

Parte A. DATOS PERSONALES		Fecha del CVA	10/2023
Nombre y apellidos	Raquel Martínez Unanue		
DNI/NIE/pasaporte			
Núm. identificación del investigador	Researcher ID		
	Código Orcid		

A.1. Situación profesional actual

Organismo	UNED		
Dpto./Centro	Lenguajes y Sistemas Informáticos/ETSI Informática		
Dirección			
Teléfono		correo electrónico	
Categoría profesional	Catedrática de Universidad	Fecha inicio	
Espec. cód. UNESCO	120317, 120304, 570102, 570104		
Palabras clave	Procesamiento de lenguaje natural, Minería de textos, multilingüismo		

A.2. Formación académica (título, institución, fecha)

Licenciatura/Grado/Doctorado	Universidad	Año
Lic. Informática	Universidad de Deusto	1985
Doctora Informática	Universidad de Deusto	2000

A.3. Indicadores generales de calidad de la producción científica (véanse instrucciones)

Nº sexenios de investigación: 4
 Nº sexenios de transferencia: 1

- Número de tesis doctorales dirigidas: 6 tesis con máxima calificación. De ellas, 2 han sido tesis con Mención Europea y ambas han obtenido el Premio Extraordinario de la UNED, y 2 tesis con Mención Internacional.
- Citas totales: 1188 según Google Scholar, de ellas 423 citas desde 2018 según la misma fuente, índice h: 16 e índice i10: 28.

He participado en 20 proyectos de I+D+I financiados en convocatorias públicas, siendo IP en 4 de ellos. Además, he participado en 4 contratos de I+D+I con Administraciones y empresas, en los que ha habido transferencia, siendo la IP por parte de la universidad en 2 de ellos.

Parte B. RESUMEN LIBRE DEL CURRÍCULUM (máximo 3500 caracteres, incluyendo espacios en blanco)

Soy miembro del grupo de investigación de Procesamiento del Lenguaje Natural (PLN) y Recuperación de Información (PLN&IR) de la UNED (<http://nlp.uned.es>). Mi actividad investigadora se ha desarrollado principalmente en este grupo en el ámbito del Procesamiento del Lenguaje Natural (PLN) y de las Tecnologías del Lenguaje aplicadas a diferentes dominios.

Dentro del campo del PLN, mi actividad investigadora se ha desarrollado principalmente en dos áreas: alineación de corpus paralelos y más recientemente minería de textos monolingües y multilingües abordando diferentes aspectos: representación de documentos monolingües y multilingües, cálculo de la similitud entre documentos y unidades textuales, algoritmos de clustering, modelizando diferentes problemas del estado del arte como problemas de clustering, clasificación de documentos y modelización de diferentes problemas como problemas de clasificación automática. He trabajado con técnicas supervisadas y no supervisadas, así como con técnicas de recuperación de información. He trabajado con documentos heterogéneos de diferentes dominios (documentos

administrativos, noticias, páginas web, tweets) y más recientemente con registros médicos electrónicos y ensayos clínicos.

Actualmente, mi investigación se centra principalmente en el dominio biomédico, abordando específicamente tareas como la anonimización de registros médicos, la identificación y desambiguación de abreviaturas, la identificación de entidades del dominio, la clasificación de registros médicos según el estándar CIE-10, minería de argumentación, enriquecimiento de ontologías y la detección de eventos adversos. Todo ello aplicando técnicas de aprendizaje profundo, aprendizaje no supervisado, diferentes técnicas de aumento de datos y clasificación extrema.

Soy coautora de más de 80 publicaciones científicas en el área, habiendo publicado con investigadores de diferentes universidades, centros de investigación y países. De ellas 36 son artículos indexados en el JCR o de actas de congresos CORE A y A+. He realizado dos estancias de investigación, una predoctoral en el laboratorio CRL (U. de Nuevo Méjico, USA) y otra postdoctoral en la U. James Madison (Virginia, USA). También he colaborado en la creación de recursos, en particular de datasets y goldstandars, disponibles para la comunidad investigadora. Soy evaluadora ANEP.

Parte C. MÉRITOS MÁS RELEVANTES (*ordenados por tipología*)

C.1. Publicaciones (seleccionadas en los últimos 10 años)

Revistas:

M. Almagro, **R. Martínez**, V. Fresno, S. Montalvo. ICD-10 Coding of Spanish Electronic Discharge Summaries: An Extreme Classification Problem. IEEE Access. 8, pp. 100073 - 100083. IEEE, 2020. Disponible en Internet en: <10.1109/ACCESS.2020.2997241>.

M. Almagro; **R. Martínez**; S. Montalvo; V. Fresno. (2/4). 2019. A cross-lingual approach to automatic ICD-10 coding of death certificates by exploring machine translation (<https://doi.org/10.1016/j.jbi.2019.103207>) Journal of biomedical informatics. Elsevier. 94. ISSN 1532-0464.

Agustín Delgado, **Raquel Martinez**, Soto Montalvo , Victor Fresno (2018). Person name disambiguation on the web in a multilingual context. Journal of the Association for Information Science and Technology, vol 68, 1751-1762.

Agustín Delgado, Soto Montalvo, **Raquel Martinez**, Victor Fresno (2018). A survey of person name disambiguation on the web. IEEE ACCESS, vol. 6, 59496- 59514.

Agustín Delgado, **Raquel Martinez**, Soto Montalvo , Victor Fresno. (2017) Person Name Disambiguation in the Web Using Adaptive Threshold Clustering. Journal of the Association for Information Science and Technology <https://doi.org/10.1002/asi.23810>.

Soto Montalvo, **Raquel Martinez**, Leonardo Campillos, Agustín D. Delgado, Víctor Fresno, Felisa Verdejo. (2017) MC4WEPS: a multilingual corpus for Web people search disambiguation. Lang Resources & Evaluation, Vol 51(3), 805-832, 2017. doi:10.1007/s10579-016-9365-4.

Soto Montalvo, **Raquel Martinez**, Victor Fresno. (2015) Quality prediction of multilingual news clustering: An experimental study. Journal of Information Science, vol. 41(4), 518-530, 2015. doi:10.1177/0165551515586671.

Arkaitz Zubiaga, Damiano Spina, **Raquel Martínez**, Víctor Fresno. (2015) *Real-time classification of Twitter trends*. Journal of the Association for Information Science and Technology 66(3): 462-473.

Soto Montalvo, **Raquel Martínez**, Víctor Fresno. (2015) *Quality prediction of multilingual news clustering: An experimental study*. Journal of Information Science 41(4): 518-530.

Soto Montalvo, **Raquel Martínez**, Víctor Fresno, Agustín D. Delgado. (2015). *Exploiting named entities for bilingual news clustering*. Journal of the Association for Information Science and Technology, Volume 66(2): 363-376.

Arkaitz Zubiaga, Víctor Fresno, **Raquel Martínez**, Alberto Pérez García-Plaza. (2013). *Harnessing Folksonomies to Produce a Social Classification of Resources*. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol 25(8): 1801-1813, ISSN: 1041-4347.

Mari-Sanna Paukkeri, Alberto Pérez García-Plaza, Víctor Fresno, **Raquel Martínez Unanue**, Timo Honkela. (2012) "Learning a taxonomy from a set of text documents". *Applied Soft Computing*, vol. 12, pág. 1138–1148. ISSN: 1568-4946.

Alberto Pérez García-Plaza, Arkaitz Zubiaga, Víctor Fresno, **Raquel Martínez**. (2012). *Reorganizing clouds: A study on tag clustering and evaluation*. Expert Systems with Applications 39(10): 9483-9493.

C.2. Proyectos

Título: *DESCUBRIMIENTO DE INFORMACION Y CATEGORIZACION BASADA EN PROCESAMIENTO DEL LENGUAJE PARA EL DOMINIO MEDICO.*

Referencia: (PID2019-106942RB-C32) **Entidad** financiadora y convocatoria: *Ministerio de Ciencia e Innovación. Universidades.*

Nombre del investigador principal: Lourdes Araujo Serna y Juan Martínez Romo.

Fechas inicio y finalización: 01/01/2020 - 31/12/2022.

Tipo de Participación: investigadora.

Título: MODELADO, AUTOMATIZACION DE EXTRACCION DE RELACIONES Y CATEGORIZACION DE INFORMES MEDICOS PARA LA RECOMENDACION DE CODIGOS CIE-10.

Referencia: TIN2016-77820-C3-2-R. **Entidad** financiadora y convocatoria: MINECO 2016.

Nombre del investigador principal: Raquel Martínez Unanue (UNED) y Víctor Fresno Fernández (UNED).

Fechas inicio y finalización: Desde 30/12/2016 hasta: 29/12/2019. N° total de años: 3

Cuantía de la subvención: 94.743 €

Tipo de Participación: Investigador Principal

Título: Recuperacion, organizacion y descubrimiento de relaciones en el contexto de la web, redes sociales y bases de datos médicas.

Referencia: TIN2013-46616-C2-2-R. **Entidad** financiadora y convocatoria: MINECO 2013.

Nombre del investigador principal: Lourdes Araujo Serna (UNED) y Raquel Martínez Unanue (UNED)

Fechas inicio y finalización: Desde 01/01/2014 hasta: 31/12/2016. N° total de meses: 36

Cuantía de la subvención: 101.037 €

Tipo de Participación: Investigador Principal

Título: The automatic encyclopedia of people and organizations.

Referencia: TIN2010-21128-C02. **Entidad** financiadora y convocatoria: MICINN convocatoria nacional.

Nombre del investigador principal: M^a Felisa Verdejo (UNED)

Fechas inicio y finalización: Desde 01/01/2010 hasta: 31/12/2012. N° total de meses: 36

Cuantía de la subvención: 234.982 €

Tipo de Participación: Investigador

Título: Mejorando el Acceso, el Análisis y la Visibilidad de la Información y los Contenidos Multilingüe y Multimedia en Red para la Comunidad de Madrid.

Referencia: S2009/TIC-1542. **Entidad** financiadora y convocatoria: PRICIT. Comunidad Autónoma de Madrid. Proyecto I+D.

Nombre del investigador principal: M^a Felisa Verdejo (UNED)

Fechas inicio y finalización: Desde 01/01/2010 hasta: 31/05/2014. N° total de meses: 53

Cuantía de la subvención: 829.239 €

Tipo de Participación: Investigador

Título: Qualitative Evaluation of Academic Websites Visibility. Multilingual Web Categorization and Information Extraction.

Referencia: TIN2007-67581-C02, TIN2007-67581-C01. **Entidad** financiadora y convocatoria: CICYT Convocatoria nacional.

Nombre del investigador principal: M^a Felisa Verdejo (UNED)

Fechas inicio y finalización: Desde 01/12/2007 hasta: 31/12/2010. N° total de meses: 37

Cuantía de la subvención: 131.890 €

Tipo de Participación: Investigador

Título: Tratamiento de contenidos de páginas web: representación, categorización, estructuración, búsqueda y visualización de relevancia.

Referencia: 07T/0030/2003. **Entidad** financiadora y convocatoria: Comunidad Autónoma de Madrid. Proyectos de I+D.

Nombre del investigador principal: Raquel Martínez Unanue (UNED)

Fechas inicio y finalización: Desde 01/10/2003 hasta: 30/09/2004. N° total de meses: 12

Cuantía de la subvención: 24.150 €

Tipo de Participación: Investigador Principal

C.3. Contratos

Contratos con empresas y de transferencia.

Título: Asignación automática de códigos CIE-10

Tipo: Artículo 83 de la LOU

Empresa o entidad: Pertikos Ltda. Santiago, Chile.

Fechas inicio y finalización: mayo de 2018 hasta agosto de 2019

Investigador Responsable: Raquel Martínez Unanue.

Cuantía de la subvención: 27.280€.

Título: Redacción Cross-Media en Gaceta de los Negocios.

Referencia: FIT-350201-2004-0001

Empresa o entidad: Proyectos PROFIT del Ministerio de Industria, Turismo y Comercio.

Nombre del investigador principal de la universidad: Raquel Martínez Unanue (URJC)

Fechas inicio y finalización: Desde 25/10/2004 hasta: 31/12/2005. N° total de meses: 14

Cuantía de la subvención: 46.150 €

Tipo de Participación: Investigador Principal de la URJC

Título: Sistema Inteligente de Búsqueda on-line de Profesionales.

Empresa o entidad: Cátenon Worldwide Executive Search.

Nombre del investigador principal de la universidad: Raquel Martínez Unanue (URJC)

Fechas inicio y finalización: Desde 27/01/2014 hasta: 27/07/2014. N° total de meses: 6

Cuantía de la subvención: 48.543 €

Tipo de Participación: Investigador por artículo 83 de la LOU.

Título: Plan de Formación y Sensibilización Medioambiental de PYMES en la Zona Sur de la Comunidad de Madrid.

Empresa o entidad: Ministerio de Medio Ambiente.

Nombre del investigador principal de la universidad: Raquel Martínez Unanue (URJC)

Fechas inicio y finalización: Mi participación en el proyecto se desarrolló en dos etapas que en total tuvieron una duración de 19 meses.

Cuantía de la subvención: 1.330.276 €

Tipo de Participación: Investigador por artículo 83 de la LOU.

C.4. Patentes

C.5. Software en explotación

Como resultado del Contrato de Investigación "Sistema Inteligente de Búsqueda on-line de Profesionales" con la empresa Cátenon Worldwide Executive Search se diseñó y desarrolló el buscador especializado KM-Crawler para la búsqueda on-line de profesionales. Actualmente la empresa Cátenon lo tiene en explotación en su filial de la India.

C.6. Participación en la creación de Datasets y Goldstandards de libre distribución:

- DeliciousT140. Autores: A. Zubiaga, A. Pérez García-Plaza, V. Fresno, R. Martínez. (<http://www.zubiaga.org/datasets/delicioust140/>).
- SocialBM0311. Autores: A. Zubiaga, V. Fresno, R. Martínez, A. P. García-Plaza. (<http://www.zubiaga.org/datasets/socialbm0311/>).
- Twitter Trending Topics.. Autores: A. Zubiaga, D. Spina, V. Fresno, R. Martínez. (<http://nlp.uned.es/~damiano/datasets/TT-classification.html>).
- Social-OPD-2k9. Autores: A. Zubiaga, V. Fresno, R. Martínez. (<http://www.zubiaga.org/datasets/socialodp2k9>)
- MC4WEPS: Multilingual Corpus for WEb People Search. Autores: R. Martínez, A. Delgado, V. Fresno, S. Montalvo, F. López Ostenero. (<http://nlp.uned.es/resources/>).

Otros

- Tramos de actividad docente: reconocimiento de **seis** tramos.
- Secretaria Académica de la ETSI Informática desde 1/11/2014 hasta el 5 de julio de 2017.
- Coordinadora del Grado en Tecnologías de la Información de la ETSI Informática desde 15/07/2012 hasta el 31/10/2014.
- Secretaria Adjunta de Grado de la Escuela Técnica Superior de Ingeniería Informática de la UNED desde 15 de febrero de 2010 hasta el 14 de junio de 2012.
- Directora del Dpto. lenguajes y Sistemas Informáticos desde septiembre de 2021.
- Miembro Vocal de la Junta Directiva de la Sociedad Española del Procesamiento del Lenguaje Natural desde 2008 hasta septiembre 2019.
- Miembro del Comité Científico de la revista Procesamiento del Lenguaje Natural desde 2008.
- Miembro del Comité Científico del congreso LREC desde 2008.
- Revisora externa de revistas y congresos internacionales.
- Coorganizadora del 2nd workshop on Human Language Technologies for Iberian languages, IberEval'2017, del 3rd workshop on Human Language Technologies for Iberian languages, IberEval'2018; y del IberLEF 2019 e IberLEF 2020.

