

Parte A. DATOS PERSONALES		Fecha del CVA	09/2024
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	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, el último en el periodo 2015-2020.

Nº sexenios de transferencia: 1.

Número de tesis doctorales dirigidas en los últimos 10 años: 6 con máxima calificación. De ellas, 2 han sido tesis europeas y han obtenido el Premio Extraordinario de la UNED y 2 tesis con mención internacional.

Citas totales: 1230 según Google Scholar, y 363 citas desde 2019 según la misma fuente.

Índice h: 16 e índice i10 de 28, según Google Scholar.

Participación en 20 proyectos y contratos de investigación financiados en convocatorias públicas. De ellos he sido IP en 6.

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

Fermo parte del grupo de investigación en Procesamiento del Lenguaje Natural y Recuperación de Información NLP&IR (<http://nlp.uned.es>). He realizado dos estancias de investigación, una predoctoral en el laboratorio CRL (U. de Nuevo Méjico) y otra postdoctoral en la U. James Madison.

Mi actividad investigadora se ha desarrollado fundamentalmente en dicho grupo. El grueso de los proyectos de investigación en los que he participado se ha centrado en la investigación en Procesamiento de Lenguaje Natural en diferentes dominios. Mi investigación se condensa en dos campos: la alineación de corpus paralelos bilingües y en la minería de textos multilingües. En el campo de la alineación de corpus paralelos, tema de mi tesis doctoral, realicé aportaciones relevantes y publiqué en los congresos más prestigiosos del campo. Mi investigación posterior se ha centrado en diferentes aspectos relacionados con la minería de textos: representación de documentos monolingües y multilingües, cálculo de la similitud entre documentos y entre tokens, y algoritmos de clustering, logrando el modelado de diferentes problemas del estado del arte como problemas de clustering, y aprendizaje supervisado para clasificación automática de documentos. He trabajado con documentos heterogéneos (noticias, artículos, tuits, páginas web, informes médicos) y de diferentes dominios, general, administrativo y biomédico. En el

campo de la minería de textos también he realizado aportaciones destacadas que han sido publicadas en congresos y revistas de prestigio internacional en esos campos. Soy coautora de más de 80 publicaciones científicas en el área, habiendo publicado con investigadores de diferentes universidades, centros de investigación y países.

He participado en 20 proyectos y contratos de investigación financiados en convocatorias públicas. De ellos 10 son de ámbito nacional, 4 de comunidades autónomas, 2 subvencionados con planes propios de la universidad, 2 contratos de transferencia y 3 contratos Art. 83 con empresas. En lo que respecta a la gestión de la investigación, he sido Investigadora Principal en 6 proyectos de investigación competitivos, de ellos 2 de ámbito nacional, 1 autonómico y 1 de la universidad y en 2 contratos de investigación.

En lo que respecta a transferencia de resultados de investigación, he participado en 4 proyectos y contratos en los que se ha realizado transferencia de resultados. Como resultado del proyecto-contrato de investigación con la empresa Cátenon Worldwide Executive Search se generó y se transfirió una herramienta para la búsqueda inteligente de profesionales on-line en redes sociales de reclutamiento, que ha estado en explotación.

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:

Early diagnosis of HIV cases by means of text mining and machine learning models on clinical notes. *Computers in Biology and Medicine* 179 (2024). Rodrigo Morales-Sánchez, Soto Montalvo, Adrián Riaño, Raquel Martínez, María Velasco.

Villaplana, A.; **Martínez, R.**; Montalvo, S. Improving Medical Entity Recognition in Spanish by Means of Biomedical Language Models. *Electronics* 2023, 12, 4872. Special Issue [Natural Language Processing and Information Retrieval, 2nd Edition](#) <https://doi.org/10.3390/electronics12234872>

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 Martínez**, Víctor 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.

C.2. Proyectos

Título: Detección precoz de enfermedades de alto impacto mediante el procesamiento del lenguaje natural (EDHER-MED). **Referencia:** PID2022-136522OB-C21. **Entidad** financiadora y convocatoria: Ministerio de Ciencia e Innovación. **Fechas inicio y finalización:** 01/01/2023 - 31/12/2025. **Nombre del investigador principal:** Juan Martínez Romo y Lourdes Araujo Serna.

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*. **Fechas inicio y finalización:** 01/01/2020 - 31/12/2022. **Nombre del investigador principal:** *Lourdes Araujo Serna y Juan Martínez Romo*.

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^o 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^o 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^o 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: 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^o 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: Catenon 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^o 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 **cinco** tramos.
- Directora del Dpto. de Lenguajes y Sistemas Informáticos de la UNED desde septiembre de 2021.
- 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.
- Miembro Vocal de la Junta Directiva de la Sociedad Española del Procesamiento del Lenguaje Natural desde septiembre de 2007 hasta septiembre de 2019.
- Miembro del Comité Científico de la revista Procesamiento del Lenguaje Natural desde 2008. La revista recibió en 2012 el Certificado de Revista Excelente de la Federación Española de Ciencia y Tecnología (FECYT) y lo renovó en 2015.
- Miembro del Comité Científico del congreso LREC desde 2008.
- Revisora externa de revistas internacionales.
- Coorganizadora del 2nd workshop on Human Language Technologies for Iberian languages, IberEval'2017 y del 3rd workshop on Human Language Technologies for Iberian languages, IberEval'2018.