

# Franco Salvador, Marc

marc.franco@symanto.net

Valencia, 46009  
Valencian Community  
Spain



## Personal information

Nationality	Spanish
Date of birth	06/05/1988
Phone	(+34) 644 26 44 47, (+49) 01573 0342753
Websites	LinkedIn · Google Scholar

## Vocational area

R&D in Multilingual Data Science, Analytics, and Mining · Information Retrieval  
· Natural Language Processing · Computational Linguistics

## Education

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

## Professional Experience

Since Dec. 2016	<p><b>Data scientist</b> at Symanto Group, Valencia (Spain), <a href="https://www.symanto.net/">https://www.symanto.net/</a></p> <p>Reference: Dr. Yassine Benajiba, Director of Innovation - e-mail: <a href="mailto:yassine.benajiba@symanto.net">yassine.benajiba@symanto.net</a></p>
Sep. 2016 – Nov. 2016	<p><b>Research staff member</b> at Pattern Recognition and Human Language Technology (PRHLT) Research Center - Universitat Politècnica de València, Valencia (Spain), <a href="http://www.prhlt.upv.es/">http://www.prhlt.upv.es/</a></p> <p>Founded by 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>Reference: Dr. Paolo Rosso, Associate Professor - Phone: (+34) 96 387 7007 Ext.73571 - e-mail: <a href="mailto:proso@dsic.upv.es">proso@dsic.upv.es</a> URL: <a href="http://users.dsic.upv.es/~proso/">users.dsic.upv.es/~proso/</a></p>
June 2016 – Aug. 2016	<p><b>Data scientist</b> at Symanto Group, Nuremberg (Germany), <a href="https://www.symanto.net/">https://www.symanto.net/</a></p> <p>Reference: Dr. Yassine Benajiba, Director of Research - e-mail: <a href="mailto:yassine.benajiba@symanto.net">yassine.benajiba@symanto.net</a></p>
July 2014 – May 2016	<p><b>Research staff member</b> at Pattern Recognition and Human Language Technology (PRHLT) Research Center - Universitat Politècnica de València, Valencia (Spain), <a href="http://www.prhlt.upv.es/">http://www.prhlt.upv.es/</a></p> <p>Founded 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> <p>Reference: Dr. Paolo Rosso, Associate Professor - Phone: (+34) 96 387 7007 Ext.73571 - e-mail: <a href="mailto:proso@dsic.upv.es">proso@dsic.upv.es</a> URL: <a href="http://users.dsic.upv.es/~proso/">users.dsic.upv.es/~proso/</a></p>
Apr. 2014 – June 2014	<p><b>Research staff member</b> at Itálica Research Group - Universidad de Sevilla, Seville (Spain), <a href="http://www.lsi.us.es/italica/">http://www.lsi.us.es/italica/</a></p> <p>Founded 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> <p>Reference: Dr. José Antonio Troyano Jiménez, Associate Professor - Phone: (+34) 95 45 52777 - e-mail: <a href="mailto:troyano@us.es">troyano@us.es</a> URL: <a href="http://www.lsi.us.es/~troyano/">www.lsi.us.es/~troyano/</a></p>
Apr. 2013 – Mar. 2014	<p><b>Research fellow</b> at Linguistic Computing Laboratory, Computer Science Department, Sapienza University of Rome, Rome (Italy), <a href="http://lcl.uniroma1.it/index.html">http://lcl.uniroma1.it/index.html</a></p> <p>Founded by the 5-year ERC Starting Grant No. 259234 MultiJEDI: Multilingual Joint word sEnse Disambiguation.</p> <p>Reference: Dr. Roberto Navigli, Associate Professor - Phone: (+39) 064 925 51 61 - e-mail: <a href="mailto:navigli@di.uniroma1.it">navigli@di.uniroma1.it</a> URL: <a href="http://wwwusers.di.uniroma1.it/~navigli/">wwwusers.di.uniroma1.it/~navigli/</a></p>
Aug. 2012 – Dec. 2012	<p><b>Research staff member</b> at Natural Language Engineering Lab, Department of Information Systems and Computing, Universitat Politècnica de València, Valencia (Spain), <a href="http://users.dsic.upv.es/grupos/nle/">http://users.dsic.upv.es/grupos/nle/</a></p> <p>Founded 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> <p>Reference: Dr. Paolo Rosso, Associate Professor - Phone: (+34) 96 387 7007 Ext.73571 - e-mail: <a href="mailto:proso@dsic.upv.es">proso@dsic.upv.es</a> URL: <a href="http://users.dsic.upv.es/~proso/">users.dsic.upv.es/~proso/</a></p>

