

Sebastin Santy

🌐 sebast.in @ hi@sebast.in
🐙 github.com/sebastinsanty [in linkedin.com/in/sebastinsanty](https://www.linkedin.com/in/sebastinsanty) [Google Scholar](https://scholar.google.com/citations?user=...)
📍 Microsoft Research, VIGYAN, Lavelle Road, Bangalore, KA, India 560001

Education

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

Experience

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

January 2019 | *Research Intern* | Advisors: Dr. Kalika Bali, Dr. Monojit Choudhury, Dr. Sandipan Dandapat
Worked on understanding how translators can be assisted with suggestions from a machine translation system. On basis of the insights gathered, developed an interactive translation interface to make the translation process quicker and better in terms of quality.

December 2018 | **Carnegie Mellon University, Pittsburgh, USA**
July 2018 | *Research Intern (Senior Thesis)* | Advisor: Prof. David Touretzky
Worked on real-time obstacle collision avoidance and doorway navigation with limited sensor capacity of monocular vision and low computational constraints. We come to a conclusion that free space estimation along with homography-based plane detection works the best for this scenario.

June 2018 | **University College London, London, UK | Web Intelligence Group**
May 2018 | *Summer Research Intern* | Advisors: Prof. Emine Yilmaz, Dr. Rishabh Mehrotra
Worked on user tasks and needs understanding. Used Wikihow as a source of common tasks in form of an external ontology which was further leveraged for providing useful recommendations to users. Further, in order to adapt to increasing number of new tasks, developed a method to dynamically enrich this task ontology using the principles of Bayesian Rose Trees and Chinese Restaurant Processes.

August 2018 | **NumFOCUS | Julia Computing**
May 2018 | *Google Summer of Code 2018* | Mentor: Dr. Lyndon White
Worked on DataDepsGenerators.jl which is a tool to generate executable metadata for published datasets aimed at preserving their provenance information to downstream users for appropriate attribution. Added support for multiple popular data repositories such as FigShare, CKAN, UCI ML, GitHub as well as for sources which have JSON-LD metadata like Kaggle, Zenodo.

August 2017 | **Mozilla | Bugzilla**
May 2017 | *Google Summer of Code 2017* | Mentor: Dylan Hardison
Worked on Bugzilla, a bug-tracking tool to introduce a new single page bug filing interface enter_bug.cgi. This replaced the previous interface which was based on conditional forms that required navigation through multiple pages often making the overall process cumbersome.

Publications

BERTologiCoMix: Probing the Capabilities of Multilingual BERT through Code-Mixing

Sebastin Santy, Anirudh Srinivasan, Monojit Choudhury
[Under Review]

The State and Fate of Linguistic Diversity and Inclusion in the NLP World

Sebastin Santy*, Pratik Joshi*, 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

Sebastin Santy*, Devansh Mehta*, 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 [LT4A11]

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]

Software and Open Source Contributions

- › **Microsoft:** Interactive Neural Machine Translation - [commits](#), [webpage](#)
- › **Mozilla:** Bugzilla - [commits](#), Firefox - [commits](#), TreeHerder - [commits](#)
- › **Others:** scikit-learn - [commits](#), DDGenerators.jl - [commits](#), BITS-Darshini - [commits](#), UserContext - [github](#), ACL - [commits](#)
- › **University:** Setup open source organizations for different academic divisions for students to be able to contribute to their code: Academic Under Graduate Studies Division ([github](#)), Student Welfare Division ([github](#)), Open Source Development Labs ([github](#))

Academic Service

Reviewer	EACL 2021, EMNLP 2020 (Demo)
Sub-Reviewer	EMNLP 2020, ACL 2020, LREC 2020, CoNLL 2019, Interspeech 2019, ICON 2019
Volunteer	AAAI 2019, ACL 2020

Talks

PyData New York City “Repeatable Data Setup for Repeatable Science” [Video]	New York, NY, USA, Oct’18
Mozilla All-Hands “Bugzilla 2018 Milestones”	Austin, TX, USA, Dec’17
Tête-à-tête with APJ Abdul Kalam “Progress of technology around the world and its use in India” [Video]	New Delhi, India, May’09

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