Parte A. DATOS PERSONALES

Fecha del CVA	30/10/2023
----------------------	------------

Nombre y apellidos	Alberto Díaz Esteban		
DNI/NIE/pasaporte		Edad	
Núm. identificación del investigador	Researcher ID		
	Código Orcid		

A.1. Situación profesional actual

Organismo	Universidad Complutense de Madrid		
Dpto./Centro			
Dirección			
Teléfono		correo electrónico	
Categoría profesional	Profesor Titular de Universidad	Fecha inicio	3/05/2017
Espec. cód. UNESCO	120304		
Palabras clave	Procesamiento de lenguaje natural, Generación automática de resúmenes, Extracción de información		

A.2. Formación académica (título, institución, fecha)

Licenciatura/Grado/Doctorado	Universidad	Año
Licenciado en Ciencias Físicas	Universidad Complutense de Madrid	1993
Doctor Ingeniero en Informática	Universidad Complutense de Madrid	2005

A.3. Indicadores generales de calidad de la producción científica (véanse instrucciones)

- 2 sexenios de investigación (2000-2007), (2008-2013).
- Fecha del último concedido: 2014
- Tesis dirigidas y leídas en los últimos 10 años: 2
- Publicaciones
 - WOS (JCR): Total de revistas indexadas = 18 / Revistas en Q1 = 6
 - SCOPUS (SJR): Total de revistas indexadas = 20 / Revistas en Q1 = 12
- Citas totales (según Google Scholar): 1116
- Promedio citas/año durante los últimos 5 años (según Google Scholar): 436
- Índice h (Google Scholar): 20

Parte B. RESUMEN LIBRE DEL CURRÍCULUM (máximo 3500 caracteres, incluyendo espacios en blanco)

Alberto Díaz Esteban es profesor titular de universidad en el Depto. de Ingeniería del Software e Inteligencia Artificial de la Universidad Complutense de Madrid (UCM). Es Licenciado en Ciencias Físicas (especialidad de Cálculo Automático) y Doctor en Informática por la UCM.

Su trayectoria docente comenzó en 1995 y comprende tanto formación en grado como en posgrado en 3 centros distintos: la Universidad Europea de Madrid (UEM), el Centro de Estudios Superiores Felipe II, centro adscrito a la Universidad Complutense de Madrid (CES Felipe II UCM) y la Facultad de Informática de la Universidad Complutense de Madrid (UCM). Durante sus más de 25 años de experiencia docente, ha impartido más de 20 asignaturas distintas y ha dirigido más de 50 trabajos de fin de grado y fin de máster. También ha dirigido 2 tesis doctorales.

Es miembro del grupo de investigación NIL (<http://nil.fdi.ucm.es/>), grupo reconocido de la UCM, donde investiga en la aplicación de la generación automática de resúmenes, en el procesamiento del contenido textual de historiales médicos y documentos normativos, en la extracción de interacciones entre entidades biomédicas y en el desarrollo de chatbots para facilitar el acceso a la información. Más recientemente dirige una tesis relacionada con el

procesamiento digital de la lengua de signos y en la aplicación de algoritmos de deep learning para resolver distintas tareas de procesamiento del lenguaje natural.

Ha participado en más de 20 proyectos de investigación. En los últimos años ha sido investigador en proyectos europeos relacionados con la creatividad computacional (2013-2017) y en proyectos nacionales de las convocatorias de Desarrollo e innovación orientada a los retos de la sociedad: "IDyLICO: Inclusión Digital, Lenguaje Natural y Comunicación" (2016-2019) y "CANTOR: Composición automática de narrativas personales como apoyo a terapia ocupacional basada en reminiscencia" (2020-2024). Entre 2020 y 2021 además ha participado en el proyecto "Visualizando la signoescritura".

Durante los años 2017 y 2018 fue investigador principal del proyecto bianual coordinado "HiSi: Desarrollo de una herramienta de búsqueda de historias clínicas similares", financiado por la Universidad Complutense de Madrid (PR26/16-17B-1). Durante 2009 y 2010 fue investigador principal de un proyecto Avanza I+D (AutoIndexer – Investigación y desarrollo de metodologías y recursos terminológicos de apoyo para los procesos de indexación de informes clínicos en áreas de BioMedicina y Salud).

También cabe destacar la participación en proyectos artículo 83, donde lleva colaborando con la empresa ECIX GROUP S.L. de manera ininterrumpida desde el año 2019. Actualmente participa como asesor hasta diciembre de 2024 en 3 licitaciones concedidas a la empresa.

Parte C. MÉRITOS MÁS RELEVANTES

C.1. Publicaciones

AFG Sevilla, AD Esteban, JM Lahoz-Bengoechea (2023)
Automatic SignWriting Recognition: Combining Machine Learning and Expert Knowledge to Solve a Novel Problem
IEEE Access 11, 13211-13222
Citas Google Scholar: 1

B Bokharaeian, M Dehghani, A Diaz (2023)
Automatic extraction of ranked SNP-phenotype associations from text using a BERT-LSTM-based method
BMC bioinformatics 24 (1), 144
Citas Google Scholar: 3

Ping Xiao, Hannu Toivonen, Oskar Gross, Amílcar Cardoso, João Correia, Penousal Machado, Pedro Martins, Hugo Goncalo Oliveira, Rahul Sharma, Alexandre Miguel Pinto, Alberto Díaz, Virginia Francisco, Pablo Gervás, Raquel Hervás, Carlos León, Jamie Forth, Matthew Purver, Geraint A Wiggins, Dragana Miljković, Vid Podpečan, Senja Pollak, Jan Kralj, Martin Žnidaršič, Marko Bohanec, Nada Lavrač, Tanja Urbančič, Frank Van Der Velde, Stuart Battersby (2019)
Conceptual representations for computational concept creation
ACM Computing Surveys (CSUR) 52 (1), 1-33
JCR 2019: 8.7, 4/108 (Q1) COMPUTER SCIENCE, THEORY & METHODS
Citas Google Scholar: 15

B. Bokharaeian, A. Díaz, N. Taghizadeh, H. Chitsaz, R. Chavoshinejad (2017)
SNPPhenA: A Corpus for Extracting Ranked Associations of Single Nucleotide Polymorphisms and Phenotypes from Literature.
Journal of Biomedical Semantics Vol. 8, No. 14, 2017.
JCR 2017: 1.6, 32/59 (Q3) MATHEMATICAL & COMPUTATIONAL BIOLOGY.
Citas Google Scholar: 29

B. Bokharaeian, A. Díaz, H. Chitsaz (2016)

Enhancing Extraction of Drug-Drug Interaction from Literature Using Neutral Candidates, Negation, and Clause Dependency.
PLoS ONE, Vol. 11, No. 10, 2016.
JCR 2016: 2.766, 15/64 (Q1) MULTIDISCIPLINARY SCIENCES
Citas Google Scholar: 14

HG Oliveira, R Hervás, A Díaz, P Gervás (2017)
Multilingual extension and evaluation of a poetry generator
Natural Language Engineering 23 (6), 929-967
JCR 2017: 0.8, 109/132 (Q4) COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE
Citas Google Scholar: 27

Antonio Jimeno, Laura Plaza, James G. Mork, Alan R. Aronson, Alberto Díaz (2013).
MeSH indexing based on automatically generated summaries.
BMC BIOINFORMATICS 14:208 1-12.
JCR 2012: 3.024, 6/47 (Q1) MATHEMATICAL & COMPUTATIONAL BIOLOGY
Citas Google Scholar: 35

Laura Plaza, Mark Stevenson, Alberto Díaz (2012).
Resolving Ambiguity in Biomedical Text to Improve Summarization.
INFORMATION PROCESSING & MANAGEMENT 48(4): 755-766
JCR 2012: 0.817, 41/85 (Q2) INFORMATION SCIENCE & LIBRARY SCIENCE
Citas en Google Scholar: 28

JC De Albornoz, L Plaza, P Gervás, A Díaz (2011)
A joint model of feature mining and sentiment analysis for product review rating
Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011
Citas en Google Scholar: 120

L Plaza, A Díaz, P Gervás (2011)
A semantic graph-based approach to biomedical summarisation
Artificial intelligence in medicine 53 (1), 1-14
JCR 2011: 0.817, 48/111 (Q2) COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE
Citas en Google Scholar: 95

C.2. Proyectos

Proyectos europeos

Concept Creation Technology (ConCreTe)
Comisión Europea. VII Programa Marco. Grant agreement 611733 (2013-2016).
Participantes: Queen Mary College, Helsingin Yliopisto, Universidad Complutense de Madrid, Jozef Stefan Institute, Universiteit Twente, Universidade de Coimbra, Chatterbox Analytics.
Subvención: 231.264 € para la UCM (2.500.000 € en total)
IP en la UCM: Pablo Gervás Gómez Navarro
Tipo de participación: Investigador

Proyectos nacionales

HiSi: Desarrollo de una herramienta de búsqueda de historias clínicas similares.
Universidad Complutense (PR26/16-17B-1) (2017-2019)
Subvención: 20.000 €
Tipo de participación: Investigador Principal

AutoIndexer-Investigación y desarrollo de metodologías y recursos terminológicos de apoyo para los procesos de indexación de informes clínicos en áreas de BioMedicina y Salud.
Ministerio de Industria, Turismo y Comercio. PROYECTO AVANZA I+D (TSI-020100-2009-252) (2009-2010)

Subvención: 205.745 €.

Tipo de participación: Investigador Principal

IDiLyCo: Inclusión digital, lenguaje natural y comunicación.

Ministerio de Economía y Competitividad. Programa nacional de investigación (Tecnologías de la información) TIN2015-66655-R (2016-2019)

Subvención: 68.700 €

Tipo de participación: Investigador

IPs: Pablo Gervás Gómez-Navarro y Raquel Hervás Ballesteros

CANTOR: Composición automática de narrativas personales como apoyo a terapia ocupacional basada en reminiscencia (PID2019-108927RB-I00) (2020-2024)

Subvención: 112.651 €

Tipo de participación: Investigador

IPs: Pablo Gervás Gómez-Navarro y Gonzalo Méndez Pozo

C.3. Contratos

Investigación, desarrollo e implementación de un prototipo funcional de un chatbot para la interacción entre pacientes y hospitales sobre enfermedades cardíacas.

Proyecto artículo 83 con la empresa MedWhat.com, Inc. (2017)

Presupuesto: 10.500 dólares.

Tipo de participación: Investigador principal

5 proyectos artículo 83 entre ECIX GROUP S.L. y la UCM entre los años 2019 y 2023.

Presupuesto acumulado: 41.500 euros

Tipo de participación: Investigador principal

C.4. Transferencia de tecnología

En el marco del proyecto AutoIndexer se desarrollaron 3 productos software que fueron incluidos en el producto comercial SmartCoder: un sistema de corrección ortográfica, un sistema de tratamiento de acrónimos y un sistema de tratamiento de negaciones.

En el marco del proyecto realizado con la empresa MedWhat.com, Inc., se desarrolló un chatbot para la interacción entre pacientes y hospitales sobre enfermedades cardíacas.

En el marco de los proyectos con ECIX GROUP S.L. se han desarrollado prototipos experimentales que se han convertido en productos finales comercializados por la empresa.

C.5. Estancias en centros extranjeros

Natural Language Processing research group, University of Sheffield. Reino Unido. 01/07/2010. 4 semanas. Estancia post-doctoral.

C.6. Tesis doctorales dirigidas

Laura Plaza Morales. Uso de Grafos Semánticos en la Generación Automática de Resúmenes y Estudio de su Aplicación en Distintos Dominios: Biomedicina, Periodismo y Turismo. UCM. Facultad de Informática. 2011.

Behrouz Bokharaeian. Enhancing automatic extraction of biomedical relations using different linguistic features extracted from text. UCM. Facultad de Informática. Junio 2017.

C.7. Otros méritos de especial relevancia

Presidente del Comité Científico del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (septiembre-2013).

Evaluador de la ANEP.



CURRICULUM VITAE ABREVIADO (últimos 6 años) IMPRESO NORMALIZADO

Fecha: 27/10/2023

A. DATOS PERSONALES

Nombre y apellidos	Koldo Gojenola Gallettebeitia		
DNI/NIE/pasaporte		Edad	
Nº identificación del investigador	Researcher ID		
	Código Orcid	0000-0002-2116-6611	

A.1. Situación profesional actual

Entidad	Universidad del País Vasco / Euskal Herriko Unibertsitatea		
Facultad/Escuela/Instituto	Escuela Universitaria de Ingeniería de Bilbao		
Dpto./Centro	Lenguajes y Sistemas Informáticos		
Dirección	Escuela de Ingeniería de Bilbao , UPV/EHU, Paseo Rafael Moreno Pitxitxi, 3, 48013, Bilbao		
Teléfono	946014401	Correo electrónico	Koldo.gojenola@ehu.eus
Categoría profesional	Profesor Titular de Universidad		Fecha inicio 1-02-1989
Situación administrativa	Plantilla	X	Contratado
	Becario		Otra situación
Dedicación	Completa	X	Parcial
	Especialización (cód. UNESCO)		
120304 (Ciencia de los Ordenadores, Inteligencia Artificial), 120323 (Ciencia de los Ordenadores, Lenguajes de Programación), 570105 (Lingüística Aplicada, Lingüística Informatizada)			
Palabras clave	Inteligencia Artificial, Procesamiento del Lenguaje Natural, Extracción de Información, Aprendizaje automático, Aprendizaje Profundo, Procesamiento de Textos Médicos		

A.2. Formación académica

Licenciatura/Grado/Doctorado	Universidad	Año
Licenciatura en Informática	Universidad del País Vasco / Euskal Herriko Unibertsitatea	1988
Doctorado en Informática	Universidad del País Vasco / Euskal Herriko Unibertsitatea	2000

A.3. Cargos y actividades de carácter científico o profesional desempeñadas con anterioridad

Puesto	Entidad	Fechas
Programador	CIKAUTXO S. A. Sociedad Cooperativa	08-1988/02-1989
Profesor Asociado	Facultad de Informática UPV/EHU	02-1989/09-1993
Profesor Titular Interino de E.U.	Facultad de Informática UPV/EHU	09-1993/05-1996
Profesor Titular de E.U.	Facultad de Informática UPV/EHU	05-1996/07-2000
Profesor Titular Doctor de E.U.	Facultad de Informática UPV/EHU	07-2000/09-2001
Profesor Titular Doctor de E.U.	Escuela Universitaria de Ingeniería Técnica Industrial	09-2001/02-2009
Profesor Titular de Universidad	Escuela Universitaria de Ingeniería Técnica Industrial	03-2009/09-2012
Profesor Titular de Universidad	Escuela de Ingeniería de Bilbao	2012/actualidad

A.4. Indicadores generales de calidad de la producción científica

Indique el valor para el conjunto de su trayectoria científica

Número de sexenios de investigación	3
Fecha último sexenio concedido	2021
Nº de tesis doctorales dirigidas	4
Índice H (indicar fuente si no es Web of Science)	Google Scholar: 22 Web of Science: 8
Publicaciones en primer cuartil (Q1) (Web of Science)	5
Nº total de publicaciones	124
Nº total de publicaciones indexadas	Web of Science: 30
Nº de citas totales	Google Scholar: 1945 Web of Science: 195

B. RESUMEN LIBRE DEL CURRÍCULUM (máximo 3500 caracteres)

1983-1988: *Licenciatura en Informática, en la Universidad del País Vasco UPV/EHU.*

1988-1989: *Programador en Cikautxo Cooperativa (6 meses).*

1989-2018: *Profesor e investigador del departamento de Lenguajes y Sistemas Informáticos de la Universidad del País Vasco UPV/EHU. En 1990 entro a formar parte del grupo IXA de investigación en Procesamiento del Lenguaje Natural. En 2000 obtengo el Doctorado (Informática) bajo la dirección del Dr. Kepa Sarasola, en la Universidad del País Vasco, UPV/EHU.*

He participado en distintos proyectos de ámbito autonómico, estatal y europeo. El objetivo principal de estos proyectos ha sido el desarrollo de aplicaciones que permitan procesar textos de manera robusta, eficiente y con alta precisión y cobertura, haciendo uso de métodos lingüísticos y estadísticos, tanto supervisados como no supervisados.

Mi ámbito de investigación se ha dirigido al procesamiento automático del lenguaje. He trabajado en herramientas lingüísticas para el procesamiento del euskara, ayudando al desarrollo de aplicaciones basadas en morfología y sintaxis que están siendo usadas actualmente en distintas aplicaciones. Además del euskara, el enfoque de mi investigación ha sido multilingüe, incluyendo el español, inglés, y lenguas de morfología rica.

Actualmente, trabajo en diversas áreas de investigación, incluyendo el procesamiento morfológico, sintáctico y semántico del euskara, español, e inglés. Además, estoy trabajando en la aplicación de herramientas lingüísticas y estadísticas al dominio médico. Se han desarrollado diversas herramientas de procesamiento de textos del dominio médico con resultados exitosos, habiendo obtenido puestos relevantes en competiciones internacionales (shared tasks). A medio plazo se espera utilizar estas herramientas para obtener aplicaciones de alto nivel para la gestión de datos médicos.

