

## XVI Seminario de edición y crítica de textos clásicos

### Datos de la actividad

#### TÍTULO:

XVI Seminario de edición y crítica de textos clásicos

#### RESUMEN:

El objetivo propuesto es conseguir un texto transcrito de la forma más correcta posible (siempre habrá algunos errores) a partir de las imágenes de ediciones antiguas con la asistencia de sistemas informáticos, de manera que se obtenga un archivo con el texto legible para un ordenador.

Este proceso es conocido habitualmente como OCR (Optical Character Recognition). En los impresos modernos desde hace tiempo es una cuestión resuelta, para los impresos antiguos y texto manuscrito aún no se ha conseguido.

Los distintos sistemas que logran esto usan diversas técnicas de inteligencia artificial entre las cuales la parte esencial es una Red Neuronal Recurrente, que tiene la capacidad de ser entrenada y aprender.

En los impresos modernos los sistemas han sido entrenados con una cantidad enorme de datos (Big Data), que no están aún disponibles para los impresos antiguos. Por esta razón una parte del trabajo consiste en transcribir manualmente una parte del texto para entrenar el sistema y hacerlo capaz de reconocer automáticamente textos semejantes a los datos del entrenamiento.

Esta tarea requiere una dedicación más amplia de la que se puede llevar a cabo en un seminario y además es muy recomendable que una parte del trabajo sea colaborativo.

Las herramientas de software que se utilizarán durante la actividad son *Transkribus* y *ocr4all*. Con el fin de poder usar los datos y modelos en ambas herramientas, se facilitarán las ediciones, algunas transcripciones y los modelos para su aplicación. En principio veremos en particular las ediciones de las *Filípicas* Valencia, Mey 1559 y el libro VII de la *Naturalis Historia* de Plinio, Salamanca, J. M. de Terranova 1560.