

Joaquim Francisco Ferreira Silva
Assistant Professor
DI - Departamento de Informática
NOVALincs
Type of address: Visiting address.
Campus da Caparica, Monte da Caparica, FCT-UNL
Edif. II, Departamento de Informática
2829-516 Caparica
Caparica
Portugal
Email: jfs@fct.unl.pt

Qualifications

Computer Science, Doctorate, Extraction of Textual Units, Clustering, Characterization e Classification of Documents,
Universidade NOVA de Lisboa
Award Date: 4 Feb 2004

Employment

Assistant Professor

DI - Departamento de Informática
Universidade NOVA de Lisboa
Portugal
4 Feb 2004 → present

Assistant Professor

NOVALincs
Universidade NOVA de Lisboa
Portugal
1 Jan 2015 → present

Assistant Professor

CITI - Centro de Informática e Tecnologias da Informação
Universidade NOVA de Lisboa
4 Feb 2004 → 31 Dec 2014

Research outputs

How Large Corpora Sizes Influence the Distribution of Low Frequency Text n-grams

Silva, J. F. & Cunha, J. C., 25 Apr 2024, *Advances in Knowledge Discovery and Data Mining - 28th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2024, Proceedings*. Yang, D-N., Xie, X., Tseng, V. S., Pei, J., Huang, J-W. & Lin, J. C-W. (eds.). Springer Science and Business Media Deutschland GmbH, p. 210-222 13 p. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); vol. 14647 LNAI).

Improving LocalMaxs Multiword Expression Statistical Extractor

Silva, J. F. & Cunha, J. C., 2023, *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part II*. Mikyška, J., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V. V., Dongarra, J. J. & Sloot, P. M. A. (eds.). Cham: Springer, p. 154-162 9 p. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); vol. 14074 LNCS).

Vocalization Features to Recognize Small Dolphin Species for Limited Datasets

Rosário, L., Cavaco, S., Silva, J., Freitas, L. & Verborgh, P., 2023, *Progress in Artificial Intelligence: 22nd EPIA Conference on Artificial Intelligence, EPIA 2023, Faial Island, Azores, September 5–8, 2023, Proceedings, Part I*. Moniz, N., Vale, Z., Cascalho, J., Silva, C. & Sebastião, R. (eds.). Cham: Springer, p. 171-183 13 p. (Lecture Notes in Computer Science; vol. 14115 LNAI).

A Model for Predicting n-gram Frequency Distribution in Large Corpora

Silva, J. F. & Cunha, J. C., 2021, *Computational Science – ICCS 2021: 21st International Conference, Krakow, Poland, June 16–18, 2021, Proceedings, Part I*. Paszynski, M., Kranzlmüller, D., Krzhizhanovskaya, V. V., Dongarra, J. J. & Sloot, P. M. A. (eds.). 1st ed. Cham: Springer, Vol. 1. p. 699-706 8 p. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); vol. 12742 LNCS).

An Empirical Model for n-gram Frequency Distribution in Large Corpora

Silva, J. F. & Cunha, J. C., 2020, *Advances in Knowledge Discovery and Data Mining - 24th Pacific-Asia Conference, PAKDD 2020, Proceedings*. Lauw, H. W., Lim, E-P., Wong, R. C-W., Ntoulas, A., Ng, S-K. & Pan, S. J. (eds.). Cham: Springer, p. 840-851 12 p. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); vol. 12085 LNAI).

n-gram Cache Performance in Statistical Extraction of Relevant Terms in Large Corpora

Gonçalves, C., Silva, J. F. & Cunha, J. C., 2019, *Computational Science – ICCS 2019 - 19th International Conference, Proceedings*. Rodrigues, J. M. F., Cardoso, P. J. S., Monteiro, J., Lam, R., Krzhizhanovskaya, V. V., Lees, M. H., Sloot, P. M. A. & Dongarra, J. J. (eds.). Cham: Springer, p. 75-88 14 p. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); vol. 11537 LNCS).

