

Descripción de la Entidad

La Universidad Rey Juan Carlos cuenta con más de diez grupos de investigación en Humanidades digitales y tecnologías del español; concretamente, Humanidades digitales: LIYNMEDIA, HASTHGAR, CINTER, GID-LyHD e ITEM y Tecnologías del español: BigMed+, DSLAB, LeaDS, Grupo de Inteligencia Artificial, SoftDev, LITE, EfiTimeTransSS y GAVAB. En los últimos 20 años estos grupos han dirigido más de 150 proyectos de investigación y de transferencia tecnológica al sector productivo, con financiación pública y privada, nacionales y europeos.

Áreas de Actividad

Humanidades digitales, Lingüística, Aprendizaje automático, Inteligencia Artificial, Ciencia e Ingeniería de Datos e Ingeniería Biomédica

Mapa de Oferta

- Creación de bases de datos, corpus y repositorios especializados. Diseño, metodología, aplicación, preservación, reutilización y difusión. Gestión de datos de investigación y ciclo de los datos.
- Análisis de datos, visualización y creación de nuevos objetos de conocimiento. Aplicación de nuevas metodologías de investigación e innovación docente.
- Aplicación de nuevas metodologías de estudio, análisis y difusión del patrimonio histórico/artístico.
- Formación de profesionales en áreas relacionadas con la ciencia e ingeniería de datos, el aprendizaje automático y la inteligencia artificial en castellano.
- Integración en EOSC Marketplace, DARIAH/CLARIN y SSHOC Marketplace.
- Diseño y aplicación de técnicas de ciencia de datos, aprendizaje máquina e inteligencia artificial para el procesamiento y análisis automático de voz y texto en español.
- Adaptación y mejora de rendimiento de algoritmos de aprendizaje máquina aplicados al procesamiento y análisis de datos textuales en español.
- Diseño e implementación de algoritmos de búsqueda por similitud.
- Diseño e implementación de algoritmos de detección de género.
- Uso de tecnologías semánticas y las implicaciones éticas y legales de la IA en español.
- Plataformas de anotación semántica para recursos digitales en español para mejorar la accesibilidad y la reutilización de estos recursos.
- Generación de texto en español y traducción automática mediante aprendizaje automático.
- Explicabilidad de modelos de aprendizaje automático en el procesamiento del lenguaje español.
- Recuperación automática de información de redes sociales en castellano.
- Digitalización y transcripción de documentos manuscritos.
- Procesamiento de lenguaje natural aplicado al dominio de la salud para ayuda a diagnóstico, predicción de eventos adversos y ayuda a la asistencia y medicina preventiva.

Mapa de Demanda

Desarrollo y mejora de herramientas, procedimientos y buenas prácticas para el procesamiento de datos textuales en español. A pesar de que muchas herramientas y métodos propuestos son aplicables a datos textuales en múltiples lenguajes, lo cierto es que una gran mayoría de recursos están todavía muy enfocados al procesamiento de texto en inglés.

Establecer vías estables de colaboración con entidades de renombre y autoridad en el idioma español presentes en la CAM, tales como el Instituto Cervantes o la Real Academia Española, que permitan enriquecer las herramientas y métodos para procesamiento y análisis de texto en castellano. En particular, colaboración para la creación y enriquecimiento de corpus lingüísticos para entrenamiento de algoritmos y análisis de sentimiento de información textual en español.

Creación colaborativa de conjuntos de datos de referencia de alta calidad, orientados al entrenamiento de algoritmos y modelos de procesamiento de lenguaje natural y minería de texto.

C/Tulipan S/N,
Móstoles, Madrid 28933
España

Fernando Enrique García Muiña
+34914884915
vice.investigacioneinnovacion@urjc.es