

Francielle Vargas

francielleavargas@usp.br +55 31 999685891

[\[page\]](#) [\[dblp\]](#) [\[linkedin\]](#) [\[google scholar\]](#) [\[github\]](#)

Education	<i>PhD, Computer Science and Computational Mathematics</i> University of São Paulo August 2019 - current
	<i>MSc, Computer Science and Computational Mathematics</i> University of São Paulo Thesis: Semantic clustering of aspects for opinion mining. see August 2015 - November 2017
	<i>BA, Linguistics</i> Federal University of Minas Gerais August 2010 - June 2014
	<i>BSc, Information Systems</i> Pontifical Catholic University of Minas Gerais August 2006 - December 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
	<ol style="list-style-type: none">1. NLP Applications Track. <i>EMNLP 2022</i>. see2. Natural Language Generation Track. <i>EMNLP 2022</i>. see3. Discourse and Pragmatics Track. <i>EMNLP 2022</i>. see4. Resources and Evaluation Track. <i>EMNLP 2022</i>. see5. Computational Social Science and Cultural Analytics Track. <i>EMNLP 2022</i>. see6. Sentiment Analysis, Stylistic Analysis, and Argument Mining Track. <i>EMNLP 2022</i>. see7. 3rd Workshop on Computational Approaches to Discourse. <i>COLING 2022</i>. see8. 9th Workshop on Argument Mining. <i>COLING 2022</i>. see9. 6th Workshop on Online Abuse and Harms. <i>NAACL-HLT 2022</i>. see10. NAACL Student Research Workshop. <i>NAACL-HLT 2022</i>. see11. 22th Workshop on Computational Models of Natural Argument. <i>COMMA 2022</i>. see12. 1st Workshop on Argumentation & Machine Learning. <i>COMMA 2022</i>. see13. 5th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text. <i>EMNLP 2022</i>. see14. 4th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text. <i>ACL-IJCNLP 2021</i>. see

Organizing Committee

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

Publications

Journals

Vargas, F.A., Pardo, T.A.S., Benevenuto, F., Schmeisser, W., Zaczynska, Z., Rabinovich, Z. (2022). Discourse-annotation guideline using rhetorical structure theory: multilingual examples. *Dialogue & Discourse Journal*, pp. 1-63. [submitted](#)

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

Conferences

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. [pdf](#)

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. [pdf](#)

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. [pdf](#)

Vargas, F.A. Benevenuto, F. , Pardo, T.A.S (2021). Toward discourse-aware models for multilingual fake news detection. *International Conference Recent Advances in Natural Language Processing (RANLP)*, pp. 210-218. Held Online. [pdf](#)

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. [pdf](#)

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. [pdf](#)

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. [pdf](#)

Vargas, F.A. and Pardo, T.A.S. (2017). Hierarchical clustering of aspects for opinion mining: a corpus study. *9th Escola Brasileira de Linguística Computacional (EBRALC)*, pp. 69-84. São Leopoldo, Brazil. [pdf](#)

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. [pdf](#)

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. [pdf](#)

Vargas, F.A., Pardo, T.A.S. (2020) Studying dishonest intentions in texts. *st International Workshop on Deceptive AI*. In Communications in Computer and Information Science book series (CCIS,volume 1296). pp. 166-178. Held Online. [pdf](#)

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., Anchiêta, 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. [pdf](#)

Technical reports

Vargas, F.A. and Pardo, T.A.S. (2017). Empirical study on clustering and hierarchical organization of aspects for opinion mining. *Technical ICMC's Reports, USP, no.418.48p*. [pdf](#)

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

2021/2022 **Data Science Analyst.** Webex-Cisco via Telus International.
2018/2018 **Language Engineer.** Amazon via Appen.
2014/2015: **System Analyst.** Unisys Corporation.
2009/2010: **System Analyst.** Grupo Barcelos.