Automatic Classification of Impact Sounds with Rejection of Unknown Samples

da Silva, J. F., Cavaco, S. & Lopes, G. P., 1 Oct 2017, In: *New Generation Computing*. 35, 4, p. 341-363 23 p.

An n-gram cache for large-scale parallel extraction of multiword relevant expressions with LocalMaxs

Gonçalves, C., Silva, J. F. F. & Cunha, J. A. C. E., 3 Mar 2017, *Proceedings of the 2016 IEEE 12th International Conference on eScience*. Institute of Electrical and Electronics Engineers (IEEE), p. 120-129 10 p. 7870892

A theoretical model for n-gram distribution in big data corpora

Silva, J. F., Gonçalves, C. & Cunha, J. C., 1 Jan 2016, *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*. Ak, R., Karypis, G., Xia, Y., Hu, X. T., Yu, P. S., Joshi, J., Ungar, L., Liu, L., Sato, A-H., Suzumura, T., Rachuri, S., Govindaraju, R. & Xu, W. (eds.). Institute of Electrical and Electronics Engineers (IEEE), p. 134-141 8 p. 7840598

Cross-lingual Word Sense Clustering for Sense Disambiguation

Casteleiro, J., Silva, J. F. F. & Lopes, J. G. P., 25 Aug 2015, *EPIA 2015: Progress in Artificial Intelligence*. Pereira, F., Machado, P., Costa, E. & Cardoso, A. (eds.). Cham: Springer International Publishing, p. 747-758 12 p. (Lecture Notes in Computer Science (LNCS); vol. 9273).

A Parallel Algorithm for Statistical Multiword Term Extraction from Very Large Corpora

Gonçalves, C., Silva, J. F. F. & Cunha, J. A. C. E., 2015, *Proceedings - 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security and 2015 IEEE 12th International Conference on Embedded Software and Systems (HPCC-CSS-ICSS 2015)*. Institute of Electrical and Electronics Engineers (IEEE), p. 219-224 6 p. 7336167

Bilingually Learning Word Senses for Translation

Lopes, J. G. P. & Silva, J. F. F., 1 Jan 2014, *Lecture Notes in Computer Science*. p. 283 to 295

Context Sense Clustering for Translation

Lopes, J. G. P. & Silva, J. F. F., 1 Jan 2014, *EMNLP Conference, Association for Computational Linguistics*. p. 135 to 137

Automatic Extraction of Explicit and Implicit Keywords to Build Document Descriptors

Silva, J. F. F., 1 Jan 2013, *Progress in Artificial Intelligence*. Correia, L., Reis, L. P. & Cascalho, J. (eds.). Heidelberg: Springer Berlin Heidelberg, p. 492-503 (Lecture Notes in Computer Science; no. 8154).

Using Clusters of Concepts to Extract Semantic Relations from Standalone Documents

Silva, J. F. F., 1 Jan 2013, *Progress in Artificial Intelligence*. Correia, L., Reis, L. P. & Cascalho, J. (eds.). Heidelberg: Springer Berlin Heidelberg, p. 516-527 (Lecture Notes in Computer Science; no. 8154).

Finding occupational accident patterns in the extractive industry using a systematic data mining approach
Silva, J. F. F. & Jacinto, C., 1 Jan 2012, In: *Reliability Engineering & System Safety*. 108, NA, p. 108-122

Mining Causality from Non-categorical Numerical Data

Lopes, J. G. P. & Silva, J. F. F., 1 Jan 2012, *Behavior Computing. Analysis, Mining and Decision*. London: Springer London, p. 215-227

Mining Concepts from texts

Silva, J. F. F., 1 Jan 2012, *SciVerse ScienceDirect, Procedia Computer Science*. p. 27-36

Text Categorization: An extensive comparison of classifiers, feature selection metrics and document representation

