

UC3M R&D IN THE area of speech and language technologies

IDENTIFICATION OF THE RESEARCH ACTIVITY,
TECHNOLOGIES, PATENTS, INFRASTRUCTURES,
AND OTHER UC3M CAPABILITIES IN THE AREA
OF SPEECH AND LANGUAGE TECHNOLOGIES



uc3m

Universidad **Carlos III** de Madrid

Vicerrectorado de Política Científica

Servicio de Apoyo al Emprendimiento y la Innovación



The Entrepreneurship and Innovation Support Service of Universidad Carlos III of Madrid (UC3M) presents UC3M potential in this "technology map" through the Lines of research conducted within the framework of the National and international R&D projects, patents, and other results attained by UC3M researchers in the area of speech and language technologies.

The overall knowledge achieved, experience in collaborating with the industry, the existence of internally owned infrastructures and laboratories, and above all else, the multidisciplinary nature of UC3M are unique characteristics which provide an added value for the provision of a comprehensive support to innovation in institutions, large companies, and small- and medium-sized enterprises.

We would like you to learn more about the know-how at UC3M and to collaborate in new R&D&I projects.

**Entrepreneurship and Innovation Support Service,
Universidad Carlos III de Madrid**

Contact:
comercializacion@uc3m.es

Date updated
November 2022

Table of Contents

ENGINEERING

COMPUTER ENGINEERING	5
GigaBD	5
<i>PI: Jorge Luis Morato</i>	
Human Language and Accesibility Technologies (HULAT)	6
<i>PI: Paloma Martínez, Belén Ruiz</i>	
Knowledge Reusing	7
<i>PI: Juan B. Llorens</i>	
Softlab	8
<i>PI: Ángel García Crespo</i>	
SYSTEMS ENGINEERING AND AUTOMATION	9
Robotics Lab	9
<i>PI: Carlos Balaguer, Miguel A. Salichs, Luis Moreno</i>	

SIGNAL AND COMMUNICATIONS THEORY 10

Machine Learning for Data Science (ML4DS) 10
PI: Jerónimo Arenas

Multimedia Processing 11
PI: Fernando Díaz de María

HUMANITIES, DOCUMENTATION, AND COMMUNICATION

HUMANITIES: PHILOSOPHY, LANGUAGE, AND LITERATURE 12

Grammar, Lexis, Discourse, and History 12
PI: M.ª Pilar Garcés, José Antonio Pascual Rodríguez

SPANISH CENTRE FOR SUBTITLING AND AUDIO DESCRIPTION

SPANISH CENTRE FOR SUBTITLING AND AUDIO DESCRIPTION 13

Spanish Centre for Subtitling and Audio Description (CESyA) 13
PI: José Manuel Sánchez Pena

R&D GROUP

LINES OF RESEARCH

RESEARCH PROJECTS

TECHNOLOGICAL OFFER / OTHERS

ENGINEERING

COMPUTER ENGINEERING

GigaBD

PI: Jorge Luis Morato

- Deep learning applications and data enrichment
- Accessibility and comprehensibility of data
- Veracity and reliability of data
- Infrastructures to facilitate BigData

National R&D Plan Projects

- Development of a system for conceptual retrieval by means of semantic levels in metadata diagram representation

Regional R&D Plan Projects

- Development of a semantic knowledge management-based strategic smart system

Experience and capabilities

- Information retrieval and access
- Natural language processing
- Information visualisation and retrieval in semantics web

