

Francielle Vargas

francielleavargas@usp.br +55 31 999685891
[\[web page\]](#) [\[dblp\]](#) [\[linkedin\]](#) [\[google scholar\]](#) [\[github\]](#)

Education

PhD, Computer Science and Computational Mathematics
University of São Paulo
August 2019 - Current

MSc, Computer Science and Computational Mathematics
University of São Paulo
Thesis: Semantic clustering of aspects for opinion mining. [see](#)
2015 - 2017

BA, Linguistics
Federal University of Minas Gerais
2010 - 2014

BSc, Information Systems
Pontifical Catholic University of Minas Gerais
2006 - 2009

Research Groups

[LOCUS - Social Computing Lab](#), 2021/2023
Computer Science Department, Federal University of Minas Gerais, Brazil.

[NILC - Interinstitutional Center for Computational Linguistics](#), 2015/2023
Institute of Mathematical and Computer Sciences, University of São Paulo, Brazil.

Organizing and Program Committee

Program Committee

- **ACL**
Computational Social Science and Cultural Analytics Track (2022)
Discourse and Pragmatics Track (2023)
- **EMNLP**
Computational Social Science and Cultural Analytics Track ([2022](#))
Discourse and Pragmatics Track ([2022](#))
- **ICWSM**
17th International AAAI Conference on Web and Social Media ([2023](#))
- **NAACL**
NAACL Student Research Workshop ([2022](#))
- **SoCal NLP**
5th Southern California Natural Language Processing Symposium ([2022](#))
- **WOAH**
7th Workshop on Online Abuse and Harms ([ACL 2023](#))
6th Workshop on Online Abuse and Harms ([NAACL-HLT 2022](#))
- **FEVER**
6th Fact Extraction and VERification ([EACL 2023](#))

- **CASE**
6th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text ([RANLP 2022](#))
5th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text ([EMNLP 2022](#))
4th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text ([ACL-IJCNLP 2021](#)).
- **CODI**
4th Workshop on Computational Approaches to Discourse ([ACL 2023](#))
3rd Workshop on Computational Approaches to Discourse ([COLING 2022](#))
- **ARGMINING**
9th Workshop on Argument Mining ([COLING 2022](#))
22th Workshop on Computational Models of Natural Argument ([COMMA 2022](#))
1st Workshop on Argumentation & Machine Learning ([COMMA 2022](#)).

Organizing Committee

1. Data(set) Co-chair. *ICWSM 2023*. [see](#)
2. Accessibility Chair. *ICWSM 2022*. [see](#)
3. Diversity, Equity, and Inclusion Chair. *ICWSM 2021*. [see](#)
4. Steering Committee. 6th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text. (*RANLP 2023*). [see](#)
5. Steering Committee. 5th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text. (*EMNLP 2022*). [see](#)

Selected Publications

Journals

Vargas, F.A., Pardo, T.A.S., Benevenuto, F., Schmeisser, W., Zaczynska, Z., Rabinovich, Z. (2022). Discourse Annotation Guideline for Low-Resource Languages. *Natural Language Engineering Journal*, pp. 1-56. [under review](#)

Vargas, F.A., Carvalho, I., Góes, F.R., Pardo, T.A.S and Benevenuto, F. (2022). Contextual-aware and data expert resources for Brazilian Portuguese hate speech detection. *Natural Language Engineering Journal*, pp. 1-25. [under review](#)

Conferences

Vargas, F., Góes, F.R., Pardo, T.A.S., Benevenuto, F. (2023). Factual or Biased? Predicting Sentence-Level Factuality and Bias of News. *61st Annual Meeting of the Association for Computational Linguistics (ACL)*. pp.1-5. Toronto, Canada. [under review](#)

Vargas, F., Guimarães, S., Lamine, M. D., Benevenuto, F. (2023). TEAR: A Dataset of Hate Speech Comments on West African Facebook Pages. *17th International AAAI Conference on Web and Social Media (ICWSM)*. pp.1-8. Limassol, Cyprus. [under review](#)

Vargas, F.A., D'Alessandro, J., Rabinovich, Z., Benevenuto, F., Pardo, T.A.S (2022). Rhetorical structure approach online deception detection: A survey. *13th Conference on Language Resources and Evaluation (LREC)*, pp. 5906-5915. Marseille, France. [see](#)

Vargas, F.A., Carvalho, I., Góes, F.R., Pardo, T.A.S and Benevenuto, F. (2022). HateBR: A expert annotated corpus of Brazilian Instagram comments for offensive language and hate speech detection. *13th Conference on Language Resources and Evaluation (LREC)*, pp. 7174-7183. Marseille, France. [see](#)

