Enrique Flores Saez

Work experience

2012–2016   **Researcher**, *Universitat Politècnica de València (UPV)*, Spain.
- Research focused on source code monolingual and cross-lingual re-use detection.
- Source code monolingual and cross-lingual retrieval.
- Developer of source code re-use detection systems inspired in Natural Language Processing techniques
- Organiser of two editions of source code re-use detection competitions (see `http://goo.gl/ZcriU9` and `http://goo.gl/PPbtsy`)
- Participation in the project: "TEXT-ENTERPRISE 2.0: Text comprehension techniques applied to the needs of the Enterprise 2.0" collecting corpora and developing tagging tools for source code re-use detection research purposes.
- To see the list of publications related to my research work, please visit my Google Scholar at `https://goo.gl/Rhq1nz`

Feb–May 2015   **Visitor Researcher**, *INAOE, Mexico*, Research Group: Lab. of Language Technologies, Participation in the project: WIQ-EI project.

Jan 2015   **Visitor Researcher**, *Center for Computing Research, National Polytechnic Institute, Mexico*, Research Group: Lab. of NLP, Participation in the project: WIQ-EI project.

Aug–Dec 2012   **Visitor Researcher**, *Center for Computing Research, National Polytechnic Institute, Mexico*, Research Group: Lab. of NLP, Participation in the project: WIQ-EI project.

Jun–Nov 2010   **Software Engineer**, *Dept. of Computer Systems and Computation*, Universitat Politècnica de València, Participation in project: Language Technologies applied to systems modelling Information (TeLMoSis).

The main objective of TeLMoSis project is to combine advances in Human Language Technologies, particularly semantic analysis, with information system modelling.

Research proyects

Oct–Dec 2015   **Senior Technician**, *PRHLT Research Center*, Universitat Politècnica de València, JITL: Automatic internationalisation for locating JIT web interfaces, Ministry of Education, Culture and Sport (MECD) project SP20140775.

Jan–May 2015   **Senior Technician**, *PRHLT Research Center*, Universitat Politècnica de València, DIANA-APPLICATIONS - Finding Hidden Knowledge in Texts: Applications project: TIN2012-38603-C02-01.

2010–2014   **Government Scholarship Researcher Training**, *Spanish Ministry of Science and Innovation*, Universitat Politècnica de València, TEXT-ENTERPRISE 2.0: Text comprehension techniques applied to the needs of the Enterprise 2.0 project: TIN2009-13391-C04-03.

2011–2015	**WIQ-Ei Project**, *Web Information Quality Evaluation Initiative*, Universitat Politècnica de València, IRSES project (grant no. 269180) within the FP7 Marie Curie People Framework.
Dec–Mar 2011–2013	**Bilateral project**, *Applications of Natural Language Processing to the Need of the University*, UPV, AECID-PCI AP/043848/11.
Jun–Nov 2010	**Scholarship expertise**, *Dept. of Computer Systems and Computation*, Universitat Politècnica de València, The main objective of TeLMoSis project is to combine advances in Human Language Technologies, particularly semantic analysis, with information system modelling..

Teaching experience

Jan–Jun 2014	**Subject: Information storage and retrieval systems**, *School of Engineering in Computer Science*, Universitat Politècnica de València, Teaching Assistant.
Jan–Jun 2013	**Subject: Information storage and retrieval systems**, *School of Engineering in Computer Science*, Universitat Politècnica de València, Teaching Assistant.
Mar 2013	**Plagiarism and re-use detection Laboratory**, *MISC laboratory*, University of Constantine 2, 1 week.
Apr 2013	**Plagiarism detection Laboratory**, *idEC, Master in Forensic Linguistics*, IULA-Universitat Pompeu Fabra, 4h.

Education

Feb 2012 – May 2016	**PhD in Computer Science**, *Universitat Politècnica de València*. Title of the thesis: Mono- and Cross-lingual Source Code Re-use Detection. Research stays at the National Polytechnic Institute of Mexico and National Institute of Astrophysics Optics and Electronics (Mexico) (Aug - Dec 2012 and Jan - Apr 2015).
Master's Thesis	*Source code re-use across programming languages.*In this work are proposed two models for source code re-use detection and their performance at mono- and cross-lingual level. Average Grade: 9.5 out of 10. Specialisation in Pattern Recognition and Language Technologies.
2009–2012	**Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging**, *Universitat Politècnica de València*.
CS Final Project	*Development of a system for managing collaborative work*. Web application for managing tasks and questions in a workgroup. It allows different types of surveys, mailing lists, ratings and comments.
2004–2010	**Computer Science Degree (5-years degree)**, *School of Engineering in Computer Science*, Universitat Politècnica de València, *Computer Science*. Specialisation in Artificial Intelligence and Pattern Recognition.

Lectures and talks

| School of Engineering in Computer Science | **Universitat Politècnica de València**, *Computers and code: The coding world*, Jun 2015. |

