

# Franco Salvador, Marc

marc.franco@symanto.com

Valencia, 46184  
Valencian Community  
Spain



## Personal information

|             |                           |
|-------------|---------------------------|
| Nationality | Spanish                   |
| Phone       | (+34) 644 26 44 47        |
| Websites    | LinkedIn · Google Scholar |

## Vocational area

R&D in Multilingual Data Science, Analytics, and Mining · Machine and Deep Learning · Information Retrieval · Natural Language Processing

## Professional Experience

*References available upon request*

|                           |  |
|---------------------------|--|
| Since<br>Feb. 2021        | <b>Director of Research</b> at Symanto, Valencia (Spain)   |
| Since<br>April. 2021      | <b>City lead</b> at Valencia AI@Spain AI, Valencia (Spain)   |
| June. 2021 –<br>July 2021 | <b>Collaborator - Big data &amp; science master</b> at Institut de Formació Contínua IL3 - Universitat de Barcelona, Barcelona (Spain) |
| May 2018 –<br>Jan. 2021   | <b>Principal Research Scientist</b> at Symanto, Valencia (Spain)   |
| Apr. 2017 –<br>Apr. 2018  | <b>Innovation Lead</b> at Symanto, Valencia (Spain)  |
| Dec. 2016 –<br>Apr. 2017  | <b>Data scientist</b> at Symanto, Valencia (Spain)   |

|                          |   |
|--------------------------|---|
| Sep. 2016 –<br>Nov. 2016 | <p><b>Research fellow</b> at Data and Web Science Group, Universität Mannheim, Mannheim (Germany)</p> <p>Funded by DAAD (German Academic Exchange Service) in accordance with the DAAD scholarship program “STIBET-Doktoranden”.</p>  |
| June 2016 –<br>Aug. 2016 | <p><b>Data scientist</b> at Symanto, Nuremberg (Germany)</p>  |
| July 2014 –<br>May 2016  | <p><b>Senior technician</b> at Pattern Recognition and Human Language Technology (PRHLT) Research Center - Universitat Politècnica de València, Valencia (Spain)</p> <p>Funded by:</p> <p>The Spanish Ministry of Economy and Competitiveness 3-year R&amp;D project (TIN2015-71147-C2-1-P) SOMEMBED - SOcial Media language understanding - EMBEDing contexts.</p> <p>The Spanish Ministry of Economy and Competitiveness 3-year R&amp;D project (MEC TIN2012-38603-C02-01) DIANA-APPLICATIONS - Finding Hidden Knowledge in Text.</p> |
| Apr. 2014 –<br>June 2014 | <p><b>Technical specialist</b> at Itálica Research Group - Universidad de Sevilla, Seville (Spain)</p> <p>Funded by the Spanish Ministry of Economy and Competitiveness 3.5-year R&amp;D project (TIN2011-14726-E) <i>Destilado de opiniones desde contenidos generados por usuarios</i>.</p>   |
| Apr. 2013 –<br>Mar. 2014 | <p><b>Research fellow</b> at Linguistic Computing Laboratory, Computer Science Department, Sapienza University of Rome, Rome (Italy)</p> <p>Funded by the 5-year ERC Starting Grant No. 259234 MultiJEDI: Multilingual Joint word sEnse DIambiguation.</p>  |
| Aug. 2012 –<br>Dec. 2012 | <p><b>Senior technician</b> at Natural Language Engineering Lab, Department of Information Systems and Computing, Universitat Politècnica de València, Valencia (Spain)</p> <p>Funded by the Spanish Ministry of Education and Science 4-year R&amp;D project (CICYT TIN2009-13391-C04-03) TEXT-ENTERPRISE 2.0: Text comprehension techniques applied to the needs of the Enterprise 2.0.</p>   |

## Education

|           |   |
|-----------|---|
| 2013–2017 | <p><b>Doctor of Philosophy (PhD), Computer Science, Universitat Politècnica de València, Spain</b></p> <p><b>Final score:</b> Excellent with <b>cum laude</b> distinction</p> <p>PhD thesis: A Cross-domain and Cross-language Knowledge-based Representation of Text and its Meaning</p> <p>Advisor: Paolo Rosso</p>                       |
| 2011–2013 | <p><b>Master of Science (MSc), Artificial Intelligence, Pattern Recognition and Digital Image, Universitat Politècnica de València, Spain</b></p> <p><b>Avg. score: 8.7/10</b></p> <p>MSc thesis: Cross-Language Plagiarism Detection using a Multilingual Semantic Network</p> <p>Score: 10/10 - Cum laude</p> <p>Advisor: Paolo Rosso</p> |

|           |   |
|-----------|---|
| 2006–2011 | <b>Bachelor of Science (BSc), Computer Science Engineering (5-year degree), specialization in Artificial Intelligence, Universitat Politècnica de València, Spain</b><br><br><b>Avg. score: 7/10</b><br>BSc final project: Voice-based web surfing<br>Score: 10/10<br>Advisor: Carlos D. Martínez Hinarejos |
|-----------|---|