Machado, M., Pardo, T.A.S., Ruiz, E. E. S., Di Felippo, A., **Vargas, F.A.** (2022). Implicit aspect clues in Portuguese texts: analysis and categorization. *15th International Conference on Computational Processing of Portuguese (PROPOR)*, pp. 68-78. Fortaleza, Brazil. [see](#)

Vargas, F.A. Benevenuto, F. , Pardo, T.A.S (2021). Towards discourse-aware models for multilingual fake news detection. *International Conference Recent Advances in Natural Language Processing - Deep Learning for Natural Language Processing Methods and Applications - Student Research Workshop (RANLP)*, pp. 210-218. Held Online. [see](#)

Vargas, F.A., Góes, F.R., Carvalho, I., Benevenuto, F. and Pardo, T.A.S. (2021). Contextual-lexicon approach for abusive language detection. *International Conference Recent Advances in Natural Language Processing - Deep Learning for Natural Language Processing Methods and Applications (RANLP)*, pp. 1442-1451. Held Online. [see](#)

Vargas, F.A., Santos, R. R. S and Rocha, P. R. (2020) Identifying fine-grained opinion and classifying polarity on coronavirus pandemic. *9th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 511-520. Rio Grande, Brazil. [see](#)

Vargas, F.A. and Pardo, T.A.S. (2018). Aspect clustering methods for sentiment analysis. *13th International Conference on Computational Processing of Portuguese (PROPOR)*, pp. 365-374. Canela, Brazil. [see](#)

Workshops

Nurse, J. R., **Vargas, F. A.**, Hassan, N, Wang, D., Andriotis, P., Ghenai, A., Jaidka, K., Mustafaraj, E., Joseph, K., Welles, B. F. (2021) Towards a diverse, inclusive, accessible and equitable AAAI international conference on web and social media (ICWSM). *Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media (ICWSM)*, pages 1-14. Held Online. [see](#)

Vargas, F.A. and Pardo, T.A.S (2020). Linguistic rules for fine-grained opinion extraction. *Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media (ICWSM)*, pp. 1-6. Held Online. [see](#)

Pardo, T.A.S., Baptista, J., Duran, M.S., Nunes, M.G.V., Nóbrega, F.A.A., Aluísio, S.M., Di Felippo, A., Seno, E.R.M., Silva, R.R., Anchieta, R.T., Brum, H.B., Dias, M.S., Martins, R.S.O., Maziero, E.G., Souza, J.W.C., **Vargas, F.A.** (2017). The coreference annotation of the CSTNews corpus. In the *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages*. Workshop Proceedings of the 33th Conference of the Spanish Society for Natural Language Processing (SEPLN). 102-112. Murcia, Spain. [see](#)

Previous Positions

Academic Positions

2019/2022: **Graduate Research Assistant - Dept. of Computer Science.**
University of São Paulo. [see](#)
2021/2021:**Graduate Teaching Assistant in Neural Networks and Deep Learning.**
University of São Paulo. [see](#)
2020/2020:**Graduate Teaching Assistant - Computing Theory and Compilers.**
University of São Paulo. [see](#)
2016/2016: **Graduate Teaching Assistant in Computer Seminars.**
University of São Paulo. [see](#)
2015/2017: **Graduate Research Assistant - Dept. of Computer Science.**
University of São Paulo. [see](#)
2011/2014: **Undergraduate Research Assistant - Dept. of Linguistics.**
Federal University of Minas Gerais. [see](#)

Industry Positions

2021/2023 **Research Fellow**. Sinch.
2021/2022 **Data Scientist**. Webex-Cisco via Telus International.
2014/2015: **System Analyst**. Unisys Corporation.
2009/2010: **System Analyst**. Grupo Barcelos.

Conference participation

2022: 13th International Conference on Language Resources and Evaluation (LREC). [see](#)
2021: 16th Recent Advances in Natural Language Processing (RANLP). [see](#)
2021: 15th International AAAI Conference on Web and Social Media (ICWSM) . [see](#)
2020: The 2022 Empirical Methods in Natural Language Processing (EMNLP). [see](#)
2020: 9th Brazilian Conference on Intelligent Systems (BRACIS). [see](#)
2020: 12th International Conference on Language Resources and Evaluation (LREC). [see](#)
2017: 9th Brazilian School on Computational Linguistics (EBRALC). [see](#)
2013: 29th Congresso Latino Americano de Sociologia: Crise e Emergências Sociais em América Latina (ALAS). [see](#)

Mentoring

2020: **Computer Science Undergraduate Student**, Lucas Sobral Fontes Cardoso: *Title: Web resources for an aspect identification and clustering tool in fine-grained opinion mining.* University of São Paulo. [certificate](#)

