

Minería de datos e IA al servicio de los grandes misterios de la literatura

Doce instituciones europeas colaboran en la creación de herramientas digitales que leen en el XXI las obras de siglos anteriores

La UNED, junto a otras doce instituciones europeas, participa en el proyecto *Computational Literary Studies Infraestructure*, que aplica herramientas digitales y algoritmos a la creación literaria. Desvelar quienes se esconden tras los misteriosos pseudónimos de algunas obras de la literatura ya está en manos de la Inteligencia Artificial.



Para entrar en los misterios de la literatura universal, nada parece más natural en el siglo XXI que aplicar el Big Data, la Inteligencia Artificial y otros métodos computacionales a la creación literaria. Descubrir patrones de creación, arquitectura de los versos, e incluso grandes autores ocultos tras obras mediocres es posible gracias a algoritmos que leen lo que se esconde a nuestros ojos.

La UNED se ha unido con otras doce instituciones europeas en el proyecto *Computational Literary Studies Infraestructure* (CLS Infra) para experimentar nuevas maneras de investigar el ámbito literario con herramientas de la era digital. El proyecto cuenta con unos fondos de 5 millones de euros y construirá durante cuatro años un puente para compartir recursos y datos de alta calidad que pongan algo de luz a siglos de creación literaria.

Salvador Ros es uno de los investigadores de la UNED implicados en el CLS Infra. Su misión es desarrollar, junto a los investigadores de la UNED como Víctor Fresno, José Luis Fernandez Vindel o Elena Gonzalez Blanco, una infraestructura computacional para investigar la información oculta tras la poesía. “En los poemas, sirve para identificar el tipo de rima, la cantidad de versos, identificar personajes, patrones de estilo, percepciones o sentimientos. Se trata de una infraestructura que construye piezas de *software* y aplica las herramientas, por ejemplo de Inteligencia Artificial, para automatizar los estudios literarios”, explica Ros.

El experto investigaba 40 libros en una vida, el algoritmo, toda la biblioteca

El investigador de la UNED destaca que los estudios computacionales literarios están cambiando el modo en que durante siglos se ha investigado sobre prosa, poesía o teatro. “Antes el investigador iba a la biblioteca, cogía un libro y se ponía a investigar por su cuenta. La vida de una persona alcanza para leer en profundidad entre 40 y 50 libros, en cambio un algoritmo puede leerse todos los libros de la biblioteca y sacar datos relevantes, entre ellos, buscar los patrones que se repiten para encontrar relaciones que no existían”, precisa.

Aplicando estas herramientas se descubrió que la británica J.K. Rowling, la autora de la saga de Harry Potter, había escrito dos libros anteriores que había firmado con un pseudónimo. “La descubrieron gracias a estos *softwares*, identificaron los patrones estilométricos y vieron que el estilo coincidían en un 90% con el de los libros de Harry Potter. Estas investigaciones se llaman *análisis de autoría* y podrían haber contribuido a aclarar otro de los misterios de la novela reciente: quién es realmente la escritora italiana Elena Ferrante, que se desveló en 2016.

Además, **Ros** destaca que las nuevas herramientas obligan a investigar de un modo compartido en un campo donde los eruditos eran grandes solitarios que no compartían sus hallazgos. “Solían publicar un artículo con un resumen de la investigación, y el resto de la información se perdía. Con

este sistema, cambia la mentalidad. Si el investigador no está abierto a compartir, no encuentra colaboración y no puede avanzar”, asegura. “El proyecto es una gran oportunidad para unir esfuerzos con otras instituciones para construir el futuro de las infraestructuras CLS en la investigación literaria”.

El profesor asegura que la contribución de la UNED en la investigación poética es una continuación del proyecto ERC Posdata, según sus palabras, “casi un Premio Nobel que le otorgaron a la UNED en 2017”. “Lo hemos encadenado con esta investigación, porque un buen proyecto llama a otro”, indica.

Estas herramientas son “una mina” para la enseñanza de la poesía, asegura el investigador de la UNED. “Un estudiante de Primero de ESO puede meter un poema en el sistema y saber inmediatamente las estrofas que tiene, las figuras retóricas, el tipo de rima, los versos”, explica.

Uno de los grandes desafíos para los estudios literarios computacionales es que las fuentes literarias digitales están muy fragmentadas, ya que los académicos y lectores luchan por encontrar textos que sean accesibles y reutilizables de manera estandarizada. CLS INFRA abordará este déficit para permitir que este campo prospere. El objetivo general del proyecto es abrir los mejores recursos de minería de datos de Europa y abrir un gran campo de investigación en los estudios literarios computacionales, lo que permite aplicar el Big Data en el estudio de la Cultura.

Los socios europeos

- . UNED, Spain;
- . Institute of Polish Language at the Polish Academy of Sciences, Poland;
- . University of Potsdam, Germany;
- . Austrian Academy of Sciences, Austria;
- . École Normale Supérieure de Lyon, France;
- . Humboldt University of Berlin, Germany;
- . Charles University, Czech Republic;
- . Digital Research Infrastructure for the Arts and Humanities, France;
- . Ghent Centre for Digital Humanities, Ghent University, Belgium;
- . Belgrade Centre for Digital Humanities, Serbia;
- . Royal Netherlands Academy of Arts and Sciences, Netherlands;
- . Trier Center for Digital Humanities, Trier University, Germany;
- . Moore Institute, National University of Ireland Galway, Ireland.

[Más información](#)