

PARTE A. DATOS PERSONALES

Fecha del CVA	Junio 2022
----------------------	------------

Nombre	Rafael		
Apellidos	Corchuelo Gil		
Sexo		Fecha de nacimiento	
DNI, NIE, pasaporte			
Dirección email		Web	
Open Researcher and Contributor ID (ORCID)			

A.1. Situación profesional actual

Organismo	Universidad de Sevilla		
Dpto./Centro	Lenguajes y Sistemas Informáticos		
Dirección	ETSI Informática, Avda. Reina Mercedes, s/n, Sevilla E-41012, Spain		
Teléfono		Correo electrónico	
Categoría profesional	Titular de Universidad	Fecha inicio	2002
Espec. cód. UNESCO	120317 (Informática), 120318 (Sistemas de Información)		
Palabras clave	Web data science; Information extraction; Enterprise integration.		

A.2. Situación profesional anterior

Período	Puesto/Institución/País/Motivo interrupción
1996-2002	Profesor asociado, Universidad de Sevilla
1994-1996	Becario de investigación, Universidad de Sevilla

PARTE B. RESUMEN DEL CURRÍCULUM

En los últimos diez años, el trabajo de investigación del Dr. Corchuelo ha estado en la frontera entre los sistemas de información y la inteligencia artificial, con aplicaciones en el campo de la integración de aplicaciones e información empresarial. Sus resultados de investigación han dado lugar a 56 artículos en revistas del índice JCR y más 70 ponencias en conferencias. A continuación, se muestra el informe SciVal derivado de dichas publicaciones (el primer valor hace referencia al indicador del Dr. Corchuelo y el segundo al indicador medio a nivel mundial):

	Impacto normalizado	Artículos en revistas Top-10%	Colaboraciones internacionales
Sistemas de Información	0.88 / 1.10	23.80% / 13.50%	33.30% / 19.00%
Inteligencia Artificial	0.66 / 1.07	20.00% / 17.90%	21.40% / 17.40%
Aplicaciones	1.45 / 1.00	80.00% / 24.40%	14.30% / 18.90%

Es de destacar que el impacto normalizado se encuentra cerca de la media del nivel mundial en sistemas de información y alcanza la mitad de la media en inteligencia artificial, pero lo supera ampliamente en el campo de las aplicaciones. También es de destacar que su porcentaje de publicaciones en revistas top 10% es significativamente superior a la media mundial, al igual que su porcentaje de colaboración internacional, salvo en el campo particular de las aplicaciones dado que casi todos sus trabajos en dicho campo son en colaboración con la industria local. Estos buenos resultados se han obtenido en el contexto de una docena de proyectos o complementos públicos de I+D, una quincena de proyectos industriales de I+D+i, y también ha apoyado directamente la creación de cuatro empresas por parte de sus estudiantes.

PARTE C. APORTACIONES RELEVANTES EN LOS ÚLTIMOS DIEZ AÑOS

C.1. Publicaciones

1. On the design of an advanced business rule engine. Patricia Jiménez, Rafael Corchuelo. Softw. Pract. Exp, 2022. (JCR=2.03, Q2). (In press).