Awards received

2013: **Best Undergraduate Student Award**, Federal University of Minas Gerais. [see](#)
2013: **Best Undergraduate Project Award**, Federal University of Minas Gerais. [see](#)
2012: **Best Undergraduate Student Award**, Federal University of Minas Gerais. [see](#)

Short Courses

2019: Introduction on the Stars - Astrophysics. [certificate](#)

Hackathon, Datathon and Summer Schools

- The 2022 IEEE Spoken Language Technology Workshop Hackathon, Qatar, 2022. [see](#)
- 2nd Advanced Language Processing School, Université Grenoble Alpes, France, 2022. [see](#)
- 1st Datathon and Hackathon “Antisemitism on Social Media”, Indiana University Bloomington, [NLP Lab](#), United States, 2020. [see](#)
- 10th Lisbon Machine Learning School, Portugal, 2020. [see](#)
- 4th Brazilian Advanced School in Big Data Analysis, University of São Paulo, 2020. [see](#)
- 22th Computer Science Summer School, University of São Paulo, 2019. [certificate](#)
- 4th Computer Science Summer School, Federal University of Minas Gerais, 2015. [see](#)
- 10th Linguistics Summer School, Federal University of Minas Gerais, 2011. [see](#)

Participation in examination committees

2021: **Computer Science Graduate Course**. University of São Paulo. Vivian Vieira. *Aprendendo sobre veracidade de informações: Desenvolvimento e avaliação de curso EAD que aplica método científico na análise de notícias.* [certificate](#)
2021: **Computer Science Graduate Course**. University of São Paulo. Laudeir José Lopes de Oliveira. *Explorando o pensamento computacional com aprendizado de máquina: Elaboração de um material didático para uma oficina introdutória.* [certificate](#)
2021: **Computer Science Undergraduate Course**. University of São Paulo. Lucas Silva Loureiro. *Sistema de Recomendação de Livros.* [certificate](#)
2020: **Computer Science Undergraduate Course**. University of São Paulo. Luana Balador Belisario. *Análise discursiva aplicada à mineração de opinião: classificação de subjetividade para a língua portuguesa.* [see](#)
2020: **Computer Science Undergraduate Course**. University of São Paulo. Paulo André

Pimenta Aragão. *Revisão Sistemática da Literatura sobre Software Educativos para Ensino de Engenharia de Software*. [see](#)
 2020: **Computer Science Undergraduate Course**. University of São Paulo. Tarcisio Ladeia de Oliveira. *Mapeamento de Ambientes Externos com o Uso de Drones*. [see](#)
 2020: **Computer Science Undergraduate Course**. University of São Paulo. Rafael Farias Pinho. *Planejamento de rotas para VANTs com desvio de obstáculos e incertezas*. [see](#)
 2019: **27th USP International Symposium on Scientific and Technological Initiation**. *Member of the examination committee of the undergraduate projects*. [certificate](#)

Talks as an invited speaker 2023: Day of Women and Girls in Science, Sinch Latam. [see](#)
 2018: Natural Language Processing and Data Science, University Center Methodist Izabela Hendrix. [see](#)
 2016: Opinion Mining, University of Sao Paulo. [see](#)

Technical-Scientific Resources **OpCluster**: Fine-grained opinion extraction and clustering tool ([web system](#)).
NoHateBrazil: A fairness-aware offensive comments analysis web system ([web System](#))
OPCovidBR: Aspect-based sentiment analysis dataset of Covid-19 Tweets ([dataset](#)).
AspectPRO: Aspect-based annotated corpus of web consumer reviews ([dataset](#)).
HateBR: Large-scale abusive language dataset of Brazilian Instagram comments ([dataset](#)).
Deceiver: Multilingual discourse-annotated corpus of fake and truth news ([dataset](#)).
FactNews: Sentence-level annotated dataset to predict factually and media bias ([dataset](#)).
PRO: Taxonomies for aspect-based sentiment analysis in Portuguese ([lexicon](#)).
MOL: Multilingual offensive lexicon ([lexicon](#)).

Recommendation Letters **Prof. Dr. Fabricio Benevenuto**. *Computer Science Department, Federal University of Minas Gerais*. [letter](#)
Dra. Norah Abokhodair. *Microsoft Research*. [letter](#)
Prof. Dra. Roseli Romero. *Institute of Mathematics and Computer Sciences, University of São Paulo*. [letter](#)
Prof. Dr. Carlos Ramisch. *Laboratoire d'Informatique et Systèmes – LIS – UMR CNRS 7020 – Aix Marseille Université*. [letter](#)

Languages Python, SQL, RDF(s), OWL, XML, JSON, Unix/Bash.