Conference participation

2022: 13th International Conference on Language Resources and Evaluation (LREC). [pdf](#)
2021: 16th Recent Advances in Natural Language Processing (RANLP). [pdf](#)
2021: 15th International AAAI Conference on Web and Social Media (ICWSM) . [pdf](#)
2020: The 2022 Empirical Methods in Natural Language Processing (EMNLP). [pdf](#)

	<p>2020: 9th Brazilian Conference on Intelligent Systems (BRACIS). pdf</p> <p>2020: 12th International Conference on Language Resources and Evaluation (LREC). pdf</p> <p>2017: 9th Brazilian School on Computational Linguistics (EBRALC). pdf</p> <p>2013: 29th Congresso Latino Americano de Sociologia: Crise e Emergências Sociais em América Latina (ALAS). pdf</p> <p>2013: 4th International Symposium on Letters and Linguistics (SILEL) pdf</p>
Mentoring	<p>2020: Computer Science Undergraduate Student, Lucas Sobral Fontes Cardoso: <i>Title: Web resources for an aspect identification and clustering tool in fine-grained opinion mining.</i> University of São Paulo. certificate</p>
Awards received	<p>2013: Best Undergraduate Student Award, Federal University of Minas Gerais. pdf</p> <p>2013: Best Undergraduate Project Award, Federal University of Minas Gerais. pdf</p> <p>2012: Best Undergraduate Student Award, Federal University of Minas Gerais. pdf</p>
Short Courses	<p>2019: Introduction on the Stars - Astrophysics. certificate</p>
Hackathon, Datathon and Summer Schools	<p>2nd Advanced Language Processing School, Université Grenoble Alpes, France, 2022.pdf</p> <p>1st Datathon and Hackathon “Antisemitism on Social Media”, Indiana University Bloomington, NLP Lab, United States, 2020.pdf</p> <p>10th Lisbon Machine Learning School, Portugal, 2020.pdf</p> <p>4th Brazilian Advanced School in Big Data Analysis, University of São Paulo, 2020. pdf</p> <p>22th Computer Science Summer School, University of São Paulo, 2019.pdf</p> <p>4th Computer Science Summer School, Federal University of Minas Gerais, 2015. pdf</p> <p>10th Linguistics Summer School, Federal University of Minas Gerais, 2011.pdf</p>
Participation in examination committees	<p>2021: Computer Science Graduate Course. University of São Paulo. Vivian Vieira. <i>Aprendendo sobre veracidade de informações: Desenvolvimento e avaliação de curso EAD que aplica método científico na análise de notícias.</i> pdf</p> <p>2021: Computer Science Graduate Course. University of São Paulo. Laudeir José Lopes de Oliveira. <i>Explorando o pensamento computacional com aprendizado de máquina: Elaboração de um material didático para uma oficina introdutória.</i> pdf</p> <p>2021: Computer Science Undergraduate Course. University of São Paulo. Lucas Silva Loureiro. <i>Sistema de Recomendação de Livros.</i> pdf</p> <p>2020: Computer Science Undergraduate Course. University of São Paulo. Luana Balador Belisario. <i>Análise discursiva aplicada à mineração de opinião: classificação de subjetividade para a língua portuguesa.</i> see</p> <p>2020: Computer Science Undergraduate Course. University of São Paulo. Paulo André Pimenta Aragão. <i>Revisão Sistemática da Literatura sobre Software Educativos para Ensino de Engenharia de Software.</i> see</p> <p>2020: Computer Science Undergraduate Course. University of São Paulo. Tarcisio Ladeia de Oliveira. <i>Mapeamento de Ambientes Externos com o Uso de Drones.</i>see</p> <p>2020: Computer Science Undergraduate Course. University of São Paulo. Rafael Farias Pinho. <i>Planejamento de rotas para VANTs com desvio de obstáculos e incertezas.</i> see</p> <p>2019: 27th USP International Symposium on Scientific and Technological Initiation. <i>Member of the examination committee of the undergraduate projects.</i> pdf</p>

Talks as an invited speaker	Natural Language Processing and Data Science, University Center Methodist Izabela Hendrix, 2018. pdf Opinion Mining, University of Sao Paulo, 2016. pdf
-----------------------------	--

Technical-Scientific Resources	OpCluster : Fine-grained opinion extraction and clustering tool. (tool) (web version) 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) PRO : Taxonomies for aspect-based sentiment analysis in Portuguese (lexicon) MOL : Multilingual offensive lexicon (lexicon)
--------------------------------	---

Recommendation Letters	Prof. Dr. Fabricio Benevenuto . <i>Computer Science Department, Federal University of Minas Gerais</i> . letter Dra. Norah Abokhodair . <i>Microsoft Research</i> . letter Prof. Dra. Roseli Romero . <i>Institute of Mathematics and Computer Sciences, University of São Paulo</i> . letter Prof. Dr. Carlos Ramisch . <i>Laboratoire d'Informatique et Systèmes – LIS – UMR CNRS 7020 – Aix Marseille Université</i> . letter Prof. Dr. Thiago A. S. Pardo . <i>Institute of Mathematics and Computer Sciences, University of São Paulo</i> . letter
------------------------	--

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