



Becas colaboración curso 2022/2023

Fecha: 01 Junio 2022

Vicerrectorado de Investigación

Subcomisión de I+D+i

Propuesta del departamento *SISTEMAS INFORMÁTICOS Y COMPUTACIÓN*

Núm Proyecto: 2022/32/00015

Responsable

Segarra Soriano, Encarnación

E-mail

esegarra@dsic.upv.es

Ext.

77738

Responsable

Hurtado Oliver, Lluís Felip

E-mail

lhurtado@dsic.upv.es

Ext

83500

Título proyecto

Aplicación de Técnicas de Procesamiento del Lenguaje Natural para la extracción de información en noticias periodísticas en valenciano

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 en un campo de gran interés para la aplicación de técnicas de Procesamiento de Lenguaje Natural: la extracción de conocimiento a partir de noticias periodísticas en valenciano. El proyecto está enmarcado en la colaboración con la Corporació Valenciana de Mitjants de Comunicació.

La automatización de la tarea de clasificación de textos en un conjunto de categorías predeterminado, incluyendo la detección de textos que quedan fuera de las categorías definidas a priori, así como la detección de entidades nombradas son temas de gran interés en diversos campos de aplicación del Procesamiento del Lenguaje Natural. Uno de esos campos de aplicación es el trabajo de los documentalistas sobre textos periodísticos en un medio de comunicación audiovisual.

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 dos líneas. En primer lugar, se llevará a cabo un estudio del comportamiento de un sistema de reconocimiento de entidades nombradas sobre textos periodísticos de un medio de comunicación en valenciano.



Becas colaboración curso 2022/2023

Fecha: 01 Junio 2022

En segundo lugar, el trabajo consistirá en la construcción y comparación de diversos sistemas de clasificación, incluyendo también la detección de textos fuera del conjunto predeterminado de categorías. Se probarán técnicas basadas en redes neuronales.

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