## Languages

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

## Achievements and Awards

- |      |   |
|------|---|
| 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 | M.Sc. thesis Cross-language Plagiarism Detection using a Multilingual Semantic Network, cum laude.  |
| 2013 | Third prize winner in the poster session of PROMISE Winter School 2013.   |

## Publications

- |      |  |
|------|--|
| 2017 | Rosa M. Giménez-Pérez, M. Franco-Salvador, and P. Rosso. <b>Single and Cross-domain Polarity Classification using String Kernels</b> . In <i>Proc. of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)</i> , Valencia, Spain, 3-7 April, 2017, pp. n/a. <b>(CORE A)</b>                  |
| 2016 | M. Franco-Salvador, P. Gupta, P. Rosso, and R. E. Banchs. <b>Cross-language Plagiarism Detection Over Continuous-space- and Knowledge Graph-based Representations of Language</b> . In <i>Knowledge-Based Systems</i> , vol. 111, pp. 87-99. DOI: 10.1016/j.knosys.2016.08.004 <b>(Impact Factor: 2.92)</b>                                      |
| 2016 | M. Franco-Salvador, P. Rosso, and M. Montes-y-Gómez. <b>A Systematic Study of Knowledge Graph Analysis for Cross-language Plagiarism Detection</b> . In <i>Information Processing &amp; Management</i> , vol. 52(4), pp. 550-570. DOI: 10.1016/j.ipm.2015.12.004 <b>(Impact Factor: 1.26)</b>  |
| 2016 | M. Franco-Salvador, S. Kar, T. Solorio, and P. Rosso. <b>UH-PRHLT at SemEval-2016 Task 3: Combining Lexical and Semantic-based Features for Community Question Answering</b> . In <i>Proc. of the 10th International Workshop on Semantic Evaluation, SemEval '16</i> , San Diego, California. Association for Computational Linguistics.        |
| 2016 | F. Rangel, M. Franco-Salvador, and P. Rosso. <b>A Low Dimensionality Representation for Language Variety Identification</b> . In <i>Proc. of the 17th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2016)</i> , Springer-Verlag, LNCS(n/a), pp. n/a.  |
| 2015 | M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso. <b>Cross-domain Polarity Classification using a Knowledge-enhanced Meta-classifier</b> . In <i>Knowledge-Based Systems</i> , vol. 86, pp. 46-56. DOI: 10.1016/j.knosys.2015.05.020 <b>(Impact Factor: 2.92)</b>   |
| 2015 | M. Franco-Salvador, F. Rangel, P. Rosso, M. Taulé, and A. Martí. <b>Language Variety Identification using Distributed Representations of Words and Documents</b> . In <i>6th International Conference of CLEF on Experimental IR meets Multilinguality, Multimodality, and Interaction (CLEF 2015)</i> , 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.

## Research Internships

- 2015 National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico  
Mentor: Dr. Manuel Montes-y-Gómez  
Founded by the Web Information Quality Evaluation Initiative (WIQ-EI) IRSES 4-year project (grant no. 269180) within the FP 7 Marie Curie People Framework.  
Date: Feb. - Mar. 2015
- 2014 International Institute of Information Technology (IIIT), Hyderabad, India  
\*During this internship I have been also research intern at Veooz - URL: [www.veooz.com](http://www.veooz.com)  
Mentor: Prof. Dr. Vasudeva Varma  
Founded by the Web Information Quality Evaluation Initiative (WIQ-EI) IRSES 4-year project (grant no. 269180) within the FP 7 Marie Curie People Framework.  
Date: Oct - Dec. 2014

2012	<p>Computational Research Center (CIC), Mexico D.F., Mexico &amp; National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico</p> <p>Mentor: Prof. Dr. Grigori Sidorov &amp; Dr. Manuel Montes-y-Gómez</p> <p>Founded 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	<p>Knowledge-based Cross-language Natural Language Processing.</p> <p>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</p>
-----------	---

## 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 2017 (Annual Meeting of the Association for Computational Linguistics), TextGraphs@ACL 2017, 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. <a href="http://www.promise-noe.eu/events/winter-school-2013/">http://www.promise-noe.eu/events/winter-school-2013/</a>
Jul. 2009	UIMP English Language Immersion Course - Intensive English Course with Native Teachers and Monitors. Barcelona, Spain. <a href="http://www.uimp.es/immersion/immersion.html">http://www.uimp.es/immersion/immersion.html</a>

## Certifications

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	Machine learning, data mining, natural language processing, information retrieval, deep learning, machine translation
Data Mining & Machine Learning	Weka, Gensim, TensorFlow, Deeplearning4j, Torch
Programming Languages	C, C++, Java, Python, Bash, Haskell, Prolog
Software Development	Eclipse, Visual Studio, Maven, Git, Subversion
Databases	SQL, Oracle