2. A coral-reef approach to extract information from HTML tables. Patricia Jiménez, Juan C. Roldán, Rafael Corchuelo. *Appl. Soft Comput.* 115:107980, 2022. (JCR=6.73, Q1)
3. An experimental study of neural approaches to multi-hop inference in question answering. Patricia Jiménez, Rafael Corchuelo. *Int. J. Neural Syst.* 32(4):2250011:1-2250011:19, 2022. (JCR=5.87, Q1)
4. On validating web information extraction proposals. Patricia Jiménez, Rafael Corchuelo. *Expert Syst. Appl.* 199:116700, 2022. (JCR=6.95, Q1)
5. On exploring data lakes by finding compact, isolated clusters. Patricia Jiménez, Juan C. Roldán, Rafael Corchuelo. *Inf. Sci.* 591: 103-127, 2022. (JCR=6.80, Q1)
6. A hybrid quantum approach to leveraging data from HTML tables. Patricia Jiménez, Juan C. Roldán, Rafael Corchuelo. *Knowl. Inf. Syst.* 64(2): 441-474, 2022. (JCR=2.83, Q2)
7. On learning context-aware rules to link RDF datasets. Andrea Cimmino, Rafael Corchuelo. *Log. J. IGPL* 29(2): 151-166, 2021. (JCR= 0.86, Q1)
8. A clustering approach to extract data from HTML tables. Patricia Jiménez, Juan C. Roldán, Rafael Corchuelo. *Inf. Process. Manag.* 58(6):102683, 2021. (JCR=6.22, Q1)
9. TOMATE: A heuristic-based approach to extract data from HTML tables. Juan C. Roldán, Patricia Jiménez, Pedro A. Szekely, Rafael Corchuelo. *Inf. Sci.* 577: 49-68, 2021. (JCR=6.80, Q1)
10. On the synthesis of metadata tags for HTML files. Patricia Jiménez, Juan C. Roldán, Fernando O. Gallego, Rafael Corchuelo. *Softw. Pract. Exp.* 50(12): 2169-2192, 2020. (JCR=2.03, Q2)
11. An encoder-decoder approach to mine conditions for engineering textual data. F.O. Gallego, R. Corchuelo. *Eng. Appl. Artif. Intell.* 91:103568, 2020. (JCR=6.21, Q1)
12. A scheduler for SCADA-based multi-source fusion systems. R. Corchuelo, M. Toro. *Inf. Fusion* 63:41-55, 2020. (JCR=12.98, Q1)
13. On extracting data from tables that are encoded using HTML. J.C. Roldán, P. Jiménez, R. Corchuelo. *Knowl.-Based Syst.* 190:105157, 2020. (JCR=8.04, Q1)
14. A deep-learning approach to mining conditions. F.O. Gallego, R. Corchuelo. *Knowl.-Based Syst.* 193:105422, 2020. (JCR=8.04, Q1)
15. Torii: an aspect-based sentiment analysis system that can mine conditions. F.O. Gallego, R. Corchuelo. *Softw. Pract. Exp.* 50(1): 47-64, 2020. (JCR=2.03, Q2)
16. On learning web information extraction rules with TANGO. Patricia Jiménez, Rafael Corchuelo. *Inf. Syst.* 62: 74-103, 2016. (JCR=7.45, Q1)
17. On the design of a maintainable software development kit to implement integration solutions. Rafael Z. Frantz, Rafael Corchuelo, Fabricia Roos-Frantz. *J. Syst. Softw.* 111: 89-104, 2016. (JCR=2.83, Q1)
18. Classifying links automatically based on their URL. I. Hernández, C.R. Rivero, D. Ruiz, R. Corchuelo. *J. Syst. Softw.* 115:130-143, 2016. (JCR=2.83, Q1)
19. Roller: a novel approach to Web information extraction. Patricia Jiménez, Rafael Corchuelo. *Knowl. Inf. Syst.* 49(1): 197-241, 2016. (JCR=7.45, Q1)
20. Mapping RDF knowledge bases using exchange samples. C.R. Rivero, I. Hernández, D. Ruiz, R. Corchuelo. *Knowl.-Based Syst.* 93: 47-66, 2016. (JCR=8.04, Q1)
21. ARIEX: automated ranking of information extractors. P. Jiménez, R. Corchuelo, H.A. Sleiman. *Knowl.-Based Syst.* 93: 84-108, 2016. (JCR=8.04, Q1)
22. Discovering and analysing ontological models from big RDF data. Carlos R. Rivero, Inma Hernández, David Ruiz, Rafael Corchuelo. *J. Database Manag.* 26(2): 48-61, 2015 (JCR=1.23, Q4)
23. MostoDEx: a tool to exchange RDF data using exchange samples. C.R. Rivero, I. Hernández, D. Ruiz, R. Corchuelo. *J. Syst. Softw.* 100:67-79, 2015. (JCR=2.83, Q1)
24. A class of neural-network-based transducers for web information extraction. Hassan A. Sleiman, Rafael Corchuelo. *Neurocomputing* 135: 61-68, 2014.(JCR=5.72, Q1)
25. CALA: an unsupervised URL-based web page classification system. Inma Hernández, Carlos R. Rivero, David Ruiz, Rafael Corchuelo. *Knowl.-Based Syst.* 57: 168-180, 2014. (JCR=8.04, Q1)
26. Trinity: on using trinary trees for unsupervised web data extraction. Hassan A. Sleiman, Rafael Corchuelo. *IEEE Trans. Knowl. Data Eng.* 26(6): 1544-1556, 2014. (JCR=6.98, Q1)
27. MostoDE: a tool to exchange data amongst semantic-web ontologies. Carlos R. Rivero, Inma Hernández, David Ruiz, Rafael Corchuelo. *J. Syst. Softw.* 86(6): 1517-1529, 2013. (JCR=2.83, Q1)
28. Exchanging data amongst linked data applications. Carlos R. Rivero, Inma Hernández, David Ruiz, Rafael Corchuelo. *Knowl. Inf. Syst.* 37(3): 693-729, 2013. (JCR=2.82, Q2)
29. TEX: An efficient and effective unsupervised Web information extractor. Hassan A. Sleiman, Rafael Corchuelo. *Knowl.-Based Syst.* 39: 109-123, 2013. (JCR=8.04, Q1)

