

Parte A. DATOS PERSONALES

Fecha del CVA	24/01/2020
----------------------	------------

Nombre y apellidos	Alberto Díaz Esteban		
DNI/NIE/pasaporte	11818724J	Edad	49
Núm. identificación del investigador	Researcher ID		
	Código Orcid	0000-0003-1966-3421	

A.1. Situación profesional actual

Organismo	Universidad Complutense de Madrid		
Dpto./Centro	Departamento de Ingeniería del Software e Inteligencia Artificial Facultad de Informática		
Dirección	C/ Profesor José García Santesmases, s/n Ciudad Universitaria 28040 - MADRID		
Teléfono	913945761	correo electrónico	albertodiaz@fdi.ucm.es
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 = 11 / Revistas en Q1 = 5
 - SCOPUS (SJR): Total de revistas indexadas = 20 / Revistas en Q1 = 8
- Citas totales (según Google Scholar): 828
- Promedio citas/año durante los últimos 5 años (según Google Scholar): 371
- Índice h (Google Scholar): 17

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

Licenciado en Ciencias Físicas por la Universidad Complutense de Madrid en junio de 1993 obtuvo el Grado de Doctor en Informática en un doctorado con mención de calidad con la calificación de sobresaliente cum laude en julio de 2005. Profesor contratado doctor desde mayo de 2008 y acreditado para profesor titular desde septiembre de 2011. Profesor titular desde mayo de 2017. Tiene 2 sexenios de investigación.

Su trayectoria científica comenzó en la Universidad Europea de Madrid en 1995 donde participó como investigador en 4 proyectos de investigación relacionados con el acceso personalizado a la información en Internet. Estos proyectos eran los primeros que se conseguían asociados a la Ingeniería Informática en la Universidad Europea de Madrid. Fruto de estos proyectos se obtuvieron varias publicaciones asociadas a diversas líneas de investigación: clasificación de texto, personalización de la información, modelado de usuario, generación de resúmenes.

A partir de 2001 su investigación estuvo ligada a la Facultad de Informática de la Universidad Complutense de Madrid donde presentó su tesis doctoral en 2005: Integración de técnicas de clasificación de texto y modelado de usuario para la personalización en servicios de noticias.

Posteriormente empezó a investigar más en profundidad en la generación automática de resúmenes utilizando grafos conceptuales para representar y extraer información de los documentos analizados, especialmente en el dominio de la biomedicina. Esta línea de investigación fue seguida en la tesis “Uso de Grafos Semánticos en la Generación Automática de Resúmenes y Estudio de su Aplicación en Distintos Dominios: Biomedicina, Periodismo y Turismo” defendida en 2011. Asociada a esta tesis se obtuvieron numerosas publicaciones en congresos y revistas indexadas. También se mantuvieron dos colaboraciones internacionales: por un lado, con el Grupo de Procesamiento de Lenguaje Natural de la Universidad de Sheffield, donde se realizó una estancia y se generaron 2 publicaciones, y por otro lado con la Biblioteca de Medicina de los Estados Unidos, a partir de la cual se obtuvieron otras 2 publicaciones.

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) donde se colaboró con la empresa Indizen Technologies S.L. en la investigación de tecnologías de procesamiento de lenguaje natural para la mejora la indexación de documentos clínicos. En particular se desarrollaron 3 productos software que fueron incluidos en el producto comercial SmartCoder para análisis de informes clínicos: un sistema de corrección ortográfica, un sistema de tratamiento de acrónimos y un sistema de tratamiento de negaciones. Continuando con este tema ha profundizado en el tema de detección de la negación y extracción de interacciones entre fármacos en textos biomédicos, dirigiendo una tesis doctoral.

En los últimos años (2013-2017) ha participado en proyectos europeos relacionados con la creatividad computacional en la evaluación de un generador de poesía multilingüe. Por otro lado, participa en un proyecto nacional de la convocatoria de Desarrollo e innovación orientada a los retos de la sociedad titulado “Inclusión Digital, Lenguaje Natural y Comunicación” (2016-2019) donde participa en la investigación de técnicas de generación de resúmenes para facilitar el acceso a la información de personas con diversidad funcional. En la actualidad dirige el 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).

También cabe destacar la participación en proyectos artículo 83 con empresas, donde ha participado durante 2017 en el desarrollo de un prototipo funcional de un chatbot para la interacción entre pacientes y hospitales sobre enfermedades cardíacas con la empresa MedWhat, y durante el año 2019 en el desarrollo de un modelo experimental para la calificación de documentos normativos con la empresa ECIX GROUP.

Es miembro del grupo de investigación NIL, grupo reconocido de la UCM, donde investiga en la aplicación de la generación automática de resúmenes usando grafos de conceptos creados automáticamente, 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.

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

C.1. Publicaciones

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.
SJR 2017: 0.952, 73/251 (Q1) INFORMATION SYSTEMS.
Citas Google Scholar: 11

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
SJR 2016: 1.164, 463/2863 (Q1) MEDICINE (MISCELLANEOUS).
Citas Google Scholar: 5

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
SJR 2012: 1.524, 22/229 (Q1) COMPUTER SCIENCE APPLICATIONS.
Citas Google Scholar: 26

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
SJR 2012: 0.869, 55/229 (Q1) COMPUTER SCIENCE APPLICATIONS
Citas en Google Scholar: 19

Laura Plaza, Alberto Díaz, Pablo Gervás (2011). A semantic graph-based approach to biomedical summarization. ARTIFICIAL INTELLIGENCE IN MEDICINE 53(1): 1-14.
JCR 2012: 1.345, 48/111 (Q2) COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE
SJR 2012: 0.589, 75/118 (Q2) ARTIFICIAL INTELLIGENCE
Citas en Google Scholar: 54

Laura Plaza, Antonio Jimeno, Alberto Díaz, Alan R. Aronson (2011). Studying the correlation between different word sense disambiguation methods and summarization effectiveness in biomedical texts. BMC BIOINFORMATICS 12:355 1-13.
JCR 2011: 2.751, 7/47 (Q1) MATHEMATICAL & COMPUTATIONAL BIOLOGY
SJR 2011: 1.377, 23/229 (Q1) COMPUTER SCIENCE APPLICATIONS
Citas en Google Scholar: 21

Alberto Díaz, Antonio García, Pablo Gervás (2008). User-centred versus system-centred evaluation of a personalization system. INFORMATION PROCESSING & MANAGEMENT 44(3): 1293-1307
JCR 2008: 1.852, 15/61 (Q1) INFORMATION SCIENCE & LIBRARY SCIENCE
SJR 2008: 1.406, 15/229 (Q1) COMPUTER SCIENCE, INFORMATION SYSTEMS
Citas en Google Scholar: 38

Alberto Díaz, Pablo Gervás (2007). User-model based personalized summarization. INFORMATION PROCESSING & MANAGEMENT 43(6): 1715-1734
JCR 2007: 1.500, 11/56 (Q1) INFORMATION SCIENCE & LIBRARY SCIENCE
SJR 2007: 1.523, 22/338 (Q1) INFORMATION SYSTEMS
Citas en Google Scholar: 75

Antonio FG Sevilla, Alberto Fernández-Isabel, Alberto Díaz (2016). Grafeno: Semantic graph extraction and operation. 2016 Eleventh international conference on digital information management (ICDIM): 133-138.
Citas en Google Scholar: 1

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

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

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

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)