Actualmente tengo tres sexenios de investigación. He trabajado en la organización de conferencias, siendo el presidente del comité organizador de la XXV edición del Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) en 2009, como presidente del comité local en la organización de la International Conference on Parsing Technologies (IWPT) en 2015, en el comité organizador de la XXXV edición del Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) en 2019, y en el comité organizador de la XXXVII edición del Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) en 2021.

Estoy acreditado e imparto docencia en español, euskera e inglés.

C. MÉRITOS MÁS RELEVANTES

Indique los méritos correspondientes únicamente a los últimos 6 años. En los casos en que se hubiera producido alguna interrupción en la labor investigadora motivada por alguna de las situaciones previstas en el segundo párrafo del punto 13 del artículo 26 de la Orden de convocatoria se podrá incrementar el periodo de actividad investigadora a tener en cuenta según lo establecido en dicho párrafo.

C.1. Publicaciones indexadas

Autores: Iker de la Iglesia, María Vivó, Paula Chocrón, Gabriel de Maeztu, Koldo Gojenola, Aitziber Atutxa

Título: An Open Source Corpus and Automatic Tool for Section Identification in Spanish Health Records

Nombre revista: Journal of Biomedical Informatics

Volumen: Pág. inicial: Pág. final: Año: 2023

ISSN: 1532-0464 Lugar de publicación: Elsevier

Indicios de calidad: Base de datos Web of Science

JCR Cuartil Q2

Autores: Iakes Goenaga, Edgar Andrés, Koldo Gojenola, Aitziber Atutxa

Título: Advances in Monolingual and Crosslingual Automatic Disability Annotation in Spanish

Nombre revista: BMC Bioinformatics

Volumen: 24 Pág. inicial: Pág. final: Año: 2023

ISSN: 1471-2105 Lugar de publicación: DOI: 10.1186/s12859-023-05372-3

Indicios de calidad: Base de datos Web of Science

JCR Cuartil Q2

Autores: Iakes Goenaga, Xabier Lahuerta, Aitziber Atutxa, Koldo Gojenola

Título: A Section Identification Tool: towards HL7 CDA/CCR Standardization in Spanish Discharge Summaries

Nombre revista: Journal of Biomedical Informatics

Volumen: 121 Pág. inicial: Pág. final: Año: 2021

ISSN: 1532-0464 Lugar de publicación: Elsevier

Indicios de calidad: Base de datos Web of Science

JCR Cuartil Q1

Autores: A. Atutxa, A. Díaz de Ilarraza, K. Gojenola, M. Oronoz, O. Perez de Viñaspre

Título: Interpretable Deep Learning to Map Diagnostic Texts to ICD10 Codes

Nombre revista: International Journal of Medical Informatics

Volumen: 129 Pág. inicial: 49 Pág. final: 59 Año: 2019

ISSN: 1386-5056 1386-5056 Lugar de publicación: Elsevier

Indicios de calidad: Base de datos Web of Science

JCR Cuartil Q2

Autores: Alicia Pérez; Aitziber Atutxa; Arantza Casillas; Koldo Gojenola; Alvaro Sellart

Título: Inferred joint multigram models for medical term normalization according to ICD

Nombre revista: International Journal of Medical Informatics

Volumen: 110 Pág. inicial: 111 Pág. final: 117 Año: 2018

ISSN: 1386-5056 Lugar de publicación: Elsevier

Indicios de calidad: Base de datos Web of Science

JCR Cuartil Q2

Autores: Alicia Pérez; Rebecka Weegar; Arantza Casillas; Koldo Gojenola; Maite Oronoz; Hercules Dalianis

Título: Semi-supervised medical entity recognition: a study on Spanish and Swedish clinical corpora

Nombre revista: Journal of Biomedical Informatics

Volumen: 71 Pág. inicial: 16 Pág. final: 30 Año: 2017

ISSN: 1532-0464 Lugar de publicación: Elsevier

Indicios de calidad: Base de datos Web of Science

JCR Cuartil Q2

C.2. Otras publicaciones

CLAVE: L= libro completo, CL= capítulo de libro, A= artículo, R= revisión, E= editor, S= documento científico-técnico restringido, O= otras (especificar).

Autores: Itziar Aduriz, Jose Mari Arriola, Xabier Artola, Zuhaitz Beloki, Nerea Ezeiza, Koldo Gojenola

Título: Morfeus+: Word Parsing in Basque beyond Morphological Segmentation

<http://doi.org/10.3366/word.2020.0172>

Nombre revista/Título libro: WORD STRUCTURE Online ISSN: 1755-2036

Clave: A

Volumen: 13.3 Pág. inicial: 283 Pág. final: 315 Año: 2020

Editorial (si libro):
Indicios de calidad:

Lugar de publicación:

C.3. Contribuciones a congresos, conferencias científicas y seminarios

Indique un máximo de 10 contribuciones

Autores: Iakes Goenaga, Aitziber Atutxa, Koldo Gojenola, Arantza Casillas, Arantza Díaz de Ilarraza, Nerea Ezeiza, Maite Oronoz, Alicia Pérez, Olatz Perez de Viñaspre

Título: Automatic Misogyny Identification Using Neural Networks

Tipo de participación:

Congreso:

Publicación: Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018). ISSN 1613-073

Lugar de celebración:

Fecha: 2018

Autores: Aitziber Atutxa, Arantza Casillas, Nerea Ezeiza, Iakes Goenaga, V. Fresno, Koldo Gojenola, R. Martínez, Maite Oronoz, Olatz Perez-de-Viñaspre

Título: IxaMed at CLEF eHealth 2018 Task 1: ICD10 Coding with a Sequence-to-Sequence approach

Tipo de participación:

Congreso: CLEF 2018 Online Working Notes.

Publicación: CEUR-WS

Lugar de celebración:

Fecha: 2018

Autores: Iakes Goenaga, Aitziber Atutxa, Koldo Gojenola, Arantza Casillas, Arantza Díaz de Ilarraza, Nerea Ezeiza, Maite Oronoz, Alicia Pérez, Olatz Perez de Viñaspre

Título: A Hybrid Approach For Automatic Disability Annotation

Tipo de participación:

Congreso: Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018). ISSN 1613-073.

Publicación:

Lugar de celebración:

Fecha: 2018

C.4. Proyectos de investigación

Título del proyecto: EXTRECM: EXTracción de RELaciones entre Conceptos Médicos en fuentes de información heterogéneas

Entidad financiadora (convocatoria): Ministerio de Economía y Competitividad

Subvención concedida: 164.540,64 €

Fecha inicio (dd/mm/aa): 01/01/2014

Fecha fin (dd/mm/aa): 31/12/2016

Investigador/a principal: Arantza Díaz de Ilarraza y Koldo Gojenola

Título del proyecto: Prosamed: Advanced semantic textual processing for the detection of diagnostic codes, procedures, concepts and their relationships in health records

Entidad financiadora (convocatoria): Ministerio de Economía y Competitividad

Subvención concedida: 154.638 €

Fecha inicio (dd/mm/aa): 01/01/2017

Fecha fin (dd/mm/aa): 31/12/2019

Investigador/a principal: Koldo Gojenola y Arantza Díaz de Ilarraza

Título del proyecto: DOTT-HEALTH: Advanced semantic textual processing for the detection of diagnostic codes, procedures, concepts and their relationships in health records

Entidad financiadora (convocatoria): Ministerio de Economía y Competitividad

Subvención concedida: 149.000 €

Fecha inicio (dd/mm/aa): 01/06/2020

Fecha fin (dd/mm/aa): 31/05/2023

Investigador/a principal: Koldo Gojenola y Arantza Casillas

C.5. Contratos de I+D con empresas y/o administraciones

Título del contrato: Itzulbide: Desarrollo e implementación de un sistema de traducción automática neuronal para traducir informes clínicos entre el euskera y el castellano (G/211/20/1/0199/o001/0000/012019).

Tipo de contrato:

Entidad financiadora (convocatoria): Servicio Vasco de Salud, Osakidetza.

Importe del contrato: 191,469.58 €

Fecha inicio (dd/mm/aa): 05/06/2019

Fecha fin (dd/mm/aa): 05/06/2021

Investigador/a principal: Maite Oronoz Anchordoqui

C.6. Patentes y modelos de utilidad

Inventores: Maite Oronoz, Aitziber Atutxa, Alicia Perez, Arantza Diaz de Ilarraza, Arantza Casillas, Sara Santiso

Título: Registro de la Propiedad Intelectual: IxaMedTagger: Automatic system for the automatic detection of entities in the medical domain in Spanish

Nº de solicitud: País de prioridad: España

Fecha de prioridad: 2019

Entidad titular: Universidad del País Vasco UPV/EHU

Países a los que se ha extendido: España

Empresa/s que la están explotando:

C.7. Dirección de tesis doctorales

Título: ASKHI: Análisis sintáctico computacional híbrido basado en la combinación de distintos paradigmas

Doctorando: Iakes Goenaga Azkarate

Universidad: Universidad del País Vasco / Escuela de Doctorado: Facultad de Informática de Euskal Herriko Unibertsitatea San Sebastián

Departamento/Programa de Doctorado: Análisis y Procesamiento del Lenguaje

Directores/as: Nerea Ezeiza, Koldo Gojenola

Calificación: Sobresaliente cum laude

Año: 2017

Título: Hacia el análisis de sentimiento automático: creación de recursos básicos e identificación de modificadores de valencia de diferentes niveles

Doctorando: Jon Alkorta Agirrezabala

Universidad: Universidad del País Vasco / Escuela de Doctorado: Facultad de Informática de Euskal Herriko Unibertsitatea San Sebastián

Departamento/Programa de Doctorado: Análisis y Procesamiento del Lenguaje

Directores/as: Mikel Irukieta, Koldo Gojenola

Calificación: Sobresaliente cum laude

Año: 2019

C.9. Participación en comités científicos, técnicos y/o asesores

Nombre del comité: Experto evaluador de la Agencia Estatal de Investigación

Entidad de afiliación: Agencia Estatal de Investigación

Ámbito: España

Tipo de participación: Evaluación de becas, contratos y proyectos

Fecha inicio (dd/mm/aa): 02/2017

Fecha fin (dd/mm/aa):

Nombre del comité: Comité Científico de la revista "Procesamiento del Lenguaje Natural", ISSN: 1135-594

Entidad de afiliación: Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)

Ámbito: Internacional

Tipo de participación: Evaluación de artículos científicos

Fecha inicio (dd/mm/aa): 02/2007

Fecha fin (dd/mm/aa):

C.10. Organización y gestión de actividades de I+D+i

Título: Revisor de artículos científicos en revistas de calidad (JCR y CORE A): Congreso anual de la Association for Computational Linguistics ACL (CORE A+), Congreso anual de Empirical Methods in Natural Language Processing EMNLP (CORE A+), Journal of Information Science (JIS), Journal of Natural Language Engineering (NLE), Journal BMC Medical Informatics and Decision Making – MIDM, Journal of Applied Clinical Informatics.

Tipo de actividad: Evaluación de artículos científicos

Ámbito:

Funciones desempeñadas: Revisor científico y evaluador

Fecha inicio (dd/mm/aa): 2010

Fecha fin (dd/mm/aa):

Fecha del CVA	27/10/2023
---------------	------------

Parte A. DATOS PERSONALES

Nombre	Alvaro		
Apellidos	Rodrigo Yuste		
Sexo	Hombre	Fecha de Nacimiento	
DNI/NIE/Pasaporte			
URL Web	http://nlp.uned.es/~alvarory/		
Dirección Email			
Open Researcher and Contributor ID (ORCID)	0000-0002-6331-4117		

A.1. Situación profesional actual

Puesto	Profesor Titular de Universidad		
Fecha inicio	2020		
Organismo / Institución	Universidad Nacional de Educación a Distancia		
Departamento / Centro	Lenguajes y Sistemas Informáticos / ETSI Informática		
País		Teléfono	
Palabras clave	Inteligencia artificial; Ciencias de la computación y tecnología informática		

A.2. Situación profesional anterior (incluye interrupciones en la carrera investigadora - indicar meses totales, según texto convocatoria-)

Periodo	Puesto / Institución / País
2018 - 2020	Profesor Contratado Doctor / Universidad Nacional de Educación a Distancia
2015 - 2018	Profesor Ayudante Doctor / Universidad Nacional de Educación a Distancia
2010 - 2015	Profesor Ayudante / Universidad Nacional de Educación a Distancia
2007 - 2010	Personal investigador / Universidad Nacional de Educación a Distancia
2006 - 2006	Becario FPI UNED / Universidad Nacional de Educación a Distancia
2005 - 2006	Becario de Investigación asociado a proyecto / Universidad Nacional de Educación a Distancia

A.3. Formación académica

Grado/Master/Tesis	Universidad / País	Año
Doctor por la Universidad UNED. Programa de Doctorado en Inteligencia Artificial y Sistemas Informáticos por la Universidad Nacional de Educación a Distancia dentro del Programa en Inteligencia Artificial y Sistemas Informáticos	Universidad Nacional de Educación a Distancia / España	2010
Máster Universitario en Lenguajes y Sistemas Informáticos	Universidad Nacional de Educación a Distancia / España	2007
Ingeniero en Informática	Universidad Complutense de Madrid / España	2005

Parte B. RESUMEN DEL CV

Mi carrera investigadora se ha centrado en el área de procesamiento del lenguaje natural, con una especial atención a los sistemas de búsqueda de respuestas. Dentro de este campo, he realizado varios trabajos enfocados a la evaluación y desarrollo de este tipo de sistemas,

realizando propuestas encaminadas a mejorar su rendimiento. Este trabajo se ha realizado dentro de varios proyectos de investigación nacionales, internacionales y autonómicos, dando lugar a distintas publicaciones en revistas y congresos de prestigio. Además, también he colaborado y organizado varias tareas internacionales de evaluación competitiva de este tipo de sistemas.

Parte C. LISTADO DE APORTACIONES MÁS RELEVANTES

C.1. Publicaciones más importantes en libros y revistas con “peer review” y conferencias

AC: Autor de correspondencia; (nº x / nº y): posición firma solicitante / total autores. Si aplica, indique el número de citas

- 1 **Artículo científico**. 2022. Study of a lifelong learning scenario for question answering. Expert Systems With Applications. Elsevier. 209, pp.1-9. ISSN 0957-4174.
- 2 **Artículo científico**. 2022. UNED at PoliticEs 2022: Testing Approximate Nearest Neighbors and Spanish Language Models for Author Profiling in Political Ideology. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022). CEUR. 3202.
- 3 **Artículo científico**. 2022. UNED at QuALES 2022: Testing the Performance of Transformer-Based Language Models for Spanish Question-Answering. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022). CEUR. 3202.
- 4 **Artículo científico**. 2022. appForum: Una aplicación para el procesamiento de foros (appForum: An application for forum processing). Proceedings of the Annual Conference of the Spanish Association for Natural Language Processing: Projects and Demonstrations (SEPLN-PD 2022). CEUR. 3224.
- 5 **Artículo científico**. 2021. VaxxStance@IberLEF 2021: Overview of the Task on Going Beyond Text in Cross-Lingual Stance Detection. Procesamiento del Lenguaje Natural. 67, pp.173-181. ISSN 1135-5948.
- 6 **Artículo científico**. Jan Deriu; (2/7) Alvaro Rodrigo; Arantxa Otegi; Guillermo Echegoyen; Sophie Rosset; Eneko Agirre; Mark Cieliebak. 2021. Survey on Evaluation Methods for Dialogue Systems. Artificial Intelligence Review. Springer. 54-1, pp.755-810. ISSN 1573-7462.
- 7 **Artículo científico**. 2020. DeepReading @ SardiStance 2020: Combining Textual, Social and Emotional Features. Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020).
- 8 **Artículo científico**. 2020. Analyzing User Profiles for Detection of Fake News Spreaders on Twitter. Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum.
- 9 **Artículo científico**. Jan Deriu; Katsiaryna Mlynchyk; Philippe Schläpfer; et al; Mark Cieliebak. 2020. A Methodology for Creating Question Answering Corpora Using Inverse Data Annotation. Proceedings of 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020).
- 10 **Artículo científico**. Guillermo Echegoyen; (2/3) Alvaro Rodrigo; Anselmo Peñas. 2020. Cross-lingual Training for Multiple-Choice Question Answering. Procesamiento del Lenguaje Natural. 65. ISSN 1135-5948.
- 11 **Artículo científico**. Jan Deriu; Don Tuggener; Pius von Däniken; et al; Mark Cieliebak. 2020. Spot The Bot: A Robust and Efficient Framework for the Evaluation of Conversational Dialogue Systems. Proceedings of EMNLP 2020.
- 12 **Artículo científico**. 2019. Benchmarking Entity Linking for Question Answering over Knowledge Graphs. Procesamiento del Lenguaje Natural. 63, pp.121-128. ISSN 1135-5948.
- 13 **Artículo científico**. 2019. The effect of answer validation on the performance of Question-Answering systems. Expert Systems With Applications. 116, pp.351-363. ISSN 0957-4174. <https://doi.org/10.1016/j.eswa.2018.09.014>
- 14 **Artículo científico**. 2018. Do systems pass university entrance exams?. Information Processing and Management. 54-4, pp.564-575. ISSN 0306-4573. <https://doi.org/10.1016/j.ipm.2018.03.002>