R&D GROUP	LINES OF RESEARCH	RESEARCH PROJECTS	TECHNOLOGICAL OFFER / OTHERS
COMPUTER ENGINEERING			
<p><u>Human Language and accesibility technologies (HULAT)</u></p> <hr/> <p>PI: Paloma Martínez, Belén Ruiz</p>	<ul style="list-style-type: none"> Natural language processing (NLP) Deep learning applied to natural language processing Speech processing and speaker recognition 	<p>National R&D Plan Projects</p> <ul style="list-style-type: none"> Cognitive and sensory accessibility to video conferencing systems (Access2Meet) <p>Regional R&D Plan Projects</p> <ul style="list-style-type: none"> Natural Language Processing for Rare Diseases (NLP4Rare) MAVIR consortium (Improving the access and visibility of the multilingual information in web for the Community of Madrid) <p>In-house Projects (Funded by UC3M)</p> <ul style="list-style-type: none"> Strategic action in natural language processing and information accessibility <p>Private Funding</p> <ul style="list-style-type: none"> ACCESSTRANS EASIER 	<p>Experience and capabilities</p> <ul style="list-style-type: none"> Research projects related to NLP and application thereof in the biomedical domain Incorporation of the latest information storage and retrieval technologies, linguistic technology for the extraction of information from texts in various domains, in information processing for business management. Development of applications which incorporate language technology, such as lexical resources, for information extraction systems, text classification, entity and relationship recognition, mainly in the biomedical domain <p>Technological Offer</p> <ul style="list-style-type: none"> Indexation and multilingual document search tool Response search system Advanced interaction solutions <p>Doctoral Theses</p> <ul style="list-style-type: none"> Clinical word sense disambiguation (2022) Lexical simplification for the systematic support of cognitive accessibility guidelines (2022)

R&D GROUP	LINES OF RESEARCH	RESEARCH PROJECTS	TECHNOLOGICAL OFFER / OTHERS
COMPUTER ENGINEERING			
<p>Knowledge Reusing</p> <p>PI: Juan B. Llorens</p>	<ul style="list-style-type: none"> • Knowledge management <ul style="list-style-type: none"> · Universal representation of information (digitalisation and integration) · Retrieval and reuse · Ontology generation • Natural language processing <ul style="list-style-type: none"> · Computational linguistics and neural networks · Entity recognition · Pattern extraction · Automatic text generation · Language representation systems (foundation models) · Opinion and of feeling analysis · Translation • Engineering methods: <ul style="list-style-type: none"> · Engineering of requirements · Model-based systems engineering · Product lifecycle management · Variability management • Interoperable platforms for data and knowledge management <ul style="list-style-type: none"> · Continuous quality · Continuous and collaborative engineering · Business intelligence 	<p>European Projects</p> <ul style="list-style-type: none"> • H2020-ECSEL-Intelligent Reliability 4.0 • H2020-ECSEL-Arrowhead Tools • H2020-ECSEL-NewControl • H2020-ECSEL-AMASS • ARTEMISA -CRYSTAL • JUST-AG-2019/JUST-JACC-EJU-AG-2019 -IMOLA III • JUST AG-2016-05/ Grant Agreement no. 764350 -IMOLA II <p>National and Regional Projects</p> <ul style="list-style-type: none"> • RESTART • Madrid Flight On Chip • MoMEBIA <p>Private Funding</p> <ul style="list-style-type: none"> • Semantic engine design and development. • Service for analysing the dissemination of sustainable development goals. • Social network monitoring and real-time news notification service for professionals in the field of information. • Radio Television Española (RTVE) and UC3M Chair. • TRC-Research Chair. Knowledge Centric Solutions and UC3M KR. • Alma Tecnológica - Aurelio Delgado Chair. Alma Tecnológica Foundation and UC3M KR. 	<p>Experience and capabilities</p> <p>This group is involved in the creation and development of processes, technologies, methods, and tools for improving system engineering processes through the massive and intelligent use of semantic technologies and knowledge management.</p> <p>Human language technologies</p> <p>Linking language to the world through modelling, language understanding, and dialog systems.</p> <ul style="list-style-type: none"> • Automatic knowledge discovery and extraction techniques • Traceability identification systems • Knowledge quality measurement algorithms • Knowledge visualisation • Standard-based interoperable platforms • Technology consulting and standardisation • Clear, inclusive, and non-violent language detection systems • Members on a personal level of the BoD (Board of Directors) of the Spanish Association of Systems Engineering (official chapter of INCOSE), SEPLN, CSTA, INCOSE.org • Members such as UC3M of ProSTEP, INCOSE ESPAÑA, OMG, and OSLC <p>Doctoral Theses</p> <ul style="list-style-type: none"> • Use of genetic algorithms for the automatic creation of requirements patterns • Unified representation language-based interoperability methodology