30. A survey on region extractors from web documents. Hassan A. Sleiman, Rafael Corchuelo. IEEE Trans. Knowl. Data Eng. 25(9): 1960-1981, 2013. (JCR=6.98, Q1)
31. Benchmarking data exchange amongst semantic-web ontologies. Carlos R. Rivero, Inma Hernández, David Ruiz, Rafael Corchuelo. IEEE Trans. Knowl. Data Eng. 25(9): 1997-2009, 2013. (JCR=6.98, Q1)
32. A proposal to detect errors in Enterprise Application Integration solutions. Rafael Z. Frantz, Rafael Corchuelo, Carlos Molina-Jiménez. J. Syst. Softw. 85(3): 480-497, 2012. (JCR=2.83, Q1)
33. A bargaining-specific architecture for supporting automated service agreement negotiation systems. Manuel Resinas, Pablo Fernandez, Rafael Corchuelo, Sci. Comput. Program. 77(1): 4-28, 2012. (JCR=0.86, Q4)

C.2. Investigador principal en proyectos de I+D

1. P18-RT-1060. "Ingeniería de datos aplicada a la extracción, semantización, refinamiento y explotación de grafos de conocimiento a escala web". R. Corchuelo, D. Ruiz. Junta de Andalucía. 01/2020-12/2022. 116.042,00€.
2. TIN2016-75394-R. "Herramientas para la ciencia de los datos de la Web". R. Corchuelo, D. Ruiz. MINECO. 01/2017-12/2019. 76.200,00€.
3. TIN2013-40848-R. "Semantización y publicación de datos abiertos para la integración de servicios electrónicos". R. Corchuelo, D. Ruiz. MINECO. 01/2014-12/2016. 142.375,86€.
4. TIN2010-21744-C02-01. "Integración de aplicaciones y datos en la Web". R. Corchuelo. MINECO. 01/2011-12/2013. 80.828,00€.

C.3. Investigador principal en contratos de I+D

5. 103/17-SP. "Desarrollo de la iniciativa Smart Digital Segovia", P. Jiménez, R. Corchuelo. Universidad de Sevilla, SDOS, S.L., Dinamic Area. S.L. 2018-2019. 172.886,70€.
6. P030-15/E10. "Smart Retail Ommichannel". R. Corchuelo. Universidad de Sevilla, Tier1, S.A., Dinamic Area, S.L. 2015-2016. 98.000,00€.
7. P007-15/E10. "Desarrollo la tecnología de integración Guaraná". R. Corchuelo. Universidad de Sevilla, i2Factory, S.L. 2015-2016. 15.066,72€
8. P028-15/E10. "Desarrollo de una metodología de gestión por procesos y gestión documental para empresas manufactureras". R. Corchuelo. Universidad de Sevilla, Tier1, S.A., Mc Puar, S.A., Dinamic Area, S.L. 2015-2017. 90.500,00€.
9. P033-13/E10. "Paradigmas de innovación, social y colaborativo aplicados a Ingeniería del Software". R. Corchuelo. Universidad de Sevilla, Indra, S.A., Dinamic Area, S.L. 2013-2015. 235.000,00€.
10. P010-12/E10. "Inteligencia empresarial mediante técnicas heurísticas". R. Corchuelo. Universidad de Sevilla, Plan3, S.L. Dinamic Area, S.L. 2012-2012. 53.240,00€.
11. P096-11/E10. "Soluciones avanzadas de eficiencia energética y económica en el Mercado Eléctrico Español". R. Corchuelo. Universidad de Sevilla, i2Factory, S.L. 2012-2014. 413.674,80€.