- 15 **Artículo científico.** 2017. A study about the future evaluation of Question-Answering systems. Knowledge-Based Systems. 137, pp.83-93. ISSN 0950-7051. <https://doi.org/10.1016/j.knosys.2017.09.015>
- 16 **Artículo científico.** (1/2) Alvaro Rodrigo; Anselmo Peñas. 2014. On Evaluating the Contribution of Validation for Question Answering. IEEE Transactions on Knowledge and Data Engineering. IEEE. 27, pp.1157-1161. ISSN 1041-4347. <https://doi.org/10.1109/TKDE.2014.2373363>
- 17 **Artículo científico.** 2013. Answering questions about European legislation. Expert Systems with Applications. Elsevier. 40, pp.5811-5816. ISSN 0957-4174. <https://doi.org/10.1016/j.eswa.2013.05.008>
- 18 **Artículo científico.** 2013. QA4MRE 2011-2013: Overview of Question Answering for Machine Reading Evaluation. Lecture Notes in Computer Science. 8138, pp.303-320. ISSN 0302-9743.
- 19 **Artículo científico.** 2012. Evaluating Machine Reading Systems through Comprehension Tests. Language Resources and Evaluation Conference (LREC) 2012. ISBN 978-2-9517408-7-7.
- 20 **Artículo científico.** 2012. Evaluating Question Answering Validation as a Classification Problem. Language Resources and Evaluation. SPRINGER. 46(3), pp.493-501. ISSN 1574-020X. <https://doi.org/10.1007/s10579-011-9143-2>
- 21 **Artículo científico.** 2012. Question Answering at the Cross-Language Evaluation Forum 2003-2010. Language Resources and Evaluation. Springer. 46, pp.177-217. ISSN 1574-020X. <https://doi.org/10.1007/s10579-012-9177-0>
- 22 **Artículo científico.** 2012. Temporally Anchored Relation Extraction. Proceedings of 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012). 0, pp.107-116. <https://doi.org/http://www.aclweb.org/anthology/P12-1012>
- 23 **Artículo científico.** 2011. A Simple Measure to Assess Non-response. Proceedings of 49th Annual Meeting of the Association for Computational Linguistics - Human Language Technologies (ACL-HLT 2011). 0, pp.1415-1424. <https://doi.org/http://www.aclweb.org/anthology/P11-1142>
- 24 **Artículo científico.** 2009. Comparación de Enfoques para Evaluar la Validación de Respuestas. Procesamiento del Lenguaje Natural. 43, pp.277-285. ISSN 1135-5948.
- 25 **Artículo científico.** 2008. Testing the Reasoning for Question Answering Validation. Journal of Logic and Computation. OXFORD UNIVERSITY PRESS. 18(3), pp.459-474. ISSN 0955-792X. Google scholar (33) <https://doi.org/10.1093/logcom/exm072>
- 26 **Artículo científico.** 2006. SPARTE, a Test Suite for Recognising Textual Entailment in Spanish. International Conference on Intelligent Text Processing and Computational Linguistics, Lecture Notes in Computer Science. 0, pp.275-286. ISBN 3-540-32205-1.
- 27 **Capítulo de libro.** 2012. Multilingual Information Access. Emerging Applications of Natural Language Processing: Concepts and New Research. IGI GLOBAL. pp.203-228. ISBN 9781466621695.

C.3. Proyectos o líneas de investigación

- 1 **Proyecto.** HOLISTIC ANALYSIS OF ORGANISED MISINFORMATION ACTIVITY IN SOCIAL NETWORKS. Ministerio de Ciencia e Innovación. Peñas Padilla Catedrático de Universidad. (Universidad Nacional de Educación a Distancia). 01/12/2022-30/11/2025. 260.145 €.
- 2 **Proyecto.** MODELOS DE LENGUAJE PROFUNDOS EN LA INTERPRETACION DE LA INFORMACION Y LA DESINFORMACION EN CONTEXTO. Ministerio de Ciencia e Innovación. Peñas Padilla Catedrático de universidad. (Universidad Nacional de Educación a Distancia). 01/01/2022-31/12/2024. 47.916 €.
- 3 **Proyecto.** RTI2018-096846-B-C21, DeepReading. Ministerio de Ciencia, Innovación y Universidades. Rodrigo Agerrí. (Universidad Nacional de Educación a Distancia). 01/12/2019-30/11/2021. 146.894 €.
- 4 **Proyecto.** PCIN-2017-085, LIHLITH-KIQA. Ministerio de economía, industria y competitividad. Anselmo Peñas Padilla. (Universidad Nacional de Educación a Distancia). 01/12/2017-30/11/2020. 69.000 €.

- 5 Proyecto.** TIN2015-65308-C5-1-R, TUNER, Adaptación al Dominio para Procesamiento Semántico Avanzado. Ministerio de Economía y Competitividad. German Rigau Claramunt. (Universidad Nacional de Educación a Distancia). 01/01/2016-31/12/2018. 158.600 €.
- 6 Proyecto.** TIN2013-4709-C3-1P, Voxpopuli. Ministerio de Economía y competitividad. Felisa Verdejo Maillo. (Universidad Nacional de Educación a Distancia). 01/01/2014-31/12/2016. 126.352,4 €.
- 7 Proyecto.** TIN2010-21128-C02-01, Holopedia: la enciclopedia automática de personas y organizaciones. Ministerio de Ciencia e Innovación. VERDEJO MAILLO MARIA FELISA. Desde 01/01/2011. 234.982 €. Otros.
- 8 Proyecto.** S2009/TIC-1542, Mejorando el acceso. El análisis y la visibilidad de la información y los contenidos multilingües y multimedia en red para la comunidad de Madrid. Comunidad Autónoma de Madrid. VERDEJO MAILLO MARIA FELISA. Desde 01/01/2010. 132.294,32 €. Otros.
- 9 Proyecto.** CEN-20091026, Buscamedia. Ministerio de Industria. Ana García Serrano. (ATOS,ACTIVA,GFI,INDRA,CCMA,BILBOMATICA, FRACTALIA,BMAT,ISID,ISOCO,DAEDALUS).
- 10 Proyecto.** TIN2007-67581-C02-01, Multilingual web categorization and information extraction. Ministerio de Ciencia e Innovación. VERDEJO MAILLO MARIA FELISA. (UNED, CINDOC). Desde 01/12/2007. 131.890 €. Otros.
- 11 Proyecto.** S-0505/TIC/0267, Mejorando el acceso y la visibilidad de la información multilingüe en red para la Comunidad de Madrid. Comunidad Autónoma de Madrid. VERDEJO MAILLO MARIA FELISA. (UNED, UPM, UC3M, UAM, UEM, CINDOC). Desde 01/01/2006. 240.996 €. Otros.
- 12 Proyecto.** R2D2. Comisión Interministerial de Ciencia y Tecnología. Felisa Verdejo. (UNED, Universidad de Alicante, Universidad Politécnica de Valencia, Universidad de Jaén). Desde 01/12/2003. 180.320 €. Otros.



CURRICULUM VITAE (CVA)

IMPORTANT – The Curriculum Vitae cannot exceed 4 pages. Instructions to fill this document are available in the website.

27-10-2023

Part A. PERSONAL INFORMATION

First name	Lourdes	
Family name	Araujo	
Gender (*)	Woman	http://nlp.uned.es/~lurdes
Open Research and Contributor ID (ORCID)(*)	0000-0002-7657-4794	

(*) Mandatory

A.1. Current position

Position	Full Professor
Initial date	29/05/2012
Institution	Universidad Nacional de Educación a Distancia UNED
Departament/Center	Departamento de Lenguajes y Sistemas Informáticos
Country	Spain
Key words	Natural language processing; information retrieval; web documents and social networking; biomedical domain; evolutionary algorithms

A.2. Previous positions (research activity interruptions, art. 45.2.c))

Period	Position/Institution/Country/Interruption cause
1/7/1987 –17/11/1990	System analyst ALCATEL
18/11/1990 -30/9/1992	Assistant professor (Universidad Complutense de Madrid)
30/9/1992-26/10/2006	Associate Professor (Universidad Complutense)
26/10/2006 -29/5/2012	Associate Professor (UNED)

A.3. Education

PhD, Licensed, Graduate	University/Country	Year
PhD Computer Science	Complutense (Spain)	1994
Physics Degree	Complutense (Spain)	1987

Part B. CV SUMMARY (max. 5000 characters, including spaces)

Degree in Physics with a specialty in Computer Science from the Complutense University of Madrid in 1987. I worked for more than 3 years as a systems analyst at ALCATEL. After returning to the Complutense University as Assistant Professor, I obtained my PhD at the same University in 1994 in the Official Doctoral Program in Computer Science and Automatics. I did a pre-doctoral stay at the University of Sheffield (UK). In 1996 I obtained the position of Associate Professor at the Complutense University of Madrid. I developed my work at that University until 2006, when I moved to the Universidad Nacional de Educación a Distancia, joining the Natural Language Processing and Information Retrieval research

group, where research related to my lines is developed. In 2013 and in 2017 I carried out research stays at the University of York (UK).

From 2000 to the present my research has focused on different problems related to language processing and has focused especially on the application of these techniques to the biomedical domain. In this field we collaborate with different institutions, having agreements with the Hospital Clínico San Carlos de Madrid (HCSC), and the Consejería de Sanidad de la Comunidad de Madrid. We also collaborate with the Instituto de Salud Carlos III (ISCIII). In collaboration with these institutions, we address problems of great social relevance such as the characterization of rare diseases and the improvement of their registration in the Community of Madrid. With the psychiatry service of the HCSC we investigate the early detection of problems related to mental health such as suicide, self-injury, social isolation, etc. With the ISCIII we work on the identification of comorbidity in problems such as AIDS.

I have **5 research periods** approved (last one in 2021) and 6 teaching periods. I have published more than 100 publications with more than 2000 cites, 38 in JCR indexed journals, most of them in the first positions of the rankings, and 21 CORE A conferences. I have supervised 6 PhD theses, 3 of them in doctoral programs with quality mention, 3 with European doctorate mention and 2 with extraordinary doctorate award. I am currently supervising another two theses. I have supervised 3 FPI grants. I have participated in 23 research projects being PI in 8 of them. I was coordinator of the degree in Information Technology Engineering from 2009 to 2012, the period of its implementation. I was Deputy Director of Academic Planning and Quality at the E.T.S.I Informática de la UNED from 2013 to 2017. I have been head of the Department of Computer Languages and Systems of the UNED from 2017 to 2021. I have been a member of ANECA Titular Commission for access to the Corps of University Professors in the area of Engineering and Architecture from 2011 to 2014. I have been an official collaborator of the State Research Agency (AEI) from 2019 to 2022. I am also a regular reviewer for several JCR journals and prestigious conferences. I am a member of the scientific committee of several conferences. I have carried out different divulgation activities: participation in the launching of the DiverTles community, organization of NLP round table in medicine, talks in AI center in Orense, National School of Health, among others.

Part C. RELEVANT MERITS (sorted by typology)

C.1. Publications

- 1.- Hermenegildo Fabregat; Andres Duque; Juan Martinez-Romo; Lourdes Araujo: *Negation-based transfer learning for improving biomedical Named Entity Recognition and Relation Extraction*. Journal of Biomedical Informatics 138:104279 (2023) (2023). JCR Q1
- 2.- L Araujo, J Martinez-Romo, O Bisbal, R Sanchez-de-Madariaga: *Discovering HIV related information by means of association rules and machine learning*. Scientific Reports 12 (1), 1-12 (2022). JCR Q2
- 3.- Ricardo Sánchez-de-Madariaga, Juan Martínez-Romo, José Miguel Cantero Escribano, Lourdes Araujo: *Semi-supervised incremental learning with few examples for discovering medical association rules*. BMC Medical Informatics Decis. Mak. 22(1): 20 (2022). Cites: 2. JCR Q3
- 4.- Andrés Duque, Hermenegildo Fabregat, Lourdes Araujo, Juan Martinez-Romo: *A keyphrase-based approach for interpretable ICD-10 code classification of Spanish medical reports*. Artificial Intelligence in Medicine 121: 102177 (2021). DOI: 10.1016/j.artmed.2021.102177. Cites: 4. JCR Q1
- 5.- H Fabregat, J Martínez-Romo, L Araujo : *Understanding and Improving Disability Identification in Medical Documents*. IEEE Access 8, 155399-155408. Cites: 4. JCR Q2

- 6.- Lourdes Araujo: *Genetic Programming for Natural Language Processing*. Genetic Programming and Evolvable Machines 21(1): 11-32 (2020). DOI: 10.1007/s10710-019-09361-5. Cites: 7. JCR Q2
- 7.- Juan R. Martínez-Rico, Juan Martínez-Romo, Lourdes Araujo: *Can deep learning techniques improve classification performance of vandalism detection in Wikipedia?*. Eng. Appl. of AI 78: 248-259 (2019). DOI: 10.1016/j.engappai.2018.11.012. Cites: 16. JCR Q1
- 8.-Oscar Rodriguez-Prieto, Lourdes Araujo, Juan Martínez-Romo: *Discovering related scientific literature beyond semantic similarity: a new co-citation approach*. Scientometrics 120(1): 105-127 (2019). DOI: 10.1007/s11192-019-03125-9. Cites: 11. JCR Q2
- 9.- Andres Duque, Mark Stevenson, Juan Martinez-Romo, Lourdes Araujo: *Co-occurrence graphs for word sense disambiguation in the biomedical domain*. Artificial Intelligence in Medicine 87: 9-19 (2018). DOI: 10.1016/j.artmed.2018.03.002. Cites: 42. JCR Q1
- 10.- Hermenegildo Fabregat, Lourdes Araujo, Juan Martínez-Romo: *Deep neural models for extracting entities and relationships in the new RDD corpus relating disabilities and rare diseases*. Computer Methods and Programs in Biomedicine 164: 121-129 (2018). DOI: 10.1016/j.cmpb.2018.07.007. Cites; 15. JCR Q1
- 12.- Andrés Duque Fernández, Juan Martínez-Romo, Lourdes Araujo: *Can multilinguality improve Biomedical Word Sense Disambiguation?* Journal of Biomedical Informatics 64: 320-332 (2016). DOI: 10.1016/j.jbi.2016.10.020. Cites: 11. JCR Q1.
- 13.- Juan Martinez-Romo, Lourdes Araujo, Andres Duque Fernandez: *SemGraph: Extracting keyphrases following a novel semantic graph-based approach*. JASIST 67(1): 71-82 (2016). DOI: 10.1002/asi.23365. Cites: 43. JCR Q2
- 14.- Andres Duque Fernández, Juan Martínez-Romo, Lourdes Araujo: *Choosing the best dictionary for Cross-Lingual Word Sense Disambiguation*. Knowl.-Based Syst. 81: 65-75 (2015). DOI: 10.1016/j.knosys.2015.02.007. Cites: 14. JCR Q1
- 15.- Andres Duque Fernandez, Lourdes Araujo, Juan Martinez-Romo: *CO-graph: A new graph-based technique for cross-lingual word sense disambiguation*. Natural Language Engineering 21(5): 743-772 (2015). DOI: 10.1017/S1351324915000091. Cites:12. JCR Q3
- 16.- Juan Martinez-Romo, Lourdes Araujo: *Detecting malicious tweets in trending topics using a statistical analysis of language*. Expert Syst. Appl. 40(8): 2992-3000 (2013). DOI: 10.1016/j.eswa.2012.12.015. Cites: 259. JCR Q1
17. Álvaro Rodrigo, Joaquín Pérez-Iglesias, Anselmo Peñas, Guillermo Garrido, Lourdes Araujo: *Answering questions about European legislation*. Expert Syst. Appl. 40(15): 5811-5816 (2013). DOI: 10.1016/j.eswa.2013.05.008. Cites: 15. JCR Q1

