



## Becas colaboración curso 2023/2024

Fecha: 29 Mayo 2023

### Vicerrectorado de Investigación

Subcomisión de I+D+i

Propuesta del departamento *SISTEMAS INFORMÁTICOS Y COMPUTACIÓN*

**Núm Proyecto: 2023/32/00001**

#### Responsable

Hurtado Oliver, Lluís Felip

#### E-mail

lhurtado@dsic.upv.es

#### Ext.

83500

#### Responsable

Segarra Soriano, Encarnación

#### E-mail

esegarra@dsic.upv.es

#### Ext

77738

#### Título proyecto

Aplicación de Técnicas de Procesamiento del Lenguaje Natural para la Recuperación de Información basada en representaciones vectoriales densas

#### Valoración proyecto

4

#### Descripción proyecto

El grupo de investigación ELiRF tiene una dilatada experiencia en distintas aplicaciones del Procesamiento de Lenguaje Natural. Algunas de estas aplicaciones son la monitorización de redes sociales, el análisis de artículos de prensa y programas de debate televisivos, y el resumen automático. Actualmente está trabajando dentro de un proyecto del Ministerio de Ciencia e Innovación (BEWORD-UPV: DESCUBRIENDO EL SIGNIFICADO Y LA INTENCIÓN MÁS ALLÁ DE LA PALABRA HABLADA: HACIA UN ENTORNO INTELIGENTE, PID2021-126061OB-C41), consistente en mejorar la capacidad de los sistemas autónomos para procesar información recopilada de fuentes de naturaleza muy diversa, como documentos hablados o escritos, videos o incluso sensores. Nuestro objetivo es el desarrollo de sistemas que incorporen otra información, diferente a la puramente léxica, en el proceso de comunicación.

El grupo de investigación dispone de gran cantidad de recursos hardware, software y corpus que darán soporte durante la realización del trabajo asociado a la beca.

#### Actividades a realizar por el alumno

El candidato o candidata se integrará en el grupo de investigación, que forma parte del instituto de investigación VRAIN, asistiendo a reuniones y colaborando con sus miembros en los distintos proyectos activos.

En particular trabajará en la construcción de un sistema de recuperación de información sobre una colección de documentos audiovisuales para los que se dispone de la transcripción manual y automática. Para la construcción del índice de la colección tanto los documentos como las consultas se representarán mediante vectores densos. Se abordarán diferentes aproximaciones para la obtención de la representación de las consultas y documentos en base a los embeddings de sus términos y para el cálculo de su similitud.



## Becas colaboración curso 2023/2024

*Fecha: 29 Mayo 2023*

Con todo ello, el becario o becaria aprenderá los fundamentos teóricos (modelos de lenguaje, métricas de evaluación, etc.) y tecnológicos (word embeddings, deep learning, etc.) que serán necesarios para abordar las tareas encomendadas.

### **Localización de la actividad (Campus)**

Campus de Vera

### **Horario**

3 horas diarias, a acordar con el o la estudiante