## Languages

|                |   |
|----------------|---|
| Spanish        | Mother tongue                           |
| Catalan        | Mother tongue                           |
| <b>English</b> | <b>Professional working proficiency</b> |
| Italian        | Basic competence                        |

## Achievements and Awards

|      |   |
|------|---|
| 2021 | First place in several evaluation metrics at the eRisk 2021: Early risk prediction on the Internet shared task@CLEF - subtasks on "Early detection of signs of self-harm" and "Measuring the severity of the signs of depression"                       |
| 2018 | PhD thesis award for the best dissertation in the field of Information Retrieval.<br><br>Awarded by the 5th Spanish Conference in Information Retrieval (CERI 2018).  |
| 2017 | Cum laude PhD thesis: A Cross-domain and Cross-language Knowledge-based Representation of Text and its Meaning.   |
| 2016 | First position in the International Workshop on Semantic Evaluation (SemEval) 2016, task 3: community question answering - subtask B: question-question similarity.   |
| 2013 | Winner of the MAVIR 2013 national prize to the best M.Sc. thesis on Language Technologies applied to Intelligent Systems of Access to the Multilingual and Multimedia Information.<br><br>Awarded by MAVIR Consortium. Sponsored by Córex and DAEDALUS. |
| 2013 | Cum laude MSc thesis: Cross-language Plagiarism Detection using a Multilingual Semantic Network.  |
| 2013 | Third prize winner in the poster session of PROMISE Winter School 2013.   |

## Courses

|           |   |
|-----------|---|
| Oct. 2017 | Bitcoin and Cryptocurrency Technologies.<br>Offered by the Princeton University at Coursera                               |
| Feb. 2013 | PROMISE Winter School 2013 - Bridging between Information Retrieval and Databases.<br>Bressanone, Italy.                  |
| Jul. 2009 | UIMP English Language Immersion Course - Intensive English Course with Native Teachers and Monitors.<br>Barcelona, Spain. |

## Certifications

|           |   |
|-----------|---|
| Dec. 2021 | Digitization applied to the productive sector<br>Confederación Española de Organizaciones Empresariales (CEOE)  |
| Jan. 2016 | Cambridge First Certificate in English (FCE) - B2 level of the Common European Framework of Reference for Languages (CEFR)<br>University of Cambridge |
| Jan. 2016 | Elemental Grade in Valencian - B1 level of the Common European Framework of Reference for Languages (CEFR)<br>Generalitat Valenciana                  |
| Dec. 2015 | Occupational safety: laboratory researcher profile (register number: 15/45480)<br>Universitat Politècnica de València                                 |

## Research Internships

|      |  |
|------|--|
| 2015 | National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico<br>Mentor: Dr. Manuel Montes-y-Gómez<br>Funded by the Web Information Quality Evaluation Initiative (WIQ-EI) IRSES 4-year project (grant no. 269180) within the FP 7 Marie Curie People Framework.<br>Date: Feb. - Mar. 2015  |
| 2014 | International Institute of Information Technology (IIIT), Hyderabad, India<br>*During this internship I have been also research intern at Veooz - URL: <a href="http://www.veooz.com">www.veooz.com</a><br>Mentor: Prof. Dr. Vasudeva Varma<br>Funded by the Web Information Quality Evaluation Initiative (WIQ-EI) IRSES 4-year project (grant no. 269180) within the FP 7 Marie Curie People Framework.<br>Date: Oct - Dec. 2014 |
| 2012 | Computational Research Center (CIC), Mexico D.F., Mexico & National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico<br>Mentor: Prof. Dr. Grigori Sidorov & Dr. Manuel Montes-y-Gómez<br>Funded by the Web Information Quality Evaluation Initiative (WIQ-EI) IRSES 4-year project (grant no. 269180) within the FP 7 Marie Curie People Framework.<br>Date: Nov - Dec. 2012                              |

## Competences

|                                |   |
|--------------------------------|---|
| Artificial Intelligence        | Deep learning, machine learning, data mining, natural language processing, information retrieval, machine translation |
| Data Mining & Machine Learning | TensorFlow, Keras, PyTorch, scikit-learn, huggingface-transformers, Weka  |
| Programming Languages          | Python, Java, C, C++, Bash  |
| Software Development           | Kanban, Scrum, Eclipse, Maven, Visual Studio, Visual Studio Code, Git   |
| Containers & Orchestration     | Docker, Kubernetes  |
| Big Data, DB                   | Spark, SQL, PostgreSQL, Oracle  |