C.2. Congress

- Fernando López-Ostenero, Laura Plaza, Lourdes Araujo, Juan Martínez-Romo: *Self-Assesment tool with topic-driven navigation for algorithms learning*. EDUCON 2022: 356-363.
- Juan R. Martinez-Rico, Juan Martínez-Romo, Lourdes Araujo: *NLP&IR@UNED at CheckThat! 2021: Check-worthiness estimation and fake news detection using transformer models*. CLEF 2021: 545-557 (oral)
- Hermenegildo Fabregat, Andrés Duque, Lourdes Araujo, Juan Martínez-Romo: *LSI_UNED at CLEF eHealth2021: Exploring the effects of transfer learning in negation detection and entity recognition in clinical texts*. CLEF 2021: 780-793 (oral)
- Elena Campillo Ageitos, Hermenegildo Fabregat, Lourdes Araujo, Juan Martínez-Romo: *NLP-UNED at eRisk 2021: self-harm early risk detection with TF-IDF and linguistic features*. CLEF 2021: 950-965 (oral)
- **Organización de la tarea** de evaluación DIANN:
Hermenegildo Fabregat, Juan Martínez-Romo, Lourdes Araujo:
Overview of the DIANN Task: Disability Annotation Task. IberEval@SEPLN 2018: 1-14

C.3. Research projects

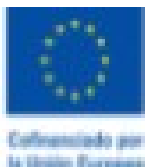
- 1.- EDHER-MED: Early Detection of HEalth Risks by textual analysis of MEDical documents (PID2022-136522OB-C21). Ministerio de Ciencia e Innovación (Transición Digital). Participation: PI. (1/9/2023 – 31/8/2026).
- 2.- GELP: GEneration with Language Processing of demographic profiles in social networks for detection of suicide risk and its relationship with other psychological problems (TED2021-130398B-C21). Ministerio de Ciencia e Innovación (Transición Digital). Participation: PI. (1/12/2022 – 30/11/2024).
- 3.- RAICES: Rules of association in the investigation of diseases of special interest. Instituto Mixto IMIENS: UNED, Instituto de Salud Carlos III (Call 2022). Participation: PI.
- 4.- INDICA-MED: INformation DIScovery and CAtegorization based on language processing for the MEDical domain (PID2019-106942RB-C32). Ministerio de Ciencia e Innovación Participation: PI. (1/6/2020 – 31/5/2023)
- 5.-MAMTRA-MED: Modelado, automatización de extracción de relaciones y categorización de informe médicos para recomendación de códigos CIE-10 (TIN2016-77820-C3-2-R). Ministerio de Economía y Competitividad. Participation: researcher (1/1/2014-30/06/2018).
- 6.- EXTRAE II: EXTRacción de Asociaciones entre Enfermedades y otros conceptos médicos II: Aumentando la complejidad de datos y modelos. Instituto Mixto IMIENS: UNED, Instituto de Salud Carlos III (Call 2019). Participation: PI.
- 7.- IABiomed-net: Red temática sobre Inteligencia Artificial en Biomedicina (RED2018-102312-T). Coordinadora: M^a del Mar Marcos López. Participation: PI (UNED node) (Call 2018).
- 8.-EXTRAЕ: Extracción de Asociaciones entre Enfermedades y otros conceptos médicos (EXTRAЕ). Instituto Mixto IMIENS: UNED, Instituto de Salud Carlos III. (Convocatoria 2017) (1/10/2017-30-9-2019). Participation: PI.
- 9.- EXTRECM: EXTRacción de RELaciones entre Conceptos Médicos en fuentes de información heterogéneas (TIN2013-46616-C2-2-R). Ministerio de Economía y Competitividad. Participation: PI (1/1/2014-30/06/2018).

C.4. Contracts, technological or transfer merits

- Collaboration con el Instituto Tecnológico de Castilla León (ITCL) 8-3-2023 (in charge)
- Collaboration agreement with the San Carlos Clinical Hospital 25-9-2020 (in charge)
- Collaboration agreement with the Consejería de Sanidad de la C. Madrid 14-12-2021 (in charge)
- Collaboration with the Instituto de Salud Carlos III (IP) (since 2017)

Scientific resources generated:

- DIANN: bilingual corpus in Spanish and English with disability annotations (2018).
- RDD: corpus of relationships between rare diseases and disabilities (2018).
- EBCRD: Elsevier Bilingual Corpus for Rare Diseases, summaries of scientific articles in English and Spanish on rare diseases (2016).



CURRICULUM VITAE (CVA)

IMPORTANT – The Curriculum Vitae cannot exceed 4 pages. Instructions to fill this document are available in the website.

Part A. PERSONAL INFORMATION

CV date

27/10/2023

First name	GERMAN		
Family name	RIGAU CLARAMUNT		
Gender (*)	MALE	Birth date (dd/mm/yyyy)	
Social Security, Passport, ID number			
e-mail	german.rigau@ehu.es	URL Web http://adimen.si.ehu.es/~rigau	
Open Researcher and Contributor ID (ORCID) (*)	H-7235-2015 0000-0003-1119-0930		

(*) Mandatory

A.1. Current position

Position	Profesor Titular de Universidad		
Initial date	05/09/2002		
Institution	Universidad del País Vasco/Euskal Herriko Unibertsitatea		
Department/Center	Lenguajes y Sistemas Informáticos	Facultad de Informática de San Sebastián	
Country	SPAIN	Teleph. number	943015151
Key words	Inteligencia Artificial, Procesamiento del Language Natural		

A.2. Previous positions (research activity interruptions, art. 14.2.b))

Period	Position/Institution/Country/Interruption cause
30/03/2001-04/09/2002	Profesor Titular Universidad / UPC / Spain
01/10/1995-29/03/2001	Profesor Asociado / UPC / Spain
01/07/1992-30/09/1995	Becario FPU

A.3. Education

PhD, Licensed, Graduate	University/Country	Year
Licenciado en Informática	UPC	1991
Doctorado en Inteligencia Artificial	UPC	1998

Part B. CV SUMMARY (max. 5000 characters, including spaces)

Bachelor Degree in Computer Science and Doctor in Artificial Intelligence from the Polytechnic University of Catalonia (UPC). Member of the Department of Computer Languages and Systems at the UPC and member of the NLP research group at the UPC as project fellow, FPU fellow, Assistant Professor and Associate Professor at the UPC and at the TALP center since its creation in 1998. Since the year 2002, Associate Professor at the Department of Computer Languages and Systems at the Computer Science Faculty in San Sebastián and member of the IXA research group of the UPV/EHU. Since its creation in February 2019, deputy director of **HITZ** (Basque Centre for Language Technology) of the UPV/EHU. Since September 2023 National Coordinator of the ESFRI European research infrastructures [CLARIN-ERIC](#) and [DARIAH-ERIC](#).



He published more than 200 journal and conference articles in the area of Natural Language Processing (NLP), and in particular, lexical semantics, semantic processing and inference. Web of Science Thomson Reuters JCR 2022: 9 publications in Q1 quartile or [SCIE-GGS Class 1](#) last 5 years. [Google Scholar Metrics 2022](#): 8182 total citations, average of 604 citations/year last 5 years, Index h: 45. [Ranking of scientists in Spain 2022](#): 2939 of 112000 (top 3%), 7 invited talks in the last 10 years. 3 Research periods recognized (1996-2001, 2002-2007, 2008-2016). Promoter from its inception of multiple tools, semantic resources and open domain NLP applications. All this with the ultimate aim of allowing the understanding of natural language. In particular, the wordnets for Spanish, Catalan, Basque and Galician, and their integration together with other lexical semantic resources in the [Multilingual Central Repository](#). Promoter in collaboration with the LOREA research group of the [AdimenSUMO](#) ontology. Promoter from its inception of [Freeling](#) and [IXA-pipes](#) developments led respectively by Lluís Padrò and Rodrigo Agerri. 2 intellectual property registrations in the last 10 years.

Advisor and co-advisor of 6 PhD theses (4 international) in the last ten years. All of them are currently professors and researchers at universities (UTEM Chile and the University of Arizona) or research centers in Spain (BSC, Vicomtech, Elhuyar). Direction of 14 Final Master's Projects and 23 Final Degree Projects in the last 10 years. Supervisor of a Ramón y Cajal researcher.

Participation in several European research projects (ACQUILEX, ACQUILEX II, EuroWordNet, NAMIC, MEANING, KYOTO, PATHS, OpeNER, READERS, NewsReader, Lihlith, MT4All, ELE, Antidote and ELE2). Principal Investigator (PI) of the European MEANING project (IST-2001-34460), PI of the UPC group in the NAMIC project, and PI of the UPV/EHU group in the European KYOTO, OpeNER, NewsReader, ELE and ELE2 projects. He has also been PI for the SkaTer, TUNER, DeepReading, CrossText and DeepKnowledge national projects. He has also participated in several national research projects (ITEM, HERMES, SENSEM, KNOW and KNOW2). He also lead the strategic research network [INTELE](#) towards the integration of Spain into the [CLARIN](#) and [DARIAH](#) research infrastructures. Currently, he is also leading one of the core partners of the European Language Equality 2 (ELE2) project that continues developing the strategic research, innovation and deployment agenda that will allow to achieve digital language equality in Europe by 2030. He also leads the coordinated national projects DeepKnowledge, DeepR3 and NEL-GAITU. Active participation in the international competitions of SENSEVAL/SEMEVAL and IberEval/IberLEF.

President of the executive committee of the Global WordNet Association (GWA), board member of the European Language Resource Association (ELRA) and member of the permanent committee of the SEPLN. He is also a member of the Association for Computational Linguistics (ACL), the Global WordNet Association (GWA) and the Spanish Society for Natural Language Processing (SEPLN). Member of the program committee and the reviewer of the main international journals, conferences and workshops in the NLP and IA. Evaluator of calls for ANEP and AEI. Participation in the definition of the call CHIST-ERA 2014. Since 2022 Spanish representative of the ESFRI Strategic Working Group on [Social and Cultural Innovation](#).

Advisor since its inception in 2015 of the [Plan de Impulso a las Tecnologías del Lenguaje](#) organized by the Secretary of State for Digital Advancement (SEAD) and co-editor of the "[Informe sobre el estado de las tecnologías del lenguaje en España dentro de la Agenda Digital para España](#)". Member of the working group of the Natural Language Processing Strategy 2020 edited by the SEPLN. Organizer of the 4 editions of the TecnoLing Language Technology exhibitions (2016-2019). PI of 3 contracts with the administration and industry in the last 10 years.

Part C. RELEVANT MERITS (sorted by typology)

C.1. Publications (see instructions)



García-Ferrero I., Altuna B., Álvarez J., González-Agirre I. and Rigau G. [This is not a Dataset: A Large Negation Benchmark to Challenge Large Language Models](#). Empirical Methods for Natural Language Processing (EMNLP 2023). Singapur. 2023. [SCIE-GGS](#) Class 1. Rating A++

García-Ferrero I., Agerri R. and Rigau G. [T-Projection: High Quality Annotation Projection for Sequence Labeling Tasks](#). Findings of the Empirical Methods for Natural Language Processing (EMNLP 2023). Singapur. 2023. [SCIE-GGS](#) Class 1. Rating A++

Pérez N., Cuadros M. and Rigau G. [Negation and speculation processing: A study on cue-scope labelling and assertion classification in Spanish clinical text](#), Artificial Intelligence in Medicine, Volume 145, ISSN 0933-3657. 2023. <https://doi.org/10.1016/j.knosys.2023.110612> JCR 2022: 8.8 **Q1**

Escribano N., Rigau G. and Agerri R. [A modular approach for multilingual timex detection and normalization using deep learning and grammar-based methods](#), Knowledge-Based Systems, Volume 273, ISSN 0950-7051. 2023. <https://doi.org/10.1016/j.artmed.2023.102682> JCR 2022: 7.5 **Q1**

García-Ferrero I., Agerri R. and Rigau G. [Model and Data Transfer for Cross-Lingual Sequence Labelling in Zero-Resource Settings](#). Findings of the Empirical Methods for Natural Language Processing (EMNLP 2022). Abu Dhabi. 2022. [SCIE-GGS](#) Class 1. Rating A++

García-Ferrero I., Agerri R. and Rigau G. [Benchmarking Meta-embeddings: What Works and What Does Not](#). Findings of the Empirical Methods for Natural Language Processing (EMNLP 2021). Punta Cana, Dominican Republic. 2021. [SCIE-GGS](#) Class 1. Rating A++

Zotova E., Agerri R. and Rigau G. [Semi-automatic generation of multilingual datasets for stance detection in Twitter](#). Expert Systems with Applications, Volume 170, 2021, 114547, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.114547>. JCR 2021: 8.665 **Q1** Google Scholar citations: 13.

Álvarez J., Gonzalez-Dios I. and Rigau G. [Applying the Closed World Assumption to SUMO-based FOL Ontologies for Effective Commonsense Reasoning](#). European Conference on Artificial Intelligence (ECAI 2020). Santiago de Compostela, Spain. 2020. [SCIE-GGS](#) Class 2. Rating A-

Perez-Miguel N., Accuosto A., Bravo A., Cuadros M., Martinez-Garcia E., Saggion H. and Rigau G. [Cross-lingual Semantic Annotation of Biomedical Literature: Experiments in Spanish and English](#). Bioinformatics. <https://doi.org/10.1093/bioinformatics/btz853> 2019. JCR 2019: 5.610 **Q1** Google Scholar citations: 11.

Alvez J., Lucio P. and Rigau G. [A Framework for the Evaluation of SUMO-based Ontologies Using WordNet](#). *IEEE Access*, vol. 7, pp. 36075-36093, 2019. <https://doi.org/10.1109/ACCESS.2019.2904835>. JCR 2019: 3.745 **Q1** Google Scholar citations: 8.

Agerri R. and Rigau G. [Language Independent Sequence Labelling for Opinion Target Extraction](#). Artificial Intelligence. ISSN: 0004-3702. Volume 268, Pages 85-95 March 2019, <https://doi.org/10.1016/j.artint.2018.12.002> JCR 2019: 6.628 **Q1** Google Scholar citations: 20.

Garcia-Pablos A., Cuadros M. and Rigau G. [W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis](#). Expert Systems with Applications. Volume 91, pages 127–137, ISSN 0957-4174, 2018. <https://doi.org/10.1016/j.eswa.2017.08.049> JCR 2018: 3.768 **Q1** Citas Google Scholar: 188.

C.3. Research projects

Reference: Real Decreto [958/2022](#), de 15 de noviembre de 2022 **Title:** NEL-GAITU **Financing entity:** Ministerio de Asuntos Económicos y Transformación Digital **Start and end date:** 01/01/2023-31/12/2025 **Grant amount:** 2.000.000€ **Participation:** PI UPV/EHU.



Reference: LANGEQ-2021 PPPA-AG Pilot Project & Preparation Action Grant **Title:** European Language Equality 2 (ELE 2) **Financing entity:** European Commission **Start and end date:** 01/07/2022-30/06/2023 **Grant amount:** 104.000€ **Participation:** PI UPV/EHU.

Reference: LANGEQ-2020 PPPA-AG Pilot Project & Preparation Action Grant **Title:** European Language Equality (ELE) **Financing entity:** European Commission **Start and end date:** 01/01/2021-30/06/2022 **Grant amount:** 124.612€ **Participation:** PI UPV/EHU.

Reference: ICT Call 8 FP7-ICT-2011-8-316404 **Title:** NewsReader: Building structured event indexes of large volumes of financial and economic data for decision making. **Entidad financiadora:** Comisión Europea. FP7 **Start and end date:** 01/01/2013-31/12/2015 **Grant amount:** 480.000€ **Participation:** PI UPV/EHU.

Reference: TED2021-130295B-C31 **Title:** DeepR3: Reducing, Reusing and Recycling large models for developing Responsible and Green Language Technologies **Financing entity:** Ministerio de Economía y Transición Digital. «Proyectos de Transición Ecológica y Transición Digital» **Start and end date:** 01/12/2022-30/11/2024 **Grant amount:** 188.600€ **Participation:** PI UPV/EHU.

