative scheme for PG Students where Best X out of Y courses can be used to calculate CPI for IITB-Monash Research Academy Students.Mentored the website work for PG Academic Council.Created an online interface for student grievances to be later clubbed with a central portal.Proposed an exit degree for IITB-Monash Research Academy students.Proposed more Faculty Advisors for PhD Students (X faculty on every Y students).Direct Transfer of CISCO Fellowship to student accounts instead of the release of a cheque.	

Research Scholars Representative		CSE Department, IIT Bombay	July 2016 – July 2017
Research Scholars Representative	<ul style="list-style-type: none"> Representing research scholars in the Departmental Post-Graduate Committee (DPGC) meetings. Organizing various meetings between potential employers, and interested research scholars. Helping organize intra-departmental meets between faculty-RS, and among RS, if need be. Organizing monthly 'Chai-meet' for RS, where they interact and share ideas across their research areas. Organizing RISC as a core-group member. 		

Academy Student Leadership Team		IITB-Monash Research Academy	Nov 2016 – Present
Member, Student Leadership Team	<ul style="list-style-type: none"> Organization of the "Pizza pe charcha" event for discussion of research happening among the IITB-Monash Research Academy students. Initiating an internal placement cell for the IITB-Monash Research Academy graduates. Discussion on various issues which are commonly faced by the Academy students. 		

PREVIOUS RESEARCH WORK & PROJECTS

"WordNet and its Applications"		Summer Internship CSE Department, IIT Bombay	Guide: Prof. Pushpak Bhattacharyya July 2011 – Aug 2011
WordNet	<ul style="list-style-type: none"> WordNet is a rich repository of knowledge of words and plays a key role in many Information Retrieval and Language Processing applications. Extensive study of WordNet (English and Hindi) and various state of the art algorithms. 		
IWSD Algorithm	<ul style="list-style-type: none"> Work based on Error Analysis of the IWSD algorithm, divided in various parts, viz. POS based, Ontology based, Ablation test on IWSD parameters. Aim being the importance of contextual part of the scoring function of IWSD algorithm. 		

"Man vs. Machine Annotation for WSD"		Research Project CSE Department, IIT Bombay	Guide: Prof. Pushpak Bhattacharyya Jan 2012 – Mar 2012
Study of Man / Machine Annotation Process	<ul style="list-style-type: none"> Our work aimed at understanding the meaning and significance of "contextual evidence", with respect to the human sense disambiguation task. The experiments probed into the disambiguation time across different POS categories and ontological categories, they also investigated into the set of clues in the neighbouring words of a pre-defined target word. 		

"Mechanisms for Sense Discrimination"		Minor Research Project CSE Department, IIT Bombay	Guide: Prof. Pushpak Bhattacharyya Jun 2012 – Aug 2012
Development of "Sense Discrimination Tool"	<ul style="list-style-type: none"> Emphasized on importance of context and clue words around a target word synset word. Developed "Sense Discrimination Tool" which facilitates the addition of clue words by lexicographers manually into the WordNet database. Laid groundwork for automation of the tool using clues present in the database, using various heuristics. 		

"Mechanisms for Sense Discrimination"		Major Research Project CSE Department, IIT Bombay	Guide: Prof. Pushpak Bhattacharyya Jan 2013 – Mar 2013
Improvements in "Sense Discrimination Tool" and its Future Implications	<ul style="list-style-type: none"> Improved the context word searching scenario by providing the lexicographers with a rich corpus. Facilitated the clue addition to database by including Google Hindi Transliteration API within the tool, and by including Concordancer Corpus search. Automated the clue searching by providing with an automatically generated possible clue set. Work on automatic clue addition in future using PMI and G² indexing of the clues. Laid groundwork for a possible future rule based framework to be called, "Discrimination Net". 		

"College Website in HTML5 and JS"		Double Major Project CSE Department, KSJIT	Guide: Mrs. Aishwarya Jain Jan 2013 – Mar 2013
Developed the College Website in HTML5 & CSS3	<ul style="list-style-type: none"> Completed College Website development. Development of website in HTML5 and CSS3. Made use of JavaScript and jQuery plugins for many features. 		

PROGRAMMING

C / C++	<ul style="list-style-type: none"> Advanced knowledge of C / C++ programming language, successful implementation of "Data Structures" and basic algorithms done during academic courses.
JAVA and Python	<ul style="list-style-type: none"> Knowledge of JAVA and Python, Implementation of various algorithms for NLP and data manipulation during research work at CFIL.
PHP, HTML, CSS, MySQL, JS	<ul style="list-style-type: none"> Proficient website designer, using the aforementioned languages and methods. Designed the complete College Website, in a short span of time. Tweaking and Maintenance of web-based interfaces in CFIL Lab, IIT Bombay. Database Normalization for WordNets, CFIL, IIT Bombay.

Diptesh Kanojia

<https://dipteshkanojia.github.io>

Email: dipteshkanojia@gmail.com, diptesh@cse.iitb.ac.in, diptesh.kanojia@monash.edu

Contact: +91 8879058668



IITB-Monash
Research Academy



EXTRA CURRICULAR ACTIVITIES

Tech-Geek	▪ Technology enthusiast. Like being up-to-date with the latest technology and gadgets around. Helps me make conscious decision on what technology / gadget is worth investing into.
Sports	▪ Played sports like Basketball, Table Tennis at school, and college level.
Reading	▪ English / Hindi Fiction & Non-fiction, novels and short stories , both in classical English Literature and in Modern day Literature .
Photography	▪ Zealous about digital photography , maintain an anthology of photography done by myself, using both Advanced Point-and-Shoot & DSLR cameras.
Singing	▪ Winner of first prize at a school level singing competition .
Debating & Quizzes	▪ Participant in various school level debating competitions, Winner of Intra College quiz competition , and many other quizzes at school level.