## Publications

- 2021 C. Cohrdes, S. Yenikent, J. Wu, B. Ghanem, M. Franco-Salvador, and F. Vogelgesang. **Indications of Depressive Symptoms During the COVID-19 Pandemic in Germany: Comparison of National Survey and Twitter Data.** In *JMIR Ment Health* 2021;8(6):e27140, DOI: 10.2196/27140 (**Impact Factor: 4.39**)
- 2021 A. Basile, G. Pérez-Torró, and M. Franco-Salvador. **Probabilistic Ensembles of Zero- and Few-Shot Learning Models for Emotion Classification.** In *Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, Bulgaria, September 1-4, pp. 128–137.
- 2021 A. Basile, M. Chineza-Rios, A.S. Uban, T. Muller, L. Rossler, S. Yenikent, B. Chulví, P. Rosso, and M. Franco-Salvador. **UPV-Symanto at eRisk 2021: Mental Health Author Profiling for Early Risk Prediction on the Internet.** In *CLEF 2021 Labs and Workshops, Notebook Papers*. CEUR-WS.org, <http://ceur-ws.org/Vol-2936/paper-75.pdf>.
- 2021 S. Stajner, S. Yenikent, B. Ghanem, and M. Franco-Salvador. **What Motivates You? Benchmarking Automatic Detection of Basic Needs From Short Posts.** In *Proc. of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)*, August 1th through August 6th, 2021, pp. 803–81. (**CORE A**)
- 2021 S. Stajner, S. Yenikent, and M. Franco-Salvador. **Five Psycholinguistic Characteristics for Better Interaction with Users.** In *Proceedings of the 8th IEEE International Conference on Behavioural and Social Computing (BESC)*, November 1, 2021.

- 2020 M. Chinea-Rios, M. Franco-Salvador, and Y. Benajiba. **Aspect On: an Interactive Solution for Post-Editing the Aspect Extraction based on Online Learning**. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France, May 13-16, pp. 4974–4981. (CORE C)
- 2019 A. Basile, M. Franco-Salvador, N. Pawar, S. Štajner, M. Chinea Rios, and Y. Benajiba. **SymantoResearch at SemEval-2019 Task 3: Combined Neural Models for Emotion Classification in Human-Chatbot Conversations**. In *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*, Minneapolis, USA, June 6-7.
- 2018 M. Franco-Salvador, and L. A. Leiva. **Multilingual phrase sampling for text entry evaluations**. In *International Journal of Human-Computer Studies*, vol. 113, pp. 15-31. DOI: 10.1016/j.ijhcs.2018.01.006 (Impact Factor: 2.13)
- 2018 R. M. Giménez-Pérez, M. Franco-Salvador, and P. Rosso. **String Kernels for Polarity Classification: A Study across Different Languages**. In *Proc. of the 23rd International Conference on Natural Language & Information Systems (NLDB 2018)*, Paris, France, 13-15 June, pp. 489–493. (CORE C)
- 2018 M. A. Álvarez Carmona, M. Franco-Salvador, M. Montes-Y-Gómez, P. Rosso, L. Villaseñor-Pineda, and E. Villatoro Tello. **Semantically-informed Distance and Similarity Measures for Paraphrase Plagiarism Identification**. In *Journal of Intelligent and Fuzzy Systems, Special Issue 4th Int. Symposium on Language & Knowledge Engineering*, vol. 34, pp. 2983–2990. DOI: 10.3233/JIFS-169483 (Impact Factor: 1.26)
- 2018 S. Štajner, M. Franco-Salvador, P. Rosso, and S. P. Ponzetto. **CATS: A Tool for Customised Alignment of Text Simplification Corpora**. In *Proc. 11th Int. Conf. on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12. (CORE C)
- 2017 G. Glavaš, M. Franco-Salvador, S. P. Ponzetto, and P. Rosso. **A Resource-light Method for Cross-lingual Semantic Textual Similarity**. In *Knowledge-Based Systems*, vol. 143, pp. 1-9. DOI: 10.1016/j.knosys.2017.11.041 (Impact Factor: 4.53)
- 2017 S. Štajner, M. Franco-Salvador, S. P. Ponzetto, P. Rosso, and H. Stuckenschmidt. **Sentence Alignment Methods for Improving Text Simplification Systems**. In *Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, Vancouver, Canada, July 30th through August 4th, 2017, pp. 97–102. (CORE A)
- 2017 M. Franco-Salvador. **A Cross-domain and Cross-language Knowledge-based Representation of Text and its Meaning**. In *PhD Thesis, Universitat Politècnica de València*. <http://hdl.handle.net/10251/84285>
- 2017 M. Franco-Salvador, N. Plotnikova, N. Pawar, and Y. Benajiba. **Subword-based Deep Averaging Networks for Author Profiling in Social Media - Notebook for PAN at CLEF 2017**. In *CLEF 2017 Labs and Workshops, Notebook Papers*. CEUR-WS.org, <http://ceur-ws.org/Vol-1866/>

