Dr. Sandeep Albert Mathias

Ph.D., IIT Bombay, Mumbai

✉ sam@cse.iitb.ac.in | mathiassandeep@gmail.com
☎ +91-8433780647
lwsam.github.io | www.cse.iitb.ac.in/~sam
lwsam | Sandeep Mathias | Sandeep Mathias
Sandeep Mathias | Sandeep-Mathias

Research Interests

My primary area of interest is the field of Natural Language Processing (NLP). Currently, I am working on problems combining gaze behaviour with NLP, in particular, in the domain of education, like automatic essay grading, text simplification and automatic short-answer scoring.

Education

Qualification	Board/University	Month & Year	CPI (out of) / %
Ph.D. (Computer Science)	IIT Bombay, Mumbai	July, 2020	8.00 (10)
M.Tech. (Information Technology)	IIIT Bangalore, Bengaluru	July, 2012	3.22 (4)
B.Tech. (Information Technology)	NITK, Surathkal	June, 2010	7.45 (10)
Higher Secondary Examination (10+2)	Karnataka PUC Board St. Joseph's Pre-University College	April, 2006	82.33%
Secondary Examination (10)	Indian Certificate of Secondary Education St. Joseph's Boys' High School	April, 2004	86.00%

Work Experience/Internship

- **Research Intern**, Kaybus India Pvt. Ltd., Bengaluru *May 2014 to August 2014*
  Responsibilities: During the internship, I worked on automatically generating questions from text, in particular user manuals.
- **Software Engineer**, Nokia India Private Ltd., Bengaluru *July 2012 to June 2013*
  Responsibilities: During my time in Nokia, I worked on user authentication and sign-in for the Nokia account.

Professional Service

- Program Committee member for:
  1. Language Resources and Evaluation Conference (LREC) *LREC 2020*
  2. Innovative Use of NLP for Building Educational Applications Workshop (BEA) *BEA 2020*
  3. AAAI Conference on Artificial Intelligence (AAAI) *AAAI 2021*
- Teaching Assistant for the following courses:
  1. Topics in Natural Language Processing @ IIT Bombay *Spring 2015*
  2. Speech, Natural Language Processing and the Web @ IIT Bombay *Fall 2014*
  3. Automata Theory and Logic @ IIT Bombay *Spring 2014*
  4. Functional Programming @ IIT Bombay *Fall 2013*
  5. Mathematical Models of Computation @ IIIT Bangalore *Spring 2012*

Thesis and Projects[1]

- **Ph.D. Thesis: Cognitively Aided Automatic Essay Grading** (*Defended: July 24, 2020*)
  Under the guidance of Prof. Pushpak Bhattacharyya.
  Description: An essay is a piece of text, written in response to a topic, called a prompt. In the thesis, cognitive information, which was recorded / learnt using a reader's gaze behaviour, was used to score the essays. As part of the thesis, we first showed that gaze behaviour can be used for predicting how a reader would score a text qualitatively. Next, we showed how we can learn the gaze behaviour of a reader to score the text. In addition to the above works, we also investigated ways of scoring essay traits - different aspects of the essay, like content, organization, style, etc. - and using the trait scores to score the essay.

[1]Only important ones listed here.

- **Automatic Text Simplification** *July 2014 to May 2016*
  Description: In this project, I worked on simplifying text by splitting sentences and substituting words and phrases. One of the outcomes of the project was coming up with a measure of text complexity, called the E-Score.

- **Automatic Question Generation from Text** *May 2014 to August 2014*
  Description: This project was done as my internship with Kaybus. As part of the project, I was to create and extract questions from texts. This was done by first splitting the sentences using a number of rules. Next, important question words (like names, places, dates, etc.) would be identified for each sub-sentence, and appropriate questions would be generated.

- **M.Tech Thesis: Automatic Classification of Folktales** *January 2012 to June 2012*
  Under the guidance of Prof. V.N. Muralidhara
  Description: The aim of the thesis was, given an input text and a set of classes, we have to find out which class the text belongs to. Tagging was done using Stanford Core NLP, after which I found out which was the most similar class that a text belonged to, using similarity measures. The results were compared using the Aarne-Thompson Tale Type Index for classifying folktales.

- **Academic Data Analytics** *January 2011 to April 2011*
  Description: Implementation of data mining algorithms to extract potentially useful information from a data set. The algorithms implemented were k-means clustering (for clustering), Dynamic Itemset Counting and Frequency Pattern Tree (for Association Rule Mining) and ID3 (for Classification). I was responsible for implementing the k-means clustering, as well as making an offline version using Java.

- **Interactive Ray Tracing Using Photon Mapping** *January 2011 to April 2011*
  Description: Photon mapping is a widely used algorithm for global illumination, where one traces photons through the scene, recording their interaction with elements in the scene in a data structure called the photon map. We implemented a GPU-based ray tracer that uses the photon map to simulate full global illumination directly on the graphics hardware, with progressive, interactive feedback to the user. I was a developer, responsible for implementing the parsers necessary for reading the scene files that described the initial scenes.

