Sebastin **Santy**

Ph.D. student, UNIVERSITY *of* WASHINGTON

🌐 sebast.in   @ ssanty@uw.edu   ⌂ github.com/sebastinsanty   🎓 Google Scholar

📍 Paul G. Allen Center, 185 E Stevens Way NE, Seattle, WA 98195

Education

Starting **Sep 2021**	**University of Washington | Paul G. Allen School of CSE** Ph.D. student in Computer Science & Engineering (CSE)	**Seattle, USA**
Dec 2018 **Aug 2015**	**Birla Institute of Technology and Science (BITS) Pilani** Bachelor in Engineering, Electronics & Instrumentation (Graduated Early)	**Goa, India**

Experience

Jul 2021 **Jul 2019**	**Microsoft Research | Center for Societal Impact through Cloud & AI** [🌐] *AI Center Fellow | Advisors: Dr. Kalika Bali, Dr. Monojit Choudhury, Tanuja Ganu* Worked on understanding the disparity between research and deployment especially for low resource languages and developing technologies to mitigate this gap.	**Bangalore, India**
Jan 2019	*Research Intern | Advisors: Dr. Kalika Bali, Dr. Monojit Choudhury* Developed Interactive Neural Machine Translation which assists translators with suggestions on-the-fly with the aim to bring forth a synergy between the pros of human and machine translation.	
Dec 2018 **Jul 2018**	**Carnegie Mellon University | School of Computer Science** *Research Intern (Bachelor Thesis) | Advisor: Prof. Dave Touretzky* Worked on real-time obstacle collision avoidance and doorway navigation with limited sensor capacity of monocular vision and low computational constraints.	**Pittsburgh, USA**
Jun 2018 **May 2018**	**University College London | Web Intelligence Group** [🌐] *Summer Research Intern | Advisors: Prof. Emine Yilmaz, Dr. Rishabh Mehrotra* Worked on user tasks and needs understanding specifically on understanding how incorporating task information can lead to better, more relevant search/retrieval.	**London, UK**
Aug 2018 **May 2018**	**NumFOCUS | Julia Computing** [🌐] *Google Summer of Code 2018 | Mentor: Dr. Lyndon White* Developed DataDepsGenerators.jl, a provenance-preserving data preprocessing tool to generate executable metadata for published datasets. Added support to popular data repositories such as FigShare, Zenodo.	**Remote**
Aug 2017 **May 2017**	**Mozilla | Bugzilla Team** [🌐] *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 which replaced the previous interface based on conditional forms that required cumbersome navigation.	**Remote / Austin, USA**

Publications

S=In Submission, C=Conference, W=Workshop, P=Poster/Demo, J=Journal

[C.6] **Language Translation as a Socio-Technical System: Case-Studies of Mixed-Initiative Interactions**
<u>Sebastin Santy</u>, Kalika Bali, Monojit Choudhury, Sandipan Dandapat, Tanuja Ganu, Anurag Shukla, Jahanvi Shah, Vivek Seshadri
ACM SIGCAS Conference on Computing and Sustainable Societies (Virtual)   `[COMPASS'21]`

[C.5] **Use of Formal Ethical Reviews in NLP Literature: Historical Trends and Current Practices** 📄
<u>Sebastin Santy</u>, Anku Rani, Monojit Choudhury
Mentions: Academic Freedom and Ethics Review
Annual Conference of the Association for Computational Linguistics (Virtual)   `[ACL'21 Findings]`

[C.4] **The State and Fate of Linguistic Diversity and Inclusion in the NLP World** 📄
<u>Sebastin Santy</u>*, Pratik Joshi*, Amar Budhiraja*, Kalika Bali, Monojit Choudhury (* = Equal Contribution)
Coverage: NLP Newsletter, NLP Beyond English, Quartz, Underrated ML, SIGTYP
Annual Conference of the Association for Computational Linguistics (Virtual)   `[ACL'20]`

[C.3] **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 (* = Equal Contribution)
Coverage: Times of India, Hindustan Times, ETV, LiveMint, The Caravan
International Conference on Language Resources and Evaluation (Virtual) `[LREC'20]`

[C.2] **Unsung Challenges of Building and Deploying Language Technologies for LRL Communities** 📄
Pratik Joshi, Christain Barnes, Sebastin Santy, Simran Khanuja, Sanket Shah, Anirudh Srinivasan, Satwik Bhattamishra, Sunayana Sitaram, Monojit Choudhury, Kalika Bali
16th International Conference on Natural Language Processing, Hyderabad, India `[ICON'19]`

[C.1] **BITS Darshini: A Modular, Concurrent Protocol Analyzer Workbench** 📄
Prasad Talasila, Mihir Kakrambe, Anurag Rai, Sebastin Santy, Neena Goveas, Bharat Deshpande
19th ACM International Conference on Distributed Computing and Networking, Varanasi, India `[ICDCN'18]`

[P.3] **Deploying Language Technologies for Underserved Communities** [Invited Poster] 📄
Kalika Bali, Monojit Choudhury, Sunayana Sitaram, Sebastin Santy
UNESCO International Conference on Language Technologies for All, Paris, France `[LT4All'20]`

[P.2] **INMT: Interactive Neural Machine Translation** 📄
Sebastin Santy, Sandipan Dandapat, Monojit Choudhury, Kalika Bali
Conference on Empirical Methods in Natural Language Processing, Hong Kong [Systems Demo] `[Demo@EMNLP'19]`