Lopes, J. G. P. & Silva, J. F. F., 1 Jan 2011, *Proceedings of the 15th Portuguese Conference in Artificial Intelligence, EPIA 2011..* p. 660 to 674

Unsupervised Music Genre Classification with a Model-Based Approach

Silva, J. F. F. & Cavaco, S. C. F. M., 1 Jan 2011, *Lecture Notes in Computer Science*. p. 268-281

Towards Automatic Building of Document Keywords

Lopes, J. G. P. & Silva, J. F. F., 1 Jan 2010, *International Conference on Computational Linguistics*. p. 1149-1157

A Document Descriptor Extractor Based on Relevant Expressions

Lopes, J. G. P. & Silva, J. F. F., 1 Jan 2009, *Lecture Notes in Computer Science*. Springer Verlag, p. 646-657

Using Covariance as a Similarity Measure for Document Language Identification in Hard Contexts

Lopes, J. G. P. & Silva, J. F. F., 19 Mar 2007, *Proceedings of the XII International Summer Conference on Probability and Statistics and Seminar on Statistical Data Analysis, SDA, 2006*. Yanev, N. (ed.). Sofia: Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, p. 341-360 20 p. (Pliska Studia Mathematica Bulgarica, Bulgaria; vol. 18).

A divide-and-conquer approach to acquire syntactic categories

Gamallo, P., Lopes, G. P. & Da Silva, J. F., 2004, *Grammatical Inference: Algorithms and Applications: 7th International Colloquium, ICGI 2004, Athens, Greece, October 11-13, 2004. Proceedings*. Paliouras, G. & Sakakibara, Y. (eds.). Berlin: Springer, p. 151-162 12 p. (Lecture Notes in Artificial Intelligence (LNAI); vol. 3264).

Document clustering and cluster topic extraction in multilingual corpora

Silva, J. F. F., Mexia, J. T. P. N., Coelho, C. A. & Lopes, J. G. P., 1 Jan 2001, *Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM 2001*. Cercone, N., Lin, T. Y. & Wu, X. (eds.). Los Alamitos, California: IEEE Computer Society, p. 513-520 8 p.

Multilingual document clustering, topic extraction and data transformations

Silva, J. F. F., Mexia, J. T. P. N., Coelho, C. A. & Lopes, J. G. P., 1 Jan 2001, *Progress in Artificial Intelligence*. Brazdil, P. & Jorge, A. (eds.). Berlin: Springer, p. 74-87 14 p.

Deep medical knowledge to design clinical guidelines

Barahona, P., Walton, R., Ilic, Z. & Silva, J. F. D., 1 Jan 1995, In: *Computer Methods and Programs in Biomedicine*. 48, 1-2, p. 27-34 8 p.

Awards

BestSupplier: BestSupplier

Marques, N. C. & Silva, J. F. F.
Fundação para a Ciência e a Tecnologia
31/12/11 → 30/07/13

Projects

BestSupplier: BestSupplier

Marques, N. C. & Silva, J. F. F.

Fundação para a Ciência e a Tecnologia

31/12/11 → 30/07/13

ISTRION: ISTRION - Improving Phrase-Based Statistical Machine TRanslation through supervisiON; ISTRION - Melhorando a Tradução Estatística via supervisão

Lopes, J. G. P., Russo, L. M. S. & Silva, J. F. F.

Fundação para a Ciência e a Tecnologia

28/02/11 → 27/02/13

Leonardo-1: Leonardo-1 - New Computer Technologies for Access to Vast Arrays of Multilingual Information

Lopes, J. G. P. & Silva, J. F. F.

Fundação para a Ciência e a Tecnologia

28/02/03 → 25/07/03

ASTROLABIUM: Portuguese Network of Mobility Centres and Portal

Lopes, J. G. P. & Silva, J. F. F.

Fundação para a Ciência e a Tecnologia

31/12/03 → 30/07/06