- **Column Data Store in C** *August 2010 to November 2010*
  Description: We had to implement a simple data base application using C as a programming language. We implemented simple database operations, such as insertion, deletion and searching of records, based on certain parameters. The biggest issue we faced was in generalizing our application to handle any type of database query. In this project, I was responsible for implementing insertion and unloading functions, as well as reducing errors.

Publications

1. **Sandeep Mathias**, Diptesh Kanojia, Abhijit Mishra, and Pushpak Bhattacharyya. A Survey on Using Gaze Behaviour for Natural Language Processing. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, pages 4907 – 4913. **IJCAI 2020**. Kyoto, Japan / Online (To appear in January, 2021).

2. **Sandeep Mathias**, and Pushpak Bhattacharyya. Can Neural Networks Automatically Score Essay Traits? In *Proceedings of the 15th Workshop on Innovatove Use of NLP for Building Educational Applications*, pages 85 – 91. Seatle, WA, USA (Online). July 10, 2020.

3. **Sandeep Mathias**, and Pushpak Bhattacharyya. Thank "Goodness"! A Way to Measure Style in Student Essays. In *Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications*, pages 35 – 41. **NLP TEA 2018**. Melbourne, Australia. July 19, 2018.

4. **Sandeep Mathias**, Diptesh Kanojia, Kevin Patel, Samarth Agrawal, Abhijit Mishra, and Pushpak Bhattacharyya. Eyes are the Windows to the Soul: Predicting the Rating of Text Quality Using Gaze Behaviour. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2352 – 2362. **ACL 2018**. Melbourne, Australia. July 16 – 18, 2018.

5. Nikhil Wani, **Sandeep Mathias**, Jayashree Aanand Gajjam, and Pushpak Bhattacharyya. The Whole is Greater than the Sum of its Parts: Towards the Effectiveness of Voting Ensemble Classifiers for Complex Word Identification. In *Proceedings of the 13th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 200 – 205. **BEA 2018** June 5, 2018.

6. **Sandeep Mathias**, and Pushpak Bhattacharyya. ASAP++: Enriching the ASAP Automated Essay Grading Dataset with Essay Attribute Scores. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pages 1169 – 1173. **LREC 2018**. Miyazaki, Japan. May 8 – 10, 2018.

7. **Sandeep Mathias**, and Pushpak Bhattacharyya. How Hard Can it Be? The E-Score - A Scoring Metric to Assess the Complexity of Text. In *Proceedings of the 1st Workshop on Quality Assessment for Text Simplification*, pages 10 – 14. **QATS 2016**. Portorož, Slovenia. May 28, 2016.

8. **Sandeep Mathias**, and Pushpak Bhattacaharyya. Using Machine Translation Evaluation Techniques to Evaluate Text Simplification Systems. In *Proceedings of the 1st Workshop on Quality Assessment for Text Simplification*, pages 38 – 41. **QATS 2016**. Portorož, Slovenia. May 28, 2016.

Technical Skills

- **Expertise**: NLP (Natural Language Processing), ML (Machine Learning).
- **Programming Language**: C/C++, Java.

Position of Responsibility

- Served as the **Joint Convenor** of the Literary, Stage and Debate Society in NITK Surathkal for the Academic Year 2009 – '10.
- Served as the **Contingent Leader** for IIT Bombay's contingent at Nihilanth 2015, 2017 and 2019. Nihilanth is the annual Inter-IIT&IIM Quiz Fest.
- Served as the **Captain for the Literary Arts** contingent of IIT Bombay at the Inter-IIT Cultural Meet in 2017 and 2019.

Students Mentored

- Nikhil Wani for his internship @ IIT Bombay (**Duration:** *January to June 2018*)
- Yashasvi Sriram for his RnD project @ IIT Bombay (**Duration:** *July to December 2018*)
- Rahul Kumar for his internship @ IIT Bombay (**Duration:** *June to July 2019*) & his B.Tech. Project @ IIT Patna (**Duration:** *August 2019 to June 2020*)

Other Achievements and Awards

- Was awarded the **Cultural Roll of Honour** at IIT Bombay. This is the highest award given by the IIT Bombay Gymkhana to graduating students for excellence in cultural activities during their stay at IIT Bombay.
- Was awarded the **ACM India IARCS Travel Grant** to attend ACL 2018 in Melbourne, Australia.
- Was awarded the **Cultural Person of the Year** (2018) and the **Cultural Freshman of the Year** (2014) for excellence in cultural activities during the Academic Years 2017 – '18, and 2013 – '14 respectively.

Date: September 1, 2020
Place: Bengaluru

Sandeep Albert Mathias