[P.1] **Towards Task Understanding in Visual Settings** 📄
Sebastin Santy, Wazeer Zulfikar, Rishabh Mehrotra, Emine Yilmaz.
33rd AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA [Student Abstract] `[Abstract@AAAI'19]`

[W.2] **BERTologiCoMix: How does Code-Mixing interact with Multilingual BERT?**
Sebastin Santy, Anirudh Srinivasan, Monojit Choudhury
Workshop on Domain Adaptation for NLP, EACL (Virtual) `[AdaptNLP@EACL'21]`

[W.1] **CoSSAT: Code-Switched Speech Annotation Tool** 📄
Sanket Shah, Pratik Joshi, Sebastin Santy, Sunayana Sitaram
Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP, EMNLP, Hong Kong `[AnnoNLP@EMNLP'19]`

[J.1] **DataDepsGenerators.jl: Automatic generation of DataDeps.jl registration code** 📄
Lyndon White, Sebastin Santy
Journal for Open Source Software `[JOSS'18]`

Honours and Awards

BITS Alumni Association (BITSAA) Travel Grant, 2019 [🌐] For attending AAAI'19 held in Honolulu, HI, USA.

BITS International Programmes Scholarship, 2018 Awarded for pursuing Summer Internship at UCL, London, UK.

Mozilla All-Hands, 2017 [🌐] Invited as one of 100 contributors to attend this employee meet held at Austin, TX, USA.

Iken Scientifica, 2009 | Student Icon of India [🎥] [🌐] Adjudged as Student Icon of India among $\sim$ 200k participants. Invited for a tête-à-tête with the former President of India, and renowned rocket scientist, Dr. APJ Abdul Kalam at his residence. The entire competition was aired on The National Geographic Channel and was supported by the Ministry of Science And Technology, Government Of India.

National Standard Examination in Chemistry 2015 | State Topper Prelims to International Chemistry Olympiad.

C.Subramaniam Award for Character Excellence, 2013 Among Secondary School students across the school system.

International Computer Fair And Seminar (CoFAS), 2012 | Runner's Up Web design and implementation Round.

Talks

"The State and Fate of Linguistic Diversity in the NLP World"
  › NLP with Friends [🌐] [🎥] .... Nov '20 (Remote)
  › ACL 2020 Presentation [🌐] [🎥] .... Jul '20 (Remote)
  › ML Invited Authors Series, Ada Support .... Jun '20 (Remote)
  › Lab Sabha, Microsoft Research India .... May '20 (Remote)

"Repeatable Data Setup for Repeatable Science"
  › PyData New York City 2018 [🌐] [🎥] .... Oct '18 (Microsoft NYC, USA)

Teaching and Leadership Roles

SNLP Reading Group, MSR India   *Organizer*   Jul'19 - Present
> Organize a weekly lab-wide Reading Group focused on research taking place in the area of Speech and Natural Language Processing (SNLP). We read recent and classical papers as well as arrange for invited talks in related areas.

Open Source Development (OSD) Labs, BITS Goa   *Founding Member*   [○]   Aug'16 - Apr'18
> Initiative to foster the open source culture at BITS Goa. Hosted several informative and technical sessions.
> Setup new open source organizations for different academic divisions such as Academic Registration and Counselling Division [○], Student Welfare Division [○] to enable students to actively contribute to university softwares.

Introduction to Computer Programming (CS F111)   *Teaching Assistant*   Jan'18 - May'18
> Responsibilities included evaluating labs, and helping students with the coursework and home/lab assignments.

Introduction to Programming, Center for Technical Education (CTE), BITS Goa   *Co-Instructor*   Dec'17 - Aug'17
> Voluntarily teaching students to object oriented concepts and functional programming using Python.

Waves, Quark, BITSMUN (Festivals at BITS Goa)   *Lead Developer*   Dec'15 - Dec'17
> Responsible for building and maintaining websites of Waves (cultural fest), Quark (technical fest), BITS Model UN.

Software and Open Source Contributions

> **Microsoft**: Interactive Neural Machine Translation - *commits*, *webpage*
> **Mozilla**: Bugzilla - *commits*, Firefox (Gecko Engine) - *commits*, TreeHerder - *commits*
> **Other External**: scikit-learn - *commits*, DDGenerators.jl - *commits*, Virtual ACL 2020 - *commits*, UserContext - *github*
> **@BITS**: BITS-Darshini - *commits*, SWD - *commits*, ARC - *commits*, Quark 2017 - *commits*, Waves 2016 - *commits*

Academic Service

Reviewer	EMNLP '21, ACL'21, EACL'21, EMNLP'21 Demo, NAACL'21 Demo, EMNLP'20 Demo, MLADS'20, ICON'20
Sub-Reviewer	EMNLP'20, ACL'20, LREC'20, CODS-COMAD'20, CoNLL'19, Interspeech'19, ICON'19
Volunteer	ACL'20, AAAI'19, Panini Linguistics Olympiad (PLO) '19/'20
Tech Staff	ACL Rolling Review, ACL'20 Virtual Conference

References

> Dr. Kalika Bali .... *Principal Researcher, Microsoft Research, India* [◉]
> Dr. Monojit Choudhury .... *Principal Researcher, Microsoft Research, India* [◉]
> Prof. Dave Touretzky .... *Research Professor, Carnegie Mellon University, USA* [◉]
> Prof. Emine Yilmaz .... *Professor & Turing Fellow, University College London, UK* [◉]
> Dr. Rishabh Mehrotra .... *Staff Research Scientist & Research Lead, Spotify Research, UK* [◉]
> Dylan Hardison .... *Principal Software Engineer, Fastly, USA* [◉]