

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

Education

2013–2017	Doctor of Philosophy (PhD), Computer Science, Universitat Politècnica de València, Spain Final score: Excellent with cum laude distinction PhD thesis: A Cross-domain and Cross-language Knowledge-based Representation of Text and its Meaning Advisor: Paolo Rosso
2011–2013	Master of Science (MSc), Artificial Intelligence, Pattern Recognition and Digital Image, Universitat Politècnica de València, Spain Avg. score: 8.7/10 MSc thesis: Cross-Language Plagiarism Detection using a Multilingual Semantic Network Score: 10/10 - Cum laude Advisor: Paolo Rosso
2006–2011	Bachelor of Science (BSc), Computer Science Engineering (5-year degree), specialization in Artificial Intelligence, Universitat Politècnica de València, Spain Avg. score: 7/10 BSc final project: Voice-based web surfing Score: 10/10 Advisor: Carlos D. Martínez Hinarejos

Professional Experience

Additional references available upon request

Since Feb. 2021	Director of Research at Symanto, Valencia (Spain), https://www.symanto.com/ Reference: Khaleeq Aziz - CEO - e-mail: khaleeq.aziz@symanto.com
Since April. 2021	City lead at Valencia AI@Spain AI, Valencia (Spain), https://www.spain-ai.com/
June. 2021 – July 2021	Collaborator - Big data & science master at Institut de Formació Contínua IL3 - Universitat de Barcelona, Barcelona (Spain), https://www.il3.ub.edu/
May 2018 – Jan. 2021	Principal Research Scientist at Symanto, Valencia (Spain), https://www.symanto.com/ Reference: Dr. Yassine Benajiba, Chief Scientist - e-mail: yassine.benajiba@symanto.net
Apr. 2017 – Apr. 2018	Innovation Lead at Symanto, Valencia (Spain), https://www.symanto.com/ Reference: Dr. Yassine Benajiba, Director of Innovation - e-mail: yassine.benajiba@symanto.net
Dec. 2016 – Apr. 2017	Data scientist at Symanto, Valencia (Spain), https://www.symanto.com/ Reference: Dr. Yassine Benajiba, Director of Innovation - e-mail: yassine.benajiba@symanto.net
Sep. 2016 – Nov. 2016	Research fellow at Data and Web Science Group, Universität Mannheim, Mannheim (Germany), http://dws.informatik.uni-mannheim.de/en/home/ Funded by DAAD (German Academic Exchange Service) in accordance with the DAAD scholarship program “STIBET-Doktoranden”. Reference: Dr. Simone Paolo Ponzetto, Full Professor - e-mail: simone@informatik.uni-mannheim.de
June 2016 – Aug. 2016	Data scientist at Symanto, Nuremberg (Germany), https://www.symanto.com/ Reference: Dr. Yassine Benajiba, Director of Research - e-mail: yassine.benajiba@symanto.net
July 2014 – May 2016	Senior technician at Pattern Recognition and Human Language Technology (PRHLT) Research Center - Universitat Politècnica de València, Valencia (Spain), http://www.prhlt.upv.es/ Funded by: The Spanish Ministry of Economy and Competitiveness 3-year R&D project (TIN2015-71147-C2-1-P) SOMEMBED - Social Media language understanding - EMBEDding contexts. The Spanish Ministry of Economy and Competitiveness 3-year R&D project (MEC TIN2012-38603-C02-01) DIANA-APPLICATIONS - Finding Hidden Knowledge in Text. Reference: Dr. Paolo Rosso, Associate Professor - Phone: (+34) 96 387 7007 Ext.73571 - e-mail: proso@dsic.upv.es URL: users.dsic.upv.es/~proso/
Apr. 2014 – June 2014	Technical specialist at Itálica Research Group - Universidad de Sevilla, Seville (Spain), http://www.lsi.us.es/italica/ Funded by the Spanish Ministry of Economy and Competitiveness 3.5-year R&D project (TIN2011-14726-E) <i>Destilado de opiniones desde contenidos generados por usuarios</i> . Reference: Dr. José Antonio Troyano Jiménez, Associate Professor - Phone: (+34) 95 45 52777 - e-mail: troyano@us.es URL: www.lsi.us.es/~troyano/
Apr. 2013 – Mar. 2014	Research fellow at Linguistic Computing Laboratory, Computer Science Department, Sapienza University of Rome, Rome (Italy), http://lcl.uniroma1.it/index.html Funded by the 5-year ERC Starting Grant No. 259234 MultiJEDI: Multilingual Joint word sEnse Disambiguation. Reference: Dr. Roberto Navigli, Associate Professor - Phone: (+39) 064 925 51 61 - e-mail: navigli@di.uniroma1.it URL: wwwusers.di.uniroma1.it/~navigli/

Aug. 2012 – Dec. 2012	Senior technician at Natural Language Engineering Lab, Department of Information Systems and Computing, Universitat Politècnica de València, Valencia (Spain), http://users.dsic.upv.es/grupos/nle/ Funded by the Spanish Ministry of Education and Science 4-year R&D project (CICYT TIN2009-13391-C04-03) TEXT-ENTERPRISE 2.0: Text comprehension techniques applied to the needs of the Enterprise 2.0. Reference: Dr. Paolo Rosso, Associate Professor - Phone: (+34) 96 387 7007 Ext.73571 - e-mail: proso@dsic.upv.es URL: users.dsic.upv.es/~proso/
--------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Languages

Spanish Catalan English Italian	Mother tongue Mother tongue Professional working proficiency 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. 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. Awarded by MAVIR Consortium. Sponsored by Córrex 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.