R&D GROUP	LINES OF RESEARCH	RESEARCH PROJECTS	TECHNOLOGICAL OFFER / OTHERS
COMPUTER ENGINEERING			
<p>Softlab</p> <hr/> <p>PI: Ángel García Crespo</p>	<ul style="list-style-type: none"> Natural language processing: dialogue system, mono- and multilingual information recovery (IR), question answering, adaptation in IR systems using ontologies Natural interaction systems Automatic multilingual subtitling, both live and multilingual Automatic text generation through large-scale language models Automatic code generation 	<p>Private Funding</p> <ul style="list-style-type: none"> App for better hearing (AUDISMART) Automatic subtitling of recorded programs Subtitling of information in Spanish for territorial centres Automatic subtitling of programs with a non-homogeneous structure Subtitling quality monitoring and supervision Intelligent system for automatic subtitling management 	<p>Experience and capabilities</p> <p>This group performs the planning and development of solutions for natural interaction applications.</p> <p>Doctoral Theses</p> <ul style="list-style-type: none"> Design of a multi-domain conceptual model for recommendations by filtering semantic information in social media Influence of discourse segments on speaker discrimination <p>Equipment</p> <ul style="list-style-type: none"> <i>IRIS Laboratory</i> <p>Various research projects have been performed in the laboratory facilities utilizing the connectivity capabilities and equipment available therein.</p>

R&D GROUP	LINES OF RESEARCH	RESEARCH PROJECTS	TECHNOLOGICAL OFFER / OTHERS
SYSTEMS ENGINEERING AND AUTOMATION			
<p><u>Robotics Lab</u></p> <hr/> <p>PI: Carlos Balaguer, Miguel A. Salichs, Luis Moreno</p>	<ul style="list-style-type: none"> • Humanoid robots • Social robots • Healthcare robots 	<p>European Projects</p> <ul style="list-style-type: none"> • Rethinking Robotics for the Robot Companion of the future - RoboCom++ • ALMA: Human Centric Algebraic Machine Learning • Multi-Robot Cognitive Systems Operating in Hospitals (MONARCH) <p>National R&D Plan Projects</p> <ul style="list-style-type: none"> • IROPER - Intelligent Robotics for Personal Needs • ROSES - Social robots for physical, cognitive, and emotional stimulation in elderly <p>Regional R&D Plan Projects</p> <ul style="list-style-type: none"> • COMPANION-CM_1 - Artificial intelligence and cognitive models for symmetric human-robot interaction in the field of healthcare robotics. <p>UC3M projects</p> <ul style="list-style-type: none"> • SMM4HRI-CM-UC3M - Social Media Mining for Human-Robot Interaction 	<p>Experience and capabilities</p> <ul style="list-style-type: none"> • Dialog-based human-robot interaction <ul style="list-style-type: none"> · A grammar-based speech recognizer (Loquendo) and an open-language speech recognizer (Google ASR). Different "skills" capable of working in coordination with the robot architecture have been developed based on same. · Speech management capability-based speech synthesis system which is able to keep a queue in memory, hijack the voice, release it, silence it, etc. It is based on three systems (Loquendo, Festival, and Google TTS), through which the robot is able to speak with emotions, with speed control, with sound gestures (such as laughing, whistling, crying, etc.) and in more than 35 different languages. · Grammar- or template-based natural language generation. · Use of advanced techniques, such as Deep Learning, for natural language processing and generation. · Generation of multimodal dialogues which can be adapted to the context of the interaction. <p>Group's video</p>