C.4 Dirección de tesis doctorales

1. Enterprise data integration: on extracting data from HTML tables. J.C. Roldán, 2020 (Codirigida con P. Jiménez).
2. Enterprise information integration: on discovering links using genetic programming. A. Cimmino, 2019.
3. Aspect-based sentiment analysis: a scalable system, a condition miner, and an evaluation dataset. F.O. Gallego, 2018.
4. Enterprise Information Integration: new approaches to web information extraction. P. Jiménez, 2015.
5. Towards automated service procurement: a framework for automated service trading. P. Fernández. 2013.
6. Exchanging data amongst semantic-web ontologies: on generating mappings and benchmarking data exchange systems. C.R. Rivero. 2012. (Codirigida con D. Ruiz).
7. Enterprise information integration: unsupervised proposals for web information extraction. H.A. Sleiman. 2012.
8. Enterprise information integration: an unsupervised proposal for web page classification. I. Hernández. 2012 (Codirigida con D. Ruiz).

9. Enterprise application integration: a model-driven engineering, easy-to-maintain approach. R.Z. Frantz. 2012.

C.5 Colaboración en creación de empresas

1. Opileak, S.L. Transferencia de resultados de investigación en relación con análisis avanzado de opinión en medios sociales, búsqueda y colaboración en proyectos industriales. 2015-.
2. Dinamic Area, S.L. Transferencia de resultados de investigación en relación con análisis de medios sociales, búsqueda y colaboración en proyectos industriales. 2007-.
3. I2Factory, S.L. Transferencia de resultados de investigación en relación con integración de aplicaciones empresariales, búsqueda y colaboración en proyectos industriales. 2008-2017.

C.6 Colaboraciones internacionales

1. University of Southern California (UCS, USA). Craig Knoblock, Pedro Szekely. (2014-)
2. Commissariat a l'Énergie Nucléaire et aux Énergies Alternatives (CEA, France). Hassan A. Sleiman. (2013-)
3. Universidade de Ijuí (UNIJUI, Brasil). Rafael Z. Frantz, Fabricia C. Roos-Fratnz (2013-)

C.7 Comités de programa internacionales

1. Business Information Systems (BIS). GGS=B/C. 2016-.
2. Practical Applications of Agents and Multi-Agent Systems (PAAMS). GCC=B/C. 2009-.
3. Complex, Intelligent, and Software Intensive Systems (CISIS). GCC=B/C. 2008-.
4. International Conference on Distributed Computing and Artificial Intelligence (DCAI). GCC=C. 2008-.
5. Intelligent Data Engineering and Automated Learning (IDEAL). GGS=B/C, 2007-.
6. International Conference on Hybrid Intelligent Systems (HAIS). GCC=C. 2007-.

C.8 Comités de programa nacionales

1. Jornadas de Ingeniería del Software y Bases de Datos (JISBD). 2003-.
2. Conferencia de la Asociación Española de Inteligencia Artificial (CAEPIA). 2007-.

C.9 Comités de evaluación y prospectiva

1. European Science Foundation. 2007-.
2. Agencia Nacional de Evaluación y Prospectiva. 2005-2015.
3. Agencia de Certificación en Innovación Española. 2011-12.
4. Agencia Andaluza del Conocimiento. 2007-12.
5. Consellería de Economía e Industria de la Xunta de Galicia. 2006-12.

C.10 Actividades de gestión

1. Investigador responsable del grupo de investigación oficial de la Comunidad Andaluza TIC 134 "Sistemas Informáticos", 2007-2019.
2. Comisión de Investigación de la Universidad de Sevilla, 2009-2016.
3. Comisión de Asesoramiento del Centro de Formación Permanente de la Universidad de Sevilla, 2010-2015.
4. Subdirector de Planes de Estudio e Innovación Docente de la ETSI Informática de la Universidad de Sevilla, 2007-14.
5. Coordinador del Máster en Ingeniería Informática de la Universidad de Sevilla, 2013-2016.
6. Presidente de la Comisión de Seguimiento de Planes de Estudio de la ETSI Informática de la Universidad de Sevilla, 2010-14.
7. Coordinador del Grado en Ingeniería Informática – Ingeniería del Software de la Universidad de Sevilla, 2010-14.