Research Internships

2015	<p>National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico</p> <p>Mentor: Dr. Manuel Montes-y-Gómez</p> <p>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.</p> <p>Date: Feb. - Mar. 2015</p>
2014	<p>International Institute of Information Technology (IIIT), Hyderabad, India</p> <p>*During this internship I have been also research intern at Veooz - URL: www.veooz.com</p> <p>Mentor: Prof. Dr. Vasudeva Varma</p> <p>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.</p> <p>Date: Oct - Dec. 2014</p>
2012	<p>Computational Research Center (CIC), Mexico D.F., Mexico & National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico</p> <p>Mentor: Prof. Dr. Grigori Sidorov & Dr. Manuel Montes-y-Gómez</p> <p>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.</p> <p>Date: Nov - Dec. 2012</p>

Tutorials

Dec. 2014	<p>Knowledge Graph-based Natural Language Processing.</p> <p>International Workshop on Data and Text Analytics (DATA 2014), New Delhi, India</p>
-----------	--------------------------------------------------------------------------------------------------------------------------------------------------

Invited talks

Jan. 2016	<p>Knowledge Graph-based Cross-language and Domain Natural Language Processing.</p> <p>Symanto Group, Nuremberg, Germany.</p>
Feb. 2015	<p>Semantic Relatedness using Word Representations in Vector Space.</p> <p>National Institute of Astrophysics, Optics and Electronics (INAOE). Puebla, Mexico.</p>
Jan. 2015	<p>Learning text and word representation models.</p> <p>Language and Computation Center (CLiC), Universitat de Barcelona (UB), Barcelona, Spain</p>
Nov. 2014	<p>Using Knowledge graphs for Natural Language Processing.</p> <p>International Institute of Information Technology (IIIT), Hyderabad, India</p>
Oct. 2014	<p>Knowledge-based Cross-language Natural Language Processing.</p> <p>South Asian University (SAU). New Delhi, India.</p>
Nov. 2012	<p>Graph-based Similarity Analysis: A New Approach to Cross-Language Plagiarism Detection.</p> <p>Computational Research Center (CIC). National Polytechnic Institute. Mexico D.F., Mexico.</p>
Dec. 2012	<p>Cross-language Plagiarism Detection using a Multilingual Semantic Network.</p> <p>National Institute of Astrophysics, Optics and Electronics (INAOE). Puebla, Mexico.</p>

Doctoral Consortium

June 2014	Knowledge-based Cross-language Natural Language Processing. Presented at <i>V Jornadas de la Red Temática en Tratamiento de la Información Multilingüe y Multimodal (TIMM)</i> , Cazalla de la Sierra (Sevilla), Spain
-----------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Editorial Services to Scholarly Publications

Reviewer for several journals including: Artificial Intelligence; Information Sciences; Expert Systems With Applications; Journal of Intelligent and Fuzzy Systems; Language Resources and Evaluation; Natural Language Engineering.

Organising Committees

I have been member of the following Organising committees: Publication Co-Chair at *SEM 2015 (Joint Conference on Lexical and Computational Semantics).

Program Committees

I served as member of the following program and scientific committees: ACL 2018 (Annual Meeting of the Association for Computational Linguistics), ACL 2017, IberEval@SEPLN 2018, TextGraphs@ACL 2017, KES 2017 (International Conference on Knowledge-Based and Intelligent Information & Engineering Systems), SemEval 2016 Task 3 - Community Question Answering (International Workshop on Semantic Evaluation 2016), TIR@DEXA 2017 (International Workshop on Text-based Information Retrieval), TIR@DEXA 2016, TIR@DEXA 2015, TIR@DEXA 2014, DATA-CCIP 2015 (Special Session on DATA in International Conference on Cognitive Computing and Information Processing), SEPLN 2015 (International Workshop on Embeddings and Semantics at Spanish Conference on NLP 2015), DATA 2014 (International Workshop on Data and Text Analytics) .

Courses

Feb. 2013	PROMISE Winter School 2013 - Bridging between Information Retrieval and Databases. Bressanone, Italy. http://www.promise-noe.eu/events/winter-school-2013/
Jul. 2009	UIMP English Language Immersion Course - Intensive English Course with Native Teachers and Monitors. Barcelona, Spain. http://www.uimp.es/immersion/immersion.html

Certifications

Dec. 2021	Digitization applied to the productive sector 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) University of Cambridge
Jan. 2016	Elemental Grade in Valencian - B1 level of the Common European Framework of Reference for Languages (CEFR) Generalitat Valenciana
Dec. 2015	Occupational safety: laboratory researcher profile (register number: 15/45480) Universitat Politècnica de València

Competences

Artificial Intelligence	Deep learning, machine learning, data mining, natural language processing, information retrieval, machine translation
Data Mining & Machine Learning	TensorFlow, Keras, PyTorch, Weka, Gensim, Deeplearning4j, Theano
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. n/a.
- 2021 A. Basile, M. Chinea-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**)
- 2020 S. Stajner, S. Yenikent, and M. Franco-Salvador. **Benchmarking automatic detection of psycholinguistic characteristics for better human-computer interaction.** In *arXiv preprint arXiv:2012.09692*.
- 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.
- 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.