Reference: PID2021-127777OB-C21 **Title:** DeepKnowledge: Multilingual Language Model Technology for Language Understanding **Financing entity:** Ministerio de Economía y Transición Digital. Programa investigación fundamental. Subprograma: Retos de la Sociedad. **Start and end date:** 01/09/2022-30/08/2025 **Grant amount:** 243.900€ **Participation:** PI.

Reference: RTI2018-096846-B-C21 **Title:** DeepReading: Mining, Understanding, and Reasoning with Multilingual Content **Financing entity:** Ministerio de Economía y Competitividad. Programa investigación fundamental. Subprograma: Retos de la Sociedad. **Start and end date:** 01/01/2019-31/12/2021 **Grant amount:** 146.894,00€ **Participation:** PI.

Reference: TIN2015-65308-C5-1-R **Title:** TUNER: Multi-faceted Domain Adaptation for Advanced Textual Semantic Processing **Financing entity:** Ministerio de Economía y Competitividad. Programa investigación fundamental. Subprograma: Retos de la Sociedad. **Start and end date:** 01/01/2016-31/12/2018 **Grant amount:** 191.906€ **Participation:** PI.

Reference: RED2018-102797-E **Title:** INTELE: Red estratégica para la promoción de las infraestructuras de tecnologías del lenguaje en eHumanidades y ciencias sociales **Financing entity:** Ministerio de Economía y Competitividad. Programa investigación fundamental. **Start and end date:** 01/01/2020-31/12/2021 **Grant amount:** 23.200€ **Participation:** PI.

C.4. Contracts, technological or transfer merits

Title: PTL: Tecnologías lingüísticas. Estado del arte. **Company:** SECRETARIA DE ESTADO DE TELECOMUNICACIONES Y PARA LA SOCIEDAD DE LA INFORMACIÓN **Start and end date:** 26/06/2015-26/08/2015 **Number of researchers:** 4. **Participation:** PI.

Title: RPA-IA - Plataforma para la generación, entrenamiento y gobierno de robots inteligentes (Programa Hazitek ZL-2019/00198) **Company:** TALIO. **Start and end date:** 1/07/2019-31/12/2019 **Number of researchers:** 3 **Participation:** PI.

Title: Asesoría científica en el área del procesamiento del lenguaje natural en el ámbito de la anonimización. **Company:** NYMIZ **Start and end date:** 01/06/2020-30/11/2020 **Number of researchers:** 1. **Participation:** PI.

Paloma Moreda es Profesora Titular de Universidad desde 2009 en el Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. Su formación es en Ingeniería Informática, y ha adquirido experiencia a lo largo de veinte años en Procesamiento del Lenguaje Natural (PLN), principalmente en el campo del Análisis Semántico. Ha obtenido 2 sexenios de investigación (sexenios CNEAI) y 1 de reconocimiento de transferencia (Número total de sexenios 3).

Su actividad investigadora comenzó en 2001 cuando se incorporó al grupo GPLSI de la Universidad de Alicante, y al que sigue perteneciendo en la actualidad. Su tesis doctoral sobre Análisis Semántico de textos dio lugar a la línea de investigación relacionada con la simplificación y enriquecimiento de textos. Ha publicado más de 40 artículos en congresos internacionales y revistas de reconocido prestigio. Las citas a los artículos publicados pueden consultarse en Scopus (h-index de 7 y 177 citas) en <https://www.scopus.com/authid/detail.uri?authorId=55967519800>, en WOS (h-index de 5 y 78 citss) en <https://publons.com/researcher/2394097/paloma-moreda/>, en Google Scholar (h-index 11 y 527 citas) en <https://scholar.google.es/citations?user=GOncV98AAAAJ&hl=es>, y en DBLP en <https://dblp.org/pid/81/6318.html>. A este respecto cabe destacar las siguientes publicaciones:

- Tania Josephine Martin, José Ignacio Abreu Salas, Paloma Moreda Pozo: A Review of Parallel Corpora for Automatic Text Simplification. Key Challenges Moving Forward. pp 62-78. NLDB (2023)
- Fighting post-truth using natural language processing: A review and open challenges. Saquete, E; Tomas, D.; Moreda, P.; Martínez-Barco, P.; Palomar, M. Expert Systems Applications, 141 (2020).
- DrugSemantics: A corpus for Named Entity Recognition in Spanish Summaries of Product Characteristics. Moreno, I.; Boldrini, E.; Moreda, P.; Romá-Ferri, M.T. Journal of Biomedical Informatics, 72 (2017).
- The use of Natural Language Processing techniques to support Health Literacy: an evidence-based. Moreda, P.; Lloret, E. SYLWAN, 159 (2015).
- An overview of the Applications of Natural Language to Information Systems. Llopis, F.; Martínez-Barco, P.; Patricio Martínez-Barco; Elisabeth Métais; Fernando Llopis; Moreda, P.; Paloma Moreda. Data and Knowledge Engineering 88, (2013).
- Combining Semantic Information in Question Answering Systems. Moreda, P.; Llorens, H.; Saquete, E.; Palomar, M. Information Processing and Management, 47 (2011).
- Combining Semantic Information in Question Answering Systems. Moreda, P.; Llorens, H.; Saquete, E.; Palomar, M. Information Processing and Management, 47 (2011).

Desde su incorporación al grupo de investigación GPLSI, ha participado en más de 25 proyectos de investigación financiados con fondos europeos, fondos públicos nacionales y regionales (3 de ellos como Investigadora Principal). Destaca especialmente:

- La dirección del proyecto europeo "FIRST: A Flexible Interactive Reading Support Tool", que comenzó en 2011 con una duración de 3 años. En este proyecto se trabajó en el desarrollo de diversos prototipos de herramientas para asistir a personas con autismo. El objetivo era adaptar textos a un formato más fácil de comprender.

- La dirección del proyecto nacional "RESCATA: Canonical Representación y Transformaciones Canónicas de Textos Aplicadas a las Tecnologías del Lenguaje Humano" (TIN2015-65100-R). En este proyecto se propuso un método para extender la información proporcionada por representaciones AMR, con el fin de poder representar documentos completos.
- La dirección actualmente, del proyecto financiado mediante fondos Next Generation, "ClearText: Enhancing the modernization public sector organizations by deploying Natural Language Processing to make their digital content CLEARER to those with cognitive disabilities" (TED2021-130707B-I00). En este proyecto se está investigando en cómo abordar el proceso de simplificación de textos en el dominio de la administración pública. El proceso de simplificación es abordado desde la perspectiva de lectura fácil y no como un proceso de simplificación textual sin más.

Además, ha participado en más de 10 contratos de investigación con entidades privadas entre los que destacan:

- Convenio UA Y Diputación Provincial de Alicante para impulsar los procesos de innovación, generación y transferencia de conocimientos y tecnología en el ámbito de la inteligencia artificial, (2020, 2021, 2022 y 2023).
- Estudio de la viabilidad para aplicar técnicas y herramientas informáticas de TLH para minería de textos de la UA en los procesos de negocio de extracción de información y análisis sobre redes sociales y big data (2014-2015).
- Aplicaciones de extracción y recuperación de información especializada del BORME, adaptada a las necesidades solicitadas por ADSALSA publicidad S.L. (2014-2015).

En el apartado de transferencia de tecnología, ha contribuido como autora de 12 desarrollos que han sido inscritos en el Registro de la Propiedad Intelectual como registro de software, y que han sido transferidos mediante licencia de explotación a entidades privadas. En este sentido destacan los registros *Social Analytics Web Interface, Database, Listener and Process v2.0* y *Dossier Next*.

En lo que respecta a las labores docentes, posee actualmente 4 quinquenios. Su docencia se centra en el área de las bases de datos, impartiendo docencia en diferentes cursos y diferentes niveles. Actualmente imparte docencia en el Grado de Ingeniería Informática en primer, segundo y cuarto curso; en el Grado de Ingeniería Biomédica en segundo curso; y en el Grado de Inteligencia Artificial en primer curso. Los contenidos impartidos tienen que ver con los fundamentos, el diseño y la administración de las bases de datos. Además, ha impartido docencia en diferentes másteres, trabajando sobre contenidos relacionados con las bases de datos, así como con el procesamiento del lenguaje natural. Por último destacar que en los últimos 5 años, ha dirigido 13 trabajos de fin de grado.

Comprometida con la igualdad de oportunidades para hombres y mujeres, ha colaborado en la redacción de las Guías de Igualdad de la Universidad de Alicante y ha sido la autora del libro *Guías para una docencia universitaria con perspectiva de género. Ciencias de la Computación*.

En cuanto a las actividades de gestión, ha participado en la dirección del departamento DLSI durante 8 años, 4 como subdirectora y 4 como directora. Por último, también ha formado parte de un gran número de tribunales de tesis doctorales y de puestos docentes.

Part A. PERSONAL INFORMATION		CV date	10/2023
First and Family name	Roberto Centeno Sánchez		
ID number		Age	
Researcher numbers	Researcher ID		
	Orcid code		

A.1. Current position

Name of University/Institution	UNED		
Department	Departamento de Lenguajes y Sistemas Informáticos		
Address and Country	c/ Juan del Rosal 16, Madrid (Spain)		
Phone number	913989696	E-mail	rcenteno@lsi.uned.es
Current position	Profesor Titular de Universidad	From	27/07/2020
Espec. cód. UNESCO	120304 - Inteligencia artificial; 120317 - Informática;		
Palabras clave	Fake news, disinformation, social media, agent-based simulation		

A.2. Education

PhD	University	Year
PhD in Computer Sciences	Universidad Rey Juan Carlos	2012
Master Oficial en Tecnologías de la Información y Sistemas Informáticos	Universidad Rey Juan Carlos	2007
Ingeniero Informático	Universidad Rey Juan Carlos	2006

A.3. JCR articles, h Index, thesis supervised...

- 6 journal papers published in journals indexed by JCR and 6 conference papers indexed by GII-GRIN-SCIE GGS Conference Rating as class 1 and 2.
- 20 conference papers indexed by GII-GRIN-SCIE GGS Conference Rating as class 3 and CORE B or C.
- 1 published book and 4 book chapters
- 16 papers published in international and national conferences and workshops (5 indexed as B by CORE Ranking and 6 as C by CORE Ranking)
- h-Index: 13 (<https://scholar.google.com/citations?user=R2kzy9EAAA&hl=en>)

Part B. CV SUMMARY (max. 3500 characters, including spaces)

Research

- Positive evaluation of 2 "sexenios": 2007-2014 and 2015-2020
- More than 16 years of experience (not counting the years as a pre-licensed scholar collaborating in research projects):
 - (2006-2010) predoctoral scholar in the URJC
 - (2010-2015) teaching assistant at the UNED
 - (2015-2018) assistant professor at the UNED
 - (2018-2020) associate professor at the UNED
 - (2020-current) senior lecturer at the UNED
- Participation in several research projects: 2 European, 12 national, 4 regional and 2 contracts with companies.
- Attendance and lecture at various congresses in the area
- Reviewer at international congresses and journals
- Lecturer of several seminars
- Supervisor of 4 PhD student in progress

Teaching

- More than 16 years of experience:

- 2006/2007 to 2010/2011, as a predoctoral scholar in the URJC, collaborating in first-second cycle teachings and non-regulated teachings.
- 2010/2011 - currently, as teaching assistant, assistant professor and associate professor in the UNED, teaching in undergraduate and postgraduate courses.
- supervisor of more than 20 undergraduate thesis project and and 7 master's thesis
- participation in 14 projects of teaching innovation.
- Member of the teaching innovation group: Intelligent Systems for Learning

Education

- PhD in computer sciences with European mention, by URJC, qualification of "apto-cum laude" under the program "technologies of the information and informatic systems" with mention of quality
- official master's degree in information technology and computer systems at the URJC. Mark of 9,75 out of 10
- computer engineer at the URJC. Mark of 8,37 out of 10
- technical engineer in computer science of management at the URJC. Note of 2,24 out of 4
- research scholarship (FPI program) at the URJC
- 2-month postdoctoral research stage at the university of Essex - Colchester (UK)
- predoctoral research stage of 5 months at City University - London (UK)
- 2 predoctoral research stages in national institutions (UPV and IIIA-CSIC)

Part C. RELEVANT MERITS

C.1. Publications (including books)

[JCR Journals]

1. Roberto Centeno; Víctor Fresno; Jacobo Chaquet. **From textual reviews to Individual Reputation Rankings: Leaving ratings aside solving MPC Task.** *Expert Systems with Applications*. Vol. 114, pp. 1-14. Elsevier, 2018. ISSN 0957-4174
Índice de impacto (JCR 2017): 3.768
Posición de la revista: Computer Science, Artificial Intelligence: 20/132 (Q1, T1)
2. Roberto Centeno; Ramón Hermoso. **Estimating global opinions by keeping users from fraud in online review systems.** *Knowledge and Information Systems*. Vol. 55(2), pp. 467-491. Springer, 2018. ISSN 0219-1377
Índice de impacto (JCR 2017): 2.247
Posición de la revista: Computer Science, Artificial Intelligence: 49/132 (Q2, T2)
3. Roberto Centeno; Ramón Hermoso; Maria Fasli. **On the inaccuracy of numerical ratings: dealing with biased opinions in social networks.** *Information Systems Frontiers*. Vol. 17(4), pp. 809-825. Springer, 2015. ISSN 1572-9419
Índice de impacto (JCR 2015): 1.450
Posición de la revista: Computer Science, Information Systems: 54/144 (Q2, T2)
4. Ramón Hermoso; Roberto Centeno; Maria Fasli. **From blurry numbers to clear preferences: A mechanism to extract reputation in social networks.** *Expert Systems with Applications*. Vol. 41(5), pp. 2269-2285. Elsevier, 2014. ISSN 0957-4174
Índice de impacto (JCR 2014): 2.240
Posición de la revista: Computer Science, Artificial Intelligence: 29/123 (Q1, T1)
5. Roberto Centeno; Holger Billhardt; Ramón Hermoso. **Persuading agents to act in the right way: an incentive-based approach.** *Engineering Applications of Artificial Intelligence*. Vol. 26(1), pp. 198-210. Elsevier, 2013. ISSN 0952-1976
Índice de impacto (JCR 2013): 1.962
Posición de la revista: Engineering, Multidisciplinary: 15/87 (Q1, T1)
6. Sergio Esparcia; Roberto Centeno; Ramón Hermoso; Estefanía Argente. **Enhancing MAS Environments with Organizational Mechanisms.** *International Journal on Artificial Intelligence Tools*. Vol. 20 (4), pp. 663-690. Ed. World Scientific, 2011. ISSN 0218-2130
Índice de impacto (JCR 2011): 0.217
Posición de la revista: Computer Science, Artificial Intelligence: 105/111 (Q4, T3)

[International Conference]

1. Roberto Centeno; Ramón Hermoso. **When opinion request meets majority search: avoiding fraud in on-line review systems.** *The 14th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2015)* pp. 1805-1806. IFAAMAS, 2015. ISBN 978-1-4503-3413-6
 - A* by CORE Ranking
 - Class 1 SCIE-GGS Ranking (Rating A+)



2. Roberto Centeno; Holger Billhardt. **Adaptive regulation of open MAS: an incentive mechanism based on online modifications of the environment.** *The 10th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2011)*. Vol. 3, pp. 1243-1244. IFAAMAS, 2011. ISBN 0-9826571-8-8
 - A* by CORE Ranking
 - Class 1 SCIE-GGS Ranking (Rating A+)
3. Roberto Centeno; Holger Billhardt. **Using Incentive Mechanisms for an Adaptive Regulation of Open Multi-Agent Systems.** *The 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*. pp. 139-145. AAAI PRESS, 2011. ISBN 978-1-57735-516-8
 - A* by CORE Ranking
 - Class 1 SCIE-GGS Ranking (Rating A++)
4. Roberto Centeno; Holger Billhardt; Ramón Hermoso; Sascha Ossowski. **Organising MAS: A Formal Model Based on Organisational Mechanism.** *24th Annual ACM Symposium on Applied Computing (ACM SAC 2009)*. pp. 740-746. ACM NEW YORK, 2009. ISBN 978-1-60558-166-8
 - B by CORE Ranking
 - Class 2 SCIE-GGS Ranking (Rating A-)