School of Engineering in Computer Science	**Universitat Politècnica de València**, *Can we automatically detect plagiarism?*, Jun 2014.
Mexico city, Mexico	**Universidad Autónoma Metropolitana**, *Talk: "Source Code Re-use Plagiarism Detection"*, 30 Jan 2015.
Mexico city, Mexico	**Universidad Autónoma Metropolitana**, *Talk: "PAN@FIRE: Detection of Source Code re-use (SOCO)"*, 30 Jan 2015.
Toluca, Mexico	**Universidad Autónoma del Estado México**, *Talk: "Source Code Re-use Plagiarism Detection"*, 23 Jan 2015.
Toluca, Mexico	**Universidad Autónoma del Estado México**, *Talk: "PAN@FIRE: Detection of Source Code re-use (SOCO)"*, 23 Jan 2015.
Sevilla	**V Jornadas Tratamiento de la Información Multilingüe y Multimodal (TIMM)**, *Doctoral Consortium: "Cross-Language Source Code Re-use Detection"*, Jun 2014.
Doctoral Consortium	**Universitat Politècnica de València**, *Computers and code: The coding world*, Jun 2015.
Puebla, Mexico	**Instituto Nacional de Astrofísica, Óptica y Electrónica**, *Talk: "Cross-Language Source Code Re-use"*, 6 Dec 2012.
Mexico city, Mexico	**University of London**, *Talk: "Detection of Cross-Language Source Code Re-use"*, 9 Oct 2012.
Puebla, Mexico	**Instituto Nacional de Astrofísica, Óptica y Electrónica**, *Talk: "Cross-Language Source Code Re-use"*, 6 Dec 2012.
Puebla, Mexico	**Instituto Nacional de Astrofísica, Óptica y Electrónica**, *9th Workshop of Human Language Technologies*, 18-19 Oct 2012.
Constantine, Algeria	**University of Constantine 2**, *Talk: "Plagiarism and re-use detection"*, 2 Mar 2012.

Events and conferences

*SEM 2015	**Publication Co-Chair**	*The fourth joint conference on lexical and computational semantics*
CLEF 2013	**Member of Local organising committee**	*Conference and Labs of the Evaluation Forum*
CERI 2012	**Member of organising committee**	*2nd Spanish Conference on Information Retrieval*
SWERC 2013	**Judge Committee Member**	*Southwestern Europe Regional Contest*
SWERC 2012	**Judge Committee Member**	*Southwestern Europe Regional Contest*
CLOCALPROG 2013	**Judge Committee Member**	*Local programming competition - UPV*

CLOCALPROG **Judge Committee Member** *Local programming competition - UPV*
2012

Languages

Spanish **Mother tongue**
Valencian **Mother tongue**
English **Intermediate level**

Programming skills

Languages Java, C++, Python, Unix Shell, GNU Octave, Matlab, LISP, MySQL, HTML, XML, Haskell, Prolog
Software Lucene, Weka, NLTK, Apache, Tomcat, SVN

Publications

E. Flores, P. Rosso, E. Villatoro-Tello, L. Moreno, R. Alcover, and V. Chirivella, "PAN@FIRE: Overview of CL-SOCO track on the Detection of Cross-Language SOurce COde Re-use," in *FIRE 2015 Working Notes. Seventh International Workshop of the Forum for Information Retrieval Evaluation, Gandhinagar, India, 4–6 December*, 2015.

E. Flores, P. Rosso, L. Moreno, and E. Villatoro-Tello, "On the detection of source code re-use," in *Proceedings of the Forum for Information Retrieval Evaluation*, FIRE '14, (New York, NY, USA), pp. 21–30, ACM, 2015.

E. Flores, L. Moreno, and P. Rosso, "Detecting source code re-use with ensemble models," in *Proceedings of the 4th Spanish Conference on Information Retrieval (CERI) (accepted)*, 2016.

E. Flores, A. Barrón-Cedeño, L. Moreno, and P. Rosso, "Cross-Language Source Code Re-Use Detection Using Latent Semantic Analysis," vol. 21, no. 13, pp. 1708–1725, 2015.

E. Flores, A. Barrón-Cedeño, L. Moreno, and P. Rosso, "Uncovering source code reuse in large-scale academic environments," *Computer Applications in Engineering Education*, vol. 23, no. 3, pp. 383–390, 2015.

E. Flores, P. Rosso, L. Moreno, and E. Villatoro-Tello, "PAN@FIRE: Overview of SOCO track on the detection of SOurce COde re-use," in *Sixth Forum for Information Retrieval Evaluation (FIRE 2014), Bangalore, India*, 2014.

E. Flores, A. Barrón-Cedeño, L. Moreno, and P. Rosso, "Cross-language source code re-use detection," in *Proc. 3rd Spanish Conf. on Information Retrieval, CERI-2014*, pp. 145–156, 2014.

E. Flores, M. Ibarra-Romero, L. Moreno, G. Sidorov, and P. Rosso, "Modelos de recuperación de información basados en n-gramas aplicados a la reutilización

de código fuente," in *Proc. 3rd Spanish Conf. on Information Retrieval, CERI-2014*, pp. 185–188, 2014.

E. Flores, A. Barrón-Cedeño, P. Rosso, and L. Moreno, "DeSoCoRe: Detecting Source Code Re-use across programming languages," in *Proceedings of the Demonstration Session at the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Montréal, Canada), pp. 1–4, Association for Computational Linguistics, June 2012.

E. Flores, A. Barrón-Cedeño, P. Rosso, and L. Moreno, "Towards the detection of cross-language source code reuse," in *Natural Language Processing and Information Systems*, vol. 6716 of *Lecture Notes in Computer Science*, pp. 250–253, Springer, 2011.

E. Flores, A. Barrón-Cedeno, P. Rosso, and L. Moreno, "Detección de reuso de código fuente entre lenguajes de programación con base en la frecuencia de términos," *IV Jornadas TIMM Tratamiento de la Información Multilingüe y Multimodal 7 y 8 de abril*, p. 21, 2011.

E. Flores, A. Barrón-Cedeño, P. Rosso, and L. Moreno, "Detecting Source Code Reuse across Programming Languages," *27th Conference of the Spanish Society for Natural Language Processing (SEPLN'11)*, 2011.

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 *Working Notes Papers of the CLEF 2015 Evaluation Labs*, vol. 1391 of *CEUR Workshop Proceedings*, CLEF and CEUR-WS.org, 2015.