Sebastin SANTY

🌐 sebast.in   @ hi@sebast.in
github.com/sebastinsanty   in linkedin.com/in/sebastinsanty   🎓 Google Scholar
📍 Microsoft Research, VIGYAN, Lavelle Road, Bangalore, KA, India 560001

EDUCATION

December 2018 August 2015	**Birla Institute of Technology and Science Pilani, GOA, India** Bachelor of Engineering (Hons.), Electronics & Instrumentation

EXPERIENCE

Present
July 2019
Microsoft Research, Bangalore, India | Multilingual Systems
AI Center Fellow | Advisors: Kalika Bali, Monojit Choudhury
Working in the intersection of NLP and ICTD, developing technologies for low-resource languages.

January 2019
Research Intern | Advisors: Kalika Bali, Monojit Choudhury, Sandipan Dandapat
Developed an interactive machine translation interface on top of state-of-the-art transformer model to be used by StoryWeaver for faster turnaround of document translations. [Video]

December 2018
July 2018
Carnegie Mellon University, Pittsburgh, USA
Research Intern (Senior Thesis) | Advisor: David Touretzky
Worked on real-time obstacle avoidance and doorway navigation using a monocular camera on a low-computational robot like Anki Cozmo. [Video]

June 2018
May 2018
University College London, London, UK | Web Intelligence Group
Summer Research Intern | Advisors: Emine Yilmaz, Rishabh Mehrotra
Worked on Task-based Information Retrieval, specifically on recommendation by leveraging external ontology. Used the principles of Bayesian Rose Trees and Distance Dependent Chinese Restaurant Process to design a dynamically enriched ontology. [UserContext]

August 2018
May 2018
NumFOCUS | Julia Computing
Google Summer of Code 2018 | Mentor: Lyndon White
Added data repository support to DataDepsGenerators.jl, which in turn creates registration blocks for DataDeps.jl. DataDeps.jl helps in downloading scientific data with ease. Talk on this work delivered at PyData New York City 2018.

August 2017
May 2017
Mozilla | Bugzilla
Google Summer of Code 2017 | Mentor: Dylan Hardison
Revamped the existing bug filing interface (enter_bug.cgi) into a single page reducing the multi-hop db access and unnecessary template rendering. This enhanced the speed and performance in bug filing.

PUBLICATIONS

The State and Fate of Linguistic Diversity and Inclusion in the NLP World
Pratik Joshi*, **Sebastin Santy***, Amar Budhiraja*, Kalika Bali, Monojit Choudhury
Annual Conference of the Association for Computational Linguistics   [`ACL'20`]

Learnings from Technological Interventions in a Low Resource Language: A Case-Study on Gondi
Devansh Mehta*, **Sebastin Santy***, Ramaravind Mothilal, Brij Mohan Lal Srivastava, Alok Sharma, Anurag Shukla, Vishnu Prasad, Venkanna U, Amit Sharma and Kalika Bali
International Conference on Language Resources and Evaluation   [`LREC'20`]

Deploying Language Technologies for Underserved Communities [Invited]
Kalika Bali, Monojit Choudhury, Sunayana Sitaram, **Sebastin Santy**
UNESCO International Conference on Language Technologies for All   [`LT4All`]

Unsung Challenges of Building and Deploying Language Technologies for Low Resource Language Communities
P. Joshi, C. Barnes, **Sebastin Santy**, S. Khanuja, S. Shah, A. Srinivasan, S. Bhattamishra, S. Sitaram, M. Choudhury, K. Bali
16th International Conference on Natural Language Processing   [`ICON'19`]

CoSSAT: Code-Switched Speech Annotation Tool
Sanket Shah, Pratik Joshi, **Sebastin Santy**, Sunayana Sitaram
2019 Conference on Empirical Methods in Natural Language Processing   [`AnnoNLP@EMNLP'19`]

INMT: Interactive Neural Machine Translation
Sebastin Santy, Sandipan Dandapat, Monojit Choudhury, Kalika Bali
2019 Conference on Empirical Methods in Natural Language Processing [Systems Demo]   [**EMNLP'19**]

Towards Task Understanding in Visual Settings
Sebastin Santy, Wazeer Zulfikar, Rishabh Mehrotra, Emine Yilmaz.
33rd AAAI Conference on Artificial Intelligence [Student Abstract]   [**AAAI'19**]

DataDepsGenerators.jl: Automatic generation of DataDeps.jl registration code
Lyndon White, **Sebastin Santy**
Journal for Open Source Software   [**JOSS'18**]

BITS Darshini: A Modular, Concurrent Protocol Analyzer Workbench
Prasad Talasila, Mihir Kakrambe, Anurag Rai, **Sebastin Santy**, Neena Goveas, Bharat Deshpande
19th International Conference on Distributed Computing and Networking   [**ICDCN'18**]

OTHER SELECTED PROJECTS

Analytics Sensor Test-Kit for Advanced Process Control   NOV'17 - MAY'18
Internet of Things with Nitin Sharma | Bill & Melinda Gates Foundation

> Designed a proof-of-concept IoT scalable interface to gather data from sensors installed at the waste water treatment site. This work was in collaboration with Ohio University, Cranfield University, and Caltech.