C.2. Research projects and grants

1. **Holistic Analysis of Organised Misinformation Activity in Social Networks (HAMiSoN)**
CHIST-ERA (Foundations for Misbehavior Detection and Mitigation Strategies in Online Social Networks and Media – OSNEM)
Participants: UNED (coordinator), ZHAW University, University of Tartu and Synapse
IPs: Anselmo Peñas.
Duration: 1/1/2023 – 31/12/2025
2. **Modelos de Lenguaje Profundos en la Interpretación de la Información y la Desinformación en Contexto (DeepInfo)**
Ministerio de Economía y Competitividad: Proyecto Retos I+D+i 2021 (PID2021-127770B-C22)
Participants: Universidad País Vasco y UNED
IPs: Anselmo Peñas y Álvaro Rodrigo
Participation: Equipo de trabajo
Duration: 1/9/2022 – 31/08/2025
3. **Mining, Understanding, and Reasoning with Multilingual Content: DeepReading**
Ministerio de Economía y Competitividad: Proyecto Retos I+D+i 2021 (RTI2018-096846-B-C21)
Participants: Universidad País Vasco y UNED
IPs: German Rigau
Participation: Equipo de trabajo
Duration: 1/1/2019 – 31/12/2021
4. **Coordinación dinámica de flotas abiertas en entornos urbanos (SURF)**
Ministerio de Economía y Competitividad: Proyecto Retos I+D+i 2018
Participants: Universidad Politécnica de Valencia, Universidad Rey Juan Carlos y Universidad de Salamanca.
IPs: Alberto Fernández y Holger Billhardt (URJC)
Participation: Equipo de trabajo
Duration: 1/1/2016 – 31/12/2018
5. **Red eMadrid: Investigación y desarrollo de tecnologías educativas en la Comunidad de Madrid (eMadrid-2)**
Comunidad de Madrid (S2013/ICE-2715)
Participants: Universidad Carlos III de Madrid, Universidad Nacional de Educación a Distancia, Universidad Politécnica de Madrid, Universidad Complutense de Madrid, Universidad Rey Juan Carlos y Universidad Autónoma de Madrid.
IP: Miguel Rodríguez Artacho (UNED)
Participation: Full time researcher
Duration: 1/10/2014 – 30/09/2018
6. **Sociedades Humano-Agente: Descripción, Gestión y Regulación (iHAS)**
Ministerio de Economía y Competitividad (TIN2012-36586-C03-02)
Participants: Universidad Politécnica de Valencia, Universidad Rey Juan Carlos y Universidad de Salamanca
IP: Holger Billhardt (URJC)
Participation: half time researcher
Duration: 1/12/2012 – 30/11/2015
7. **Organizaciones Virtuales Adaptativas: Técnicas y Mecanismos de Descripción y Adaptación (OVAMAH)**
Ministerio de Ciencia e Innovación (TIN2009-13839-C03-02)
Participants: Universidad Politécnica de Valencia, Universidad Rey Juan Carlos y Universidad de Salamanca
IP: Alberto Fernández Gil (URJC)
Participation: full time researcher
Duration: 1/1/2010 – 31/12/2012



8. Agreement Technologies (AT-COST-Action)

Unión Europea. Cost Action (IC0801)
Participants: Universidad Rey Juan Carlos and several international institutions
IP: Sascha Ossowski (URJC)
Participation: full time researcher
Duration: 17/10/2008 – 16/10/2012

9. Tecnologías del Acuerdo (AT)

Ministerio de Innovación, Ciencia y Tecnología. Consolider INGENIO 2010 (CSD2007-0022)
Participants: Instituto de Investigación e Inteligencia Artificial (IIIA-CSIC), Universidad Politécnica de Valencia, Universidad Rey Juan Carlos
IP: Sascha Ossowski (URJC)
Participation: full time researcher
Duration: 1/8/2007 – 29/11/2012

10. Métodos, Técnicas y Herramientas para Sistemas MultiAgente Abiertos – Coordinación, Organización, Institución (THOMAS)

Ministerio de Ciencia e Innovación (TIN2006-14630-C03-02)
Participants: Universidad Politécnica de Valencia, Universidad Rey Juan Carlos y Universidad de Salamanca
IP: Sascha Ossowski (URJC)
Participation: Investigador a tiempo completo
Duration: 1/10/2006 – 30/09/2009

C.3. Contracts

1. MIO! – Tecnologías para prestar servicios en movilidad en el futuro universo inteligente (MIO!)

Ministerio de Industria, Turismo y Comercio (CENIT-2008-1019)
Participants: Universidad Rey Juan Carlos, Intelligent Software Components, SA
IP: Sascha Ossowski (URJC)
Participation: full time researcher
Duration: 1/3/2008 – 31/12/2011

2. Museos 2.0

Ministerio de Industria, Turismo y Comercio. Plan Avanza I+D (TSI-020302-2008-44)
Participants: Universidad Rey Juan Carlos, Fundacion European Virtual Engineering
IP: Sascha Ossowski (URJC)
Participation: full time researcher
Duration: 1/1/2008 – 31/12/2010

C.4. Reviewer and Program Committee

- Organizer of Workshop on NLP applied to MisInformation @ SEPLN 2023.
- XIII Conferencia de la Asociación Española Para la Inteligencia Artificial (CAEPIA - TTIA 2009)
- 2009 International Workshop on Coordination, Organizations, Institutions and Norms in Agent Systems V at International Joint Conference on Artificial Intelligence (COIN@IJCAI 2009)
- The 2009 IEEE/WIC/ACM International Conferences on Intelligent Agent Technology (IAT 2009)
- The 9TH International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)
- The 12th Ibero-American Conference On Artificial Intelligence (IBERAMIA 2010)
- 8th German Conference on MultiAgents System Technologies (MATES 2010)
- Student Session on The European Agent Systems Summer School. Ediciones: 2011, 2012 y 2013
- Special Session TINMAS On The 11th Conference on Practical Applications of Agents and Multi-Agent Systems (TINMAS@PAAMS 2013)
- Workshop on Intelligent Systems and Applications de la Iberian Conference on Information Systems and Technologies. 2013, 2014, 2015, 2016
- Workshop on Intelligent Human-Agent Societies on the 13th Conference on Practical Applications of Agents and Multi-Agent Systems (WiHAS@PAAMS 2015)
- Workshop on Smart Cities and Intelligent Agents on the Conference on Practical Applications of Agents and MultiAgent Systems (WOSCIA@PAAMS). 2016; 2017; 2018
- Reviewer of International journals and events such as: ACL2023, IPM journal, coling 2022, expert systems, IJCAI, AAMAS,...

C.5. Grants

1. Research scholarship (FPI Grant)

Ministerio de Ciencia e Innovación.
Grupo de Inteligencia Artificial, Universidad Rey Juan Carlos
Dates: 01/07/2007 - 30/11/2010)

C.6 PhD supervision

- **Currently supervising 4 PhD students**
- **I have already supervised 3 PhD students (PhD not finished but with relevant publications).**

SAQUETE BORO, MARIA ESTELA

Situación profesional actual

Plaza:

PROFESOR/A TITULAR UNIVERSIDAD

Dpto.:

LENGUAJES Y SISTEMAS INFORMATICOS

Institutos:

I.U. INVESTIGACION INFORMATICA

Grupos:

Procesamiento del lenguaje y sistemas de información(GPLSI)

Formación académica

DOCT. INGENIERÍA LINGÜÍSTICA, APRENDIZAJE AUTOM. Y RECONOCIMIENTO DE FORMAS

Centro de Estudios de Doctorado y Posgrado (14/06/2005)

Doctora en Ingeniería Informática

Universidad de Alicante (14/06/2005)

Biografía

Actualmente soy profesora Titular de Universidad de la Universidad de Alicante. Mi actividad docente e investigadora se inicia en el año 2002 con la incorporación a la universidad, donde entro a formar parte del grupo de investigación "Procesamiento del lenguaje y sistemas de información" que dirige el Dr. Manuel Palomar Sanz, y al que sigo perteneciendo en la actualidad. En ese momento empiezo un primer trabajo de investigación centrado en el tratamiento de la información temporal. Este trabajo inicial fue el inicio de la línea de investigación sobre la extracción y resolución de información temporal aplicada a los sistemas de Búsqueda de respuestas que se culminó con la presentación de la tesis doctoral en 2005. A raíz de este trabajo, he dirigido una tesis doctoral centrado en el procesamiento de la información temporal que fue presentado en 2011. Fruto de toda esta actividad, destaca mi participación en 29 publicaciones en revistas de alto impacto, indexadas en JCR o con criterios de calidad equivalentes y 24 ponencias en congresos internacionales de prestigio en el ámbito del procesamiento del lenguaje natural. También sobre estas temáticas he participado como investigador en 37 proyectos de investigación financiados con fondos públicos europeos, nacionales y autonómicos (4 de ellos como IP), y 5 contratos de investigación con entidades privadas (1 de ellos como IP). En el apartado de transferencia tecnológica, he contribuido como autora de 3 desarrollos que han sido inscritos en el Registro de la Propiedad Intelectual como registro software, y que han podido ser transferidos mediante licencia de explotación a entidades privadas, encontrándose el resto a disposición de la comunidad investigadora para su aprovechamiento. Así mismo, destaca la realización de tres estancias en centros de investigación internacionales, una de ellas de tres meses en la Schools of Informatics de la Universidad de Edimburgo (Reino Unido), en 2004, la segunda en la Brandeis University de Boston (USA) de 2 meses y la última en Cardiff University de Gales por 1 mes. Finalmente me

gustaría destacar mi contribución como revisora de manuscritos en diversas ocasiones para revistas de prestigio indexadas en JCR o con indicios objetivos de calidad semejantes.

Experiencia docente

Quinquenios docentes: **4**

Número de trabajos fin de grado/máster dirigidos o codirigidos (últimos 5 años): **70**

Investigación

Sexenios de investigación: **4**

Tesis doctorales dirigidas o codirigidas (totales): **3**

Publicaciones en revistas: (últimos 5 años): **14**

Capítulos de libro publicados: (últimos 5 años): **1**

Participación en proyectos de investigación: (últimos 5 años): **15**

Comunicaciones a congresos: (últimos 5 años): **5**

Gestión

Cargo:	Fecha inicio:	Fecha fin:
SECRETARIO/A DE DEPARTAMENTO TIPO 3 LENGUAJES Y SISTEMAS INFORMATICOS	25/03/2021	Vigente
COORDINADOR/A ACADEMICO/A RELACIONES INTERNACIONALES Y COOPER	05/07/2013	09/01/2014
COORDINADOR/A DE CENTRO RELACIONES INTERNACIONALES Y COOPER	16/05/2013	04/07/2013
COORDINADOR/A DE CENTRO COMUNICACION	01/10/2009	15/05/2013

PUBLICACIONES EN REVISTAS (DESTACADAS)

- **Título:** A semi-automatic annotation methodology that combines Summarization and Human-In-The-Loop to create disinformation detection resources
Autores: Bonet-Jover, A.; Sepulveda, R; Saquete, E.; Martínez-Barco, P.
Revistas: Knowledge-Based Systems
Volumen: 275
Páginas: -
Fecha: 2023
ISSN: 0950-7051
DOI: <https://doi.org/10.1016/j.knosys.2023.110723>
- **Título:** RUN-AS: a novel approach to annotate news reliability for disinformation detection.
Autores: Bonet-Jover, A.; Sepulveda, R; Saquete, E.; Martínez-Barco, P.; Nieto Pérez, Mario
Revistas: Language Resources and Evaluation
Volumen: -
Páginas: -
Fecha: 2023
ISSN: 1574-020X
DOI: <http://dx.doi.org/10.1007/s10579-023-09678-9>
- **Título:** Detecting misleading headlines through the automatic recognition of contradiction in Spanish
Autores: Sepulveda, R; Bonet-Jover, A.; Saquete, E.
Revistas: IEEE access
Volumen: -
Páginas: -
Fecha: 2023
ISSN: 2169-3536
DOI: <http://dx.doi.org/10.1109/ACCESS.2023.3295781>
- **Título:** Why are some social-media contents more popular than others? Opinion and association rules mining applied to virality patterns discovery
Autores: Saquete, E.; Zubcoff, Jose; Gutiérrez, Y.; Martínez-Barco, P.; Fernández, J.
Revistas: Expert Systems with Applications
Volumen: -
Páginas: -
Fecha: 2022
ISSN: 0957-4174
DOI: <https://doi.org/10.1016/j.eswa.2022.116676>
- **Título:** HeadlineStanceChecker: Exploiting summarization to detect headline disinformation
Autores: Sepulveda, R; Vicente, M.; Saquete, E.; Lloret, E.; Palomar, M.
Revistas: Web Semantics
Volumen: in press

Páginas: -

Fecha: 2021

ISSN: 1570-8268

DOI: <https://doi.org/10.1016/j.websem.2021.100660>

PARTICIPACIÓN EN PROYECTOS DE INVESTIGACIÓN (DESTACADOS)

- **Denominación del proyecto:** NL4DISMIS: Natural Language Technologies for dealing with dis- and misinformation
Competitivo: Sí
Europeo: No
Público: Sí
Entidad/es financiadora/s: GENERALITAT VALENCIANA
Fecha de inicio: 01/01/2022
Fecha de finalización: 31/12/2025
Tipo participación: Coordinador-a
- **Denominación del proyecto:** Recursos tecnológicos para análisis viral inteligente mediante PLN
Competitivo: Sí
Europeo: No
Público: Sí
Entidad/es financiadora/s: MINISTERIO DE CIENCIA E INNOVACION
Fecha de inicio: 01/09/2022
Fecha de finalización: 31/08/2025
Tipo participación: Coordinador-a
- **Denominación del proyecto:** Análisis de la confiabilidad en medios digitales
Competitivo: Sí
Europeo: No
Público: Sí
Entidad/es financiadora/s: MINISTERIO DE CIENCIA E INNOVACION
Fecha de inicio: 01/12/2022
Fecha de finalización: 30/11/2024
Tipo participación: Coordinador-a
- **Denominación del proyecto:** Modelado del comportamiento de entidades digitales mediante Tecnologías del Lenguaje Humano
Competitivo: Sí
Europeo: No
Público: Sí
Entidad/es financiadora/s: MINISTERIO DE ECONOMIA Y EMPRESA
Fecha de inicio: 01/01/2019
Fecha de finalización: 31/12/2021
Tipo participación: Coordinador-a
- **Denominación del proyecto:** 14th Symposium on Temporal Representation and Reasoning

Competitivo: Sí

Europeo: No

Público: Sí

Entidad/es financiadora/s: GENERALITAT VALENCIANA

Fecha de inicio: 28/06/2007

Fecha de finalización: 30/06/2007

Tipo participación: Coordinador-a

Fecha del CVA	27/10/2023
----------------------	------------

Parte A. DATOS PERSONALES

Nombre y apellidos	ISABEL SEGURA BEDMAR		
DNI/NIE/pasaporte			
Núm. identificación del investigador	Researcher ID	L-6077-2017	
	Código Orcid	orcid.org/0000-0002-7810-2360	

A.1. Situación profesional actual

Organismo	UNIVERSIDAD CARLOS III DE MADRID		
Dpto./Centro	DEPARTAMENTO DE INFORMÁTICA		
Dirección	AVD. DE LA UNIVERSIDAD, 30, LEGANÉS, 28911, MADRID		
Correo electrónico	ISEGURA@INF.UC3M.ES		
Categoría profesional	PROFESOR TITULAR	Fecha inicio	15/02/2004
Espec. cód. UNESCO	120317 Informática, 120304 Inteligencia Artificial		
Palabras clave	PROCESAMIENTO DE LENGUAJE NATURAL, EXTRACCIÓN DE INFORMACIÓN, INTELIGENCIA ARTIFICIAL, DEEP LEARNING		

A.2. Formación académica (título, institución, fecha)

Licenciatura/Grado/Doctorado	Universidad	Año
LICENCIADA EN CIENCIAS MATEMÁTICAS, CIENCIAS DE LA COMPUTACIÓN	UNIVERSIDAD COMPLUTENSE DE MADRID	1998
DOCTORA EUROPEA EN CIENCIA Y TECNOLOGÍA INFORMÁTICA	UNIVERSIDAD CARLOS III DE MADRID	2010

A.3. Indicadores generales de calidad de la producción científica (véanse instrucciones)

- Tesis doctorales dirigidas últimos 10 años: 2 tesis, 1 con premio extraordinario de doctorado
- Publicaciones Q1: 15; Q2: 7; Core A: 2.
- Índice H (según Google Scholar): 24, índice i10: 40, Citas: 3030.
- Índice H (web of science): 14, citas: 986.
- Índice H (semanticscholar): 22, citas: 2468, highly influential citations: 166.