R&D GROUP	LINES OF RESEARCH	RESEARCH PROJECTS	TECHNOLOGICAL OFFER / OTHERS
SIGNAL AND COMMUNICATIONS THEORY			
<p><u>Machine Learning for Data Science (ML4DS)</u></p> <p>PI: Jerónimo Arenas</p>	<ul style="list-style-type: none"> • Machine learning for Big Data • Machine learning based on neural networks and other classic techniques (nucleus methods, random forest, e.g.) • Algorithmic variants for topic modelling • User-in-the-loop machine learning 	<p>European Projects</p> <ul style="list-style-type: none"> • IntelComp: A Competitive Intelligence Cloud/High Performance Computing Platform for Artificial Intelligence-based Science, Technology and Innovation Policy Making. • nextProcurement: Open Harmonized and Enriched Procurement Data Platform <p>National R&D Plan Projects</p> <ul style="list-style-type: none"> • Machine learning for Big Data Analysis of Heterogeneous Document Sources <p>Private Funding</p> <ul style="list-style-type: none"> • Advisory and technical assistance activities in the cases of using language technologies in FECYT • Automated classification of R&D&I projects by themes from their abstracts, applying natural language methodology 	<p>Experience and capabilities</p> <p>This group has over twenty years of experience in R&D&I in the field of machine learning and its applications in data analysis and information processing.</p> <p>The group is currently working on important challenges to make machine learning more scalable and accessible to end users.</p> <ul style="list-style-type: none"> • Natural language processing • Topic modelling • Machine learning on science and technology texts. • Machine learning with large databases • Decision-making support by providing evidence based on data. • Strategies for incorporating expert knowledge in model training (user-in-the-loop) • Speech technologies for health

R&D GROUP	LINES OF RESEARCH	RESEARCH PROJECTS	TECHNOLOGICAL OFFER / OTHERS
SIGNAL AND COMMUNICATIONS THEORY			
<p><u>Multimedia Processing</u></p> <hr/> <p>PI: Fernando Díaz de María</p>	<ul style="list-style-type: none"> • Speech technologies • Speech recognition and processing 		<p>Experience and capabilities</p> <p>This group specialises in speech, audio, image, and video processing techniques, especially in automatic contents extraction, automatic speech recognition, and last generation video coding.</p> <ul style="list-style-type: none"> • Voice signal processing in the context of neurological diseases: <ul style="list-style-type: none"> · Alzheimer: systems for aiding diagnosis through the analysis of spontaneous speech. • Speech and natural language processing • Robust speech recognition • Voice interfaces for mobile devices • Speech transcription to support the subtitling process for the hard of hearing <p>Technological Offer</p> <p>Design of algorithmic solutions in the field of speech, audio, image, and video processing, with special emphasis on artificial vision, video coding, and speech technologies.</p> <p>Doctoral Theses</p> <ul style="list-style-type: none"> • Contributions to robust speech recognition in communication networks using transparametrisation • Support vector machines for robust speech recognition

R&D GROUP

LINES OF RESEARCH

RESEARCH PROJECTS

TECHNOLOGICAL OFFER / OTHERS

HUMANITIES: PHILOSOPHY, LANGUAGE, AND LITERATURE

HUMANITIES: PHILOSOPHY, LANGUAGE, AND LITERATURE

Grammar, Lexis, Discourse, and History

PI: M.^a Pilar Garcés, José Antonio Pascual Rodríguez

- History of Spanish, lexicography, historical lexicography, linguistic historiography
- Lexis, semantics, pragmatics, discourse analysis
- Syntactic theory, Spanish syntax, relationship between lexis and grammar, teaching Spanish as a foreign language, lexicography applied to the teaching of Spanish
- History of cities in the Middle Ages, history of women in the Middle Ages, historical semantics
- Lexical semantics, the relationship between grammar and lexis
- Historical lexicography, relationships between grammar and lexis, history of language
- The Spanish language, lexicalisation, collocations, verb-noun combinations