- 2017 M. Franco-Salvador, Greg Kondrak, and P. Rosso. **Bridging the Native Language and Language Variety Identification Tasks.** In *Proc. of the 21st International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2017)*, Marseille, France, 6-8 September, 2017, pp. 1554–1561. DOI: 10.1016/j.procs.2017.08.06 (**CORE B**)
- 2017 R. M. Giménez-Pérez, M. Franco-Salvador, and P. Rosso. **Single and Cross-domain Polarity Classification using String Kernels.** In *Proc. of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)*, Valencia, Spain, 3-7 April, 2017, pp. 558-563. (**CORE A**)
- 2016 M. Franco-Salvador, P. Gupta, P. Rosso, and R. E. Banchs. **Cross-language Plagiarism Detection Over Continuous-space- and Knowledge Graph-based Representations of Language.** In *Knowledge-Based Systems*, vol. 111, pp. 87–99. DOI: 10.1016/j.knosys.2016.08.004 (**Impact Factor: 4.53**)
- 2016 M. Franco-Salvador, P. Rosso, and M. Montes-y-Gómez. **A Systematic Study of Knowledge Graph Analysis for Cross-language Plagiarism Detection.** In *Information Processing & Management*, vol. 52(4), pp. 550–570. DOI: 10.1016/j.ipm.2015.12.004 (**Impact Factor: 2.39**)
- 2016 M. Franco-Salvador, S. Kar, T. Solorio, and P. Rosso. **UH-PRHLT at SemEval-2016 Task 3: Combining Lexical and Semantic-based Features for Community Question Answering.** In *Proc. of the 10th International Workshop on Semantic Evaluation, SemEval '16*, San Diego, California. Association for Computational Linguistics.
- 2016 F. Rangel, M. Franco-Salvador, and P. Rosso. **A Low Dimensionality Representation for Language Variety Identification.** In *Proc. of the 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2016)*, Springer, Cham, 2016. pp. 156-169.
- 2015 M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso. **Cross-domain Polarity Classification using a Knowledge-enhanced Meta-classifier.** In *Knowledge-Based Systems*, vol. 86, pp. 46–56. DOI: 10.1016/j.knosys.2015.05.020 (**Impact Factor: 4.53**)
- 2015 M. Franco-Salvador, F. Rangel, P. Rosso, M. Taulé, and A. Martí. **Language Variety Identification using Distributed Representations of Words and Documents.** In *6th International Conference of CLEF on Experimental IR meets Multilinguality, Multimodality, and Interaction (CLEF 2015)*, Springer-Verlag, LNCS(9283), pp. 28–40.
- 2015 M. Franco-Salvador, I. Bensalem, E. Flores, P. Gupta, and P. Rosso. **PAN 2015 Shared Task on Plagiarism Detection: Evaluation of Corpora for Text Alignment.** In *CLEF 2015 Labs and Workshops, Notebook Papers*. CEUR-WS.org, <http://ceur-ws.org/Vol-1391/>
- 2015 M. Franco-Salvador, P. Rosso, and F. Rangel. **Distributed Representations of Words and Documents for Discriminating Similar Languages.** In *Proc. of the RANLP Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects (LT4VarDial)*.

- 2014 M. Franco-Salvador, P. Rosso, and R. Navigli. **A Knowledge-based Representation for Cross-Language Document Retrieval and Categorization.** In *Proc. of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, Gothenburg, Sweden, 26-30 April, 2014, pp. 414–423. **(CORE A)**
- 2014 M. Franco-Salvador, P. Gupta and P. Rosso. **Knowledge Graphs as Context Models: Improving the Detection of Cross-Language Plagiarism with Paraphrasing.** In *Revised Papers PROMISE Winter School 2013*, Springer-Verlag, LNCS(8173), pp. 227–236.
- 2013 M. Franco-Salvador, P. Gupta and P. Rosso. **Graph-Based Similarity Analysis: A New Approach to Cross-Language Plagiarism Detection.** In *Sociedad Española de Procesamiento del Lenguaje Natural (SEPLN)*, num. 50, pp. 21–28.
- 2013 M. Franco-Salvador, P. Gupta and P. Rosso. **Cross-Language Plagiarism Detection using a Multilingual Semantic Network.** In *Proc. of the 35th European Conference on Information Retrieval (ECIR 2013)*, Moscow, Russia, 24-27 March, 2013, Springer-Verlag, LNCS(7814), pp. 710–713. **(CORE B)**
- 2012 M. Franco-Salvador, P. Gupta and P. Rosso. **Cross-language Plagiarism Detection Using BabelNet’s Statistical Dictionary.** In *Computación y Sistemas, Revista Iberoamericana de Computación*, ISSN 1405-5546, vol. 16, num. 4, pp. 383–390.