Parte B. RESUMEN LIBRE DEL CURRÍCULUM (máximo 3500 caracteres, incluyendo espacios en blanco)

Profesora Titular del Departamento de Informática de la Universidad Carlos III de Madrid (UC3M) e investigadora en el Grupo HULAT (<http://hulat.inf.uc3m.es>). Su actividad investigadora se centra en la disciplina del Procesamiento del Lenguaje Natural (PLN), con un especial foco en el uso de técnicas de aprendizaje automático y su aplicación al dominio biomédico. En este ámbito, ha publicado más de 100 publicaciones científicas en revistas y congresos indexados. las cuales 14 son del primer cuartil, 6 del segundo y 2 artículos en conferencias core A. Su tesis obtuvo el Premio Extraordinario de Doctorado de la UC3M en 2010, y en 2011, el Premio SEPLN, otorgado por la Sociedad Española de Procesamiento de Lenguaje Natural.

También ha participado en seis proyectos I+D en convocatorias competitivas en el campo del PLN, y en el proyecto europeo TrendMiner del 7º Programa Marco. Ha codirigido el del proyecto nacional DeepEMR: Extracción de información clínica usando deep learning y técnicas de Big Data (TIN2017-87548-C2-1-R) en el periodo 2018-2020. También fue investigador principal del proyecto NLP4RARE-CM-UC3M, concedido en la Convocatoria del Programa de Apoyo a la Realización de Proyectos Interdisciplinarios de I+D para jóvenes investigadores de la UC3M 2019-2020.

Ha dirigido dos tesis con mención internacional, y una de ellas la recibió, el premio extraordinario de doctorado de la UC3M en la convocatoria del 2015. En la actualidad, dirige otra tesis, cuyo objetivo es la investigación en métodos semisupervisados basados en arquitecturas de deep learning para tareas de PLN.

Ha sido editora del Special Issue “Mining the pharmacovigilance literatura” en la revista Journal of Biomedical Informatics, y colaboradora destacada de la revista Journal of Biomedical Informatics, así



como revisora habitual en revistas de alto impacto como *Bioinformatics*, *Drug Discovery Today*, *Plus One*, entre otras. Así mismo, ha organizado varios eventos científicos y competiciones como *SemEval-DDIExtraction 2013*, *DDIExtraction 2011*, *BioSEPLN 2010*, y participado en otras como: *DTMBIO 2011* y *SEPLN 2008*. Además, he colaborado activamente como revisora de revistas de alto factor de impacto (*Bioinformatics*, *JAMIA*, *BMC Bioinformatics*, etc) y como miembro de comités científicos de conferencias y workshops en el campo de PLN. Ha realizado una estancia en el Centro Alemán de Inteligencia Artificial (DFKI) durante 6 meses en 2009, y otra estancia de 3 meses en 2008 en el grupo de PLN de la Universidad Politécnica de Valencia, financiadas ambas con ayudas competitivas por parte del Ministerio de Educación y del programa propio de la UC3M. Más información en mi página web <http://hulat.inf.uc3m.es/nosotros/miembros/iseкура>.

Parte C. MÉRITOS MÁS RELEVANTES (ordenados por tipología)

C.1. Publicaciones

1. Claudia Martínez-de Miguel, Isabel Segura-Bedmar, Esteban Chacón-Solano, Sara Guerrero-Aspizua (2022): The RareDis corpus: A corpus annotated with rare diseases, their signs and symptoms. *J. Biomed. Informatics* 125: 103961 (**JCR Science, Impact factor: 4.5, Q2**).
2. Isabel Segura-Bedmar, David Camino-Perdonas, Sara Guerrero-Aspizua (2022): Exploring deep learning methods for recognizing rare diseases and their clinical manifestations from texts. *BMC Bioinform.* 23(1): 263. (**JCR Science, Impact factor: 3.307, Q2**).
3. Isabel Segura-Bedmar, Santiago Alonso-Bartolome (2022): Multimodal Fake News Detection. *Inf.* 13(6): 284 (**JCR Science, Impact factor: 3.1, Q2**).
4. Jose Antonio Jijon-Vorbeck, Isabel Segura-Bedmar (2021): Exploring the Impact of COVID-19 on Social Life by Deep Learning. *Inf.* 12(11): 459.
5. Cristóbal Colón-Ruiz, Isabel Segura-Bedmar (2020). Comparing deep learning architectures for sentiment analysis on drug reviews. *Journal of Biomedical Informatics.* (**JCR Science, Impact factor: 3.526**).
6. Isabel Segura-Bedmar, Pablo Raez (2019). Cohort selection for clinical trials using deep learning models. *Journal of the American Medical Informatics Association*, 26(11), 1181-1188 (**JCR Science, Impact factor: 4.112, Q1**).
7. Victor Suárez-Paniagua, Renzo Zavala, Isabel Segura-Bedmar, Paloma Martínez (2019). A two-stage deep learning approach for extracting entities and relationships from medical texts. *Journal of biomedical informatics*, 99. (**JCR Science, Impact factor: 3.526, Q2**).
8. Isabel Segura-Bedmar, Cristóbal Colón-Ruiz, Miguel A. Tejedor-Alonso, Mar Moro-Moro (2018). Predicting of anaphylaxis in big data EMR by exploring machine learning approaches. *Journal of biomedical informatics*, 87, 50-59. (**JCR Science, Impact factor: 3.526, Q2**).
9. Victor Suárez-Paniagua, Isabel Segura-Bedmar, I. (2018). Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction. *BMC bioinformatics*, 19(8), 39-47 (**JCR Science, Impact factor: 3.241, Q1**).
10. Victor Suarez-Paniagua, Isabel Segura-Bedmar, Paloma Martínez (2017). Exploring Convolutional Neural Networks for Drug-Drug Interaction Extraction, *Database. The Journal of Biological Databases and Curation*, January, 2017, ISSN: 1758-0463, Volume 2017, Issue 1. (**JCR Science , Impact factor: 2.627, Q1**).
11. Isabel Segura-Bedmar, Antonio Quiros, Paloma Martínez, (2017). Exploring Convolutional Neural Networks for Sentiment Analysis of Spanish tweets, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*; Valencia, Spain, April, 2017, ACL, Volume 1, Long Papers, pp: 1014-1022, (**Core A**).
12. María Herrero Zazo, Isabel Segura-Bedmar, Paloma Martínez, (2016). Conceptual Models of Drug-Drug Interactions: A Summary of Recent Efforts, December, 2015, *Knowledge-Based Systems*, 114, pp: 99-107, (**JCR Science, Impact Factor 2015: 3.325, Q1**).
13. Paloma Martínez, José Luis Martínez Fernández, Isabel Segura-Bedmar, Julián Moreno-Schneider, Ricardo Revert , Adrián Luna, (2016). Turning user generated health-related content into actionable knowledge through text analytics services, May, 2016, *Computers in Industry*, (**JCR Science, Impact Factor 2015: 1.685, Q2**).
14. Victor Suarez-Paniagua, Isabel Segura-Bedmar, (2016). Using Recursive Neural Networks to detect and classify drug-drug interactions from biomedical texts, *22nd European Conference on Artificial Intelligence (ECAI 2016)*, The Hague, Holland, August, 2016. (**CORE A**)
15. María Herrero Zazo, Isabel Segura-Bedmar, Janna Hastings, Paloma Martínez, (2015). DINTO: Using OWL ontologies and SWRL rules to infer drug-drug interactions and their mechanisms (DOI: 10.1021/acs.jcim.5b00119), June, 2015, *Journal of Chemical Information and Modeling*, ACS. (**JCR Science, Impact Factor 2014: 3,738, Q1**).
16. Isabel Segura-Bedmar, Paloma Martínez, Ricardo Revert , Julián Moreno-Schneider, (2015). Exploring Spanish Health Social Media for detecting drug effects,

- June, 2015, BMC Medical Informatics and Decision Making, **(JCR Impact Factor 2015: 2.042, Q2)**.
17. Isabel Segura-Bedmar, Paloma Martínez, (2015). Pharmacovigilance through the Development of Text Mining and Natural Language Processing Techniques, December, 2015, Journal of Biomedical Informatics, Elsevier, 58, pp: 288-291. **(JCR Science, Impact Factor (2014): 2,194, Q1)**
 18. Isabel Segura-Bedmar, Paloma Martínez, María Herrero Zazo, (2014). Lessons learnt from the DDIExtraction-2013 shared task, January, 2014, Journal of Biomedical Informatics, Elsevier, 51, pp: 152-164. **(JCR Science, Impact Factor 2014: 2,194, Q1)**.
 19. Isabel Segura-Bedmar, Paloma Martínez, (2014). Special Issue on Mining the Pharmacovigilance Literature, Journal of Biomedical Informatics, 49, **(JCR Science, Impact Factor (2014): 2,194, Q1)**.
 20. César de Pablo-Sánchez, Isabel Segura-Bedmar, Paloma Martínez, Ana Iglesias, (2013). Lightly supervised acquisition of named entities and linguistic patterns for multilingual text mining, April, 2013, Knowledge and Information Systems Journal, Springer London, 35, 1, pp: 87-109. **(JCR Science, Impact Factor (2013): 2,639, Q1)**.
 21. María Herrero Zazo, Isabel Segura-Bedmar, Paloma Martínez, Thierry Declerck, (2013). The DDI corpus: An annotated corpus with pharmacological substances and drug-drug interactions, Journal of Biomedical Informatics, 46(5), pp: 914-920 **(JCR Science, Impact Factor (2013): 2.482, Q1)**.
 22. Isabel Segura-Bedmar, Paloma Martínez, César de Pablo-Sánchez, (2011). A linguistic rule-based approach to extract drug-drug interactions from pharmacological documents, March, 2011, BMC Bioinformatics, 12(Suppl 2):S1. **(JCR Science, Impact Factor 2011: 2,751, Q1)**.
 23. Isabel Segura-Bedmar, Paloma Martínez, César de Pablo-Sánchez, (2011). Using a Shallow Linguistic Kernel for Drug-Drug Interaction Extraction, October, 2011, Journal of Biomedical Informatics, Volume 44, Issue 5, pp: 789 – 804. **(JCR Science, Impact Factor (2011): 1,792, Q2)**.
 24. Isabel Segura-Bedmar, Paloma Martínez, César de Pablo-Sánchez, (2010). Extracting drug-drug interactions from biomedical texts., May, 2010, BMC Bioinformatics, 11, Suppl 5. **(JCR Science, Impact Factor (2010): 3,029, Q1)**.
 25. Isabel Segura-Bedmar, Mario Crespo, César de Pablo-Sánchez, Paloma Martínez, (2010). Resolving anaphoras for the extraction of drug-drug interactions in pharmacological documents. , April, 2010, BMC Bioinformatics, 11, (Suppl 2). **(JCR Science, Impact Factor (2010): 3,029, Q1)**.
 26. Roxana Danger, Isabel Segura-Bedmar, Paloma Martínez, Paolo Rosso, (2010). A comparison of machine learning techniques for detection of drug target articles, December, 2010, Journal of Biomedical Informatics, 43, Issue 6, pp: 902-913. **(JCR Science, Impact Factor (2010): 1,724, Q2)**.
 27. Isabel Segura-Bedmar, Paloma Martínez, María Segura-Bedmar, (2008). Drug Name Recognition and classification in biomedical texts. Drug Discovery Today, 13, 17-18, pp: 816-823. **(JCR Science, Impact Factor (2008): 6,618, Q1)**

C.2. Proyectos

Referencia: PID2020-116527RB-I00

Título: Accesibilidad cognitiva y sensorial a los sistemas de videoconferencia (ACCESS2MEET)

Participación: Miembro equipo investigador

Entidad Financiadora: Convocatoria 2020 Proyectos Generación del Conocimiento y Retos Investigación. MINISTERIO DE CIENCIA E INNOVACIÓN

Duración : 2021 - 2024

Estado: concedido

Referencia: TED2021-132182A-I00

Título: Plataforma para la Impartición de Clases Online y Generación de Apuntes ACcesibles para personas con discapacidad sensorial (PICOGAC)

Participación: Miembro equipo investigador

Entidad Financiadora: Proyectos de Transición Ecológica y Transición Digital. Ministerio de Ciencia e Innovación

Duración : 2022 - 2024

Estado: concedido

Referencia: NLP4Rare-cm-uc3m

Título: Natural Language Processing for Rare Diseases

Participación: Investigador Principal

Entidad Financiadora: Universidad Carlos III de Madrid y Comunidad de Madrid

Duración : 1/1/2020-31/12/2021



Financiación: 59.972 €,
Estado: concedido

Referencia: TIN2017-87548-C2-1-R
Título: DeepEMR: Extracción de información clínica usando deep learning y técnicas de Big Data
Participación: Investigador Principal
Entidad Financiadora: Ministerio de Economía y competitividad
Duración : 1/1/2018-30/12/2020
Financiación: 106.000 €,
Estado: concedido

Referencia: TIN2014-52665-C2-2-R
Título: eGovernAbility-AccessSupport: Marco basado en modelos para el desarrollo de servicios accesibles en la e-Administración
Participación: Miembro equipo investigador
Entidad Financiadora: Ministerio de Economía y competitividad
Duración : 1/1/2015-30/12/2017
Financiación: 86.000 €,
Estado: concedido

Referencia: FP7-ICT 287863
Título: TRENDMINER: Large-scale, Cross-lingual Trend Mining and Summarisation of Real-time Media Streams
Participación: Miembro equipo investigador
Entidad Financiadora: Unión Europea
Duración : 1/11/2013-31/12/2014
Financiación: 94.000 €,
Estado: concedido

Referencia: Plan Nacional de I+D+I 2008-2011, TIN2010-20644-C03-01
Título: MULTIMEDICA: Extracción de Información Multilingüe en Sanidad y su aplicación a documentación divulgativa y científica
Participación: Miembro equipo investigador
Entidad Financiadora: Ministerio de Economía y competitividad
Duración: 1/01/2011-30/06/2014
Financiación: 86.152 €,
Estado: concedido

Referencia: S2009/TIC-1542
Título: MEJORANDO EL ACCESO, EL ANÁLISIS Y LA VISIBILIDAD DE LA INFORMACION Y LOS CONTENIDOS MULTILINGÜES Y MULTIMEDIA EN RED PARA LA COMUNIDAD DE MADRID (MAVIR)
Participación: Miembro equipo investigador
Entidad Financiadora: Comunidad de Madrid
Duración: 1/01/2010-31/05/2014
Financiación: 234.245,95 €,
Estado: concedido

Referencia: TIN2007-67407
Título: BRAVO: Búsqueda de respuestas avanzada multimodal y multilingüe
Participación: Miembro equipo investigador
Entidad Financiadora: Ministerio de Educación y Ciencia
Duración: 01/10/2007-31/12/2010
Financiación: 113.740 €,
Estado: concedido

C.3. Contratos, méritos tecnológicos o de transferencia

Título: Estudio tecnológico sobre extracción de información aplicada a la gestión de las ofertas de Infojobs.
Empresa: Infojobs
Investigador Principal: Isabel Segura Bedmar
Duración: 21/02/2011- 20/05/2011
Cuantía: 9.000 €



Título: BUSCAMEDIA: Hacia una adaptación semántica de medios digitales multired-multiterminal.
Empresa: DAEDALUS S.A.
Investigador Principal: Paloma Martínez Fernández
Duración: 10/09/2009- 10/12/2012
Cuantía: 77.500 €

C.4. Registros.

1. Inventores: Isabel Segura-Bedmar, Paloma Martínez, César de Pablo-Sánchez. Título: DrugDDI: corpus anotado con interacciones farmacológicas. N. de registro: 00766320105086. Entidad titular: UC3M.
2. Inventores: Isabel Segura-Bedmar, Paloma Martínez, César de Pablo-Sánchez. Título: Sistema de extracción de información en el dominio farmacológico. N. de registro: 00766620105086. Entidad titular: UC3M.
3. Inventores: Isabel Segura-Bedmar, Paloma Martínez, María Segura-Bedmar. Título: Corpus DrugNER. N. de registro: 00767020105086. Entidad titular: UC3M.
4. Inventores (p.o. de firma): Isabel Segura-Bedmar, Paloma Martínez, César de Pablo-Sánchez, Mario Crespo. Título: Corpus DrugNERAr. N. de registro: 00767220105086. Entidad titular: UC3M.