National R&D Plan Projects

- Diachronic study of adverbs and adverbial speech for their description in the new historical dictionary of the Spanish language (*nuevo diccionario histórico de la lengua española* - NDHE)
- Diachronic study of discourse markers in Spanish for their description in a historical dictionary
- Lexical-syntactic theories: predicate-argument relationships and their lexicographic and didactic applications
- From sentence to discourse: contrastive study

Regional R&D Plan Projects

- Methodological contributions to the preparation of a historical dictionary
- New linguistic perspectives for the preparation of a historical dictionary
- Linguistic contribution to the preparation of a historical dictionary

Experience and capabilities

- Semantics
- Pragmatics
- Discourse analysis
- Lexicography
- Spanish syntax

Doctoral Theses

- The lexis of football language. A contrastive analysis in Spanish and Chinese
- Gender in words and gender of words: an approach to the ideological gender variable from the semantic phenomenon of heteronymy
- Diachronic analysis of adverbs with discursive function and their representation in a historical dictionary
- A theoretical and applied proposal for teaching Spanish copulative and semicopulative constructions to Chinese language learners
- Theoretical and applied perspectives in teaching Spanish to sub-Saharan adults in a linguistic immersion context
- Pragmalinguistics of dysphemism and discourtesy. Hostile speech acts in virtual communication media
- Relationships between lexical morphology and lexicography: denominational adjectives with possession value and their treatment in dictionaries
- Semantic inclusion relationships

GRUPO I+D

DESCRIPTION

RESEARCH

SERVICES

SPANISH CENTRE FOR SUBTITLING AND AUDIO DESCRIPTION

SPANISH CENTRE FOR SUBTITLING AND AUDIO DESCRIPTION

Spanish Centre for Subtitling and Audio Description (CESyA)

Director: José Manuel Sánchez Pena

CESyA is a centre under the Royal Board on Disability of the Ministry of Social Rights and Agenda 2030 which is managed by Universidad Carlos III of Madrid (UC3M) and collaborates with the Spanish Committee of Representatives of Persons with Disabilities (CERMI) in the multidisciplinary project to promote accessibility in the audiovisual media environment through subtitling and audio description services.

Since its creation, CESyA has tirelessly performed very intense work in the field of research, technological development, and innovation (R&D&I).

- Awards won both at international and national level associated with prototypes and tools that facilitate the audiovisual integration of people with sensory disabilities.
- Leading and/or participating in competitive research projects funded by public organisations, as well as contracts with national and international companies and institutions.
- In addition to the transfer and innovation of research results derived from these projects, patents and software have been registered, some of them being commercially exploited under license.
- Elaboration of doctoral theses on topics totally or partially related to the field of audiovisual accessibility.

- Automatic monitoring of the broadcasting of subtitles, audio description, and sign language in DTT
- Information on accessible cultural offer through the Accessible Cultural Agenda.
- Advising companies, users, and administrations in all aspects related to audiovisual accessibility: subtitling and audio description in cinemas, theatres, museums, television, internet, and other platforms
- Citizen attention services
- Training. Offering courses and workshops on general accessibility, subtitling, audio description, culture accessibility, web accessibility, and accessible digital documentation
- Quality certification through the CESyA stamp

Co-funding:

Activity of the Project "UC3M Plan for Promoting Innovation and R&D Result Transfer in the Production Sector of the Community of Madrid with Priority in the Southern Metropolitan Area" with Ref.: OI2018/PC-UC3M-5152 and the acronym PC-UC3M. This project was awarded in the 2018 Call for Grants for fostering technological innovation and promoting technology transfer to the production sector comprised within the priorities of the Regional Research and Innovation Strategy for Smart Specialization (RIS3) of the Community of Madrid through technological innovation coordinating entities. It is co-funded by the European Regional Development Fund which provides 25% of the funding and by the Community of Madrid which provides another 25% within the framework of the FEDER 2014-2020 operational program.



uc3m

Universidad **Carlos III** de Madrid

Vicerrectorado de Política Científica

Servicio de Apoyo al Emprendimiento y la Innovación