REST API: Custom Forms & GraphQL replacement   DEC'16 - MAR'17
With Dylan Hardison, Cameron Dawson & William Lachance (Mozilla)

> Made a custom form example application which uses REST API alongwith auth delegation callbacks to file the application as bugs on bugzilla.mozilla.org. This is to track the data generated through forms as bugs.
> Reponsible for introducing the integration of GraphQL based APIs on Mozilla's automation tool, Treeherder. GraphQL queries only the required variables from the database, which improves the efficiency drastically.

AutolabCLI: Command Line Interface for AutolabJS   AUG'16 - DEC'16
CLI Application with Prasad Talasila

> Developed a CLI for AutolabJS which helps in committing the code to local gitlab server, submitting the code from gitlab server to the evaluation nodes and retrieving the results using a socket connection.
> Published on npm and currently in use for programming labs at the university.

TALKS AND CONFERENCES

PyData New York City   NEW YORK, NY, USA, OCT'18
Delivered a talk. Audience: Data Scientists
Invited to give a talk at PyData New York City 2018 on "Repeatable Data Setup for Repeatable Science" which was based on my Google Summer of Code work with Julia Computing.

Mozilla All-Hands (Group Meeting)   AUSTIN, TX, USA, DEC'17
Attended (Invite Only). Audience: Mozilla Employees
Invited to attend Mozilla's conference held at Austin, Texas. Discussion on future plans of Bugzilla, its modernization, and chalking out the milestone for the year of 2018.

Recent Advances in Computer Vision and Deep Learning   IIIT HYDERABAD, JULY'17
Attended. Audience: Academia and Industrial Researchers
Sessions taken by researchers from DeepMind, Facebook AI Research, Oxford VGG, INRIA, IITs. Various recent topics in Computer Vision and Deep Learning research were discussed.

OPEN SOURCE

> Contributed to Firefox with a globally distributed workforce of Mozilla's engineers.
> Member and Reviewer for Mozilla B-Team, which handles Bugzilla and its operations.
> Managed AutolabJS (programming labs autograder) Github Organisation.
> Actively contributed to scikit-learn, a python-based Machine Learning library, for a period of two months.

TEACHING AND LEADERSHIP ROLES

Present **August 2016**	**Open Source Development** *Founding Member* It is an initiative to foster the open source culture and innovation at BITS Goa. Hosted several informative and technical sessions. Also responsible for making Student Welfare Division software.
May 2018 **Jan 2018**	**Computer Programming Course** *Teaching Assistant* Responsibilities included evaluating labs, and helping students with the course in general.
December 2017 **August 2017**	**Introduction to Programming** *Co-Instructor* Introduced students to multiple programming paradigms, object oriented concepts and functional programming using the language Python.
December 2017 **December 2015**	**Waves, Quark, BITSMUN** *Lead Developer* Responsible for building and maintaining the websites of Waves (cultural fest), Quark (technical fest), BITS Model UN of BITS Pilani Goa.

RELEVANT COURSEWORK

Computer Vision (16720B @CMU)*, Machine Learning, Data Mining, Symbolic Logic, Microprocessors and Interfacing, Digital Design, Signals & Systems, Computer Architecture, Neural Networks and Fuzzy Logic, Data Communication and Networks,Control Systems. (* = Auditing)

HONOURS AND AWARDS

Iken Scientifica, 2009   IIT BOMBAY, '09
Student Icon of India
As a part of the winning team, got an opportunity to have a tête-à-tête with the former President of India, His Excellency, Dr. APJ Abdul Kalam. Entire show was aired on The National Geographic Channel and was supported by the Ministry of Science And Technology, Govt. Of India.

International Computer Fair And Seminar   LUCKNOW, '12
Runner's Up
Runner's up in the web design and implementation round at International Computer Fair and Seminar

National Standard Examination in Chemistry, 2015   TIFR, '15
State Topper
Secured a State Top 1% in NSEC, the prelims to International Chemistry Olympiad

SKILLS

Languages	Python, JS (Node), Julia, Perl, C, C++, SQL, VB.NET
Frameworks	OpenCV, Django, Flask, ExpressJS, ReactJS
Technologies	Heroku, AWS EC2, Travis CI, Docker, Vagrant, Grunt, GraphQL, PyTorch, LaTeX, Git, AutoCAD
Operating Systems	Ubuntu, CentOS, macOS, Windows
Spoken Languages	English, Hindi, Malayalam, Gujarati

REFERENCES

Dr. Kalika Bali *Principal Researcher* MICROSOFT RESEARCH kalikab@microsoft.com	**Dr. Monojit Choudhury** *Principal Researcher* MICROSOFT RESEARCH monojitc@microsoft.com	**Dr. Sandipan Dandapat** *Principal Applied Researcher* MICROSOFT R&D sadandap@microsoft.com	**Prof. David Touretzky** *Research Professor* CARNEGIE MELLON UNIVERSITY dst@cs.cmu.edu
Prof. Emine Yilmaz *Professor* UNIVERSITY COLLEGE LONDON emine.yilmaz@ucl.ac.uk	**Dr. Rishabh Mehrotra** *Senior Research Scientist* SPOTIFY RESEARCH rishabhm@spotify.com	**Dylan Hardison** *Principal Software Engineer* FASTLY (EX-MOZILLA) dylan@hardison.net	**Dr. Lyndon White** *Ph.D. Candidate* U OF WESTERN AUSTRALIA lyndon.white@ucc.asn.au