

Departamento de Lenguajes y Sistemas Informáticos.
Universidad de Educación a Distancia, UNED.

“España tiene que incorporar más masa crítica a la investigación en tecnologías de la lengua”

Felisa Verdejo Maillo es directora del Departamento de Lenguajes y Sistemas Informáticos de la UNED y del grupo de investigación Procesamiento del Lenguaje Natural.

Investigadora y profesora de la UNED, Felisa Verdejo Maillo dirige además un equipo de trabajo especializado en tecnologías de la lengua, aquéllas que mediante la voz o el texto permiten programar ordenadores para que se comporten como si entendiesen la lengua humana.



Felisa Verdejo Maillo

Elena Higuera Rabadán

Mientras gran parte de la sociedad moderna se afana por aprender idiomas, asumiendo la inevitable existencia de millones de lenguas y dialectos sobre la faz de la tierra, unos pocos se empeñan en buscar soluciones que faciliten la comunicación entre sus pobladores y pongan fin a las interminables murallas que separan a la humanidad y segmentan el más universal de sus bienes comunes: el conocimiento.

Felisa Verdejo Maillo representa el perfil de una auténtica ‘heroína’ de la comunicación, una investigadora que trabaja por derribar las fronteras lingüísticas del saber. Una de estas fronteras, quizá la más lejana y extendida en el tiempo, tiene que ver con el nacimiento y evolución de las diversas lenguas. Y es que el grupo de investigación en Procesamiento del Lenguaje Natural de la UNED, que dirige Felisa Verdejo Maillo, fue uno de los artífices de la creación de una herramienta básica para la eliminación o minimización de las citadas murallas del lenguaje. Se llama Eurowordnet y es un recurso léxico, una especie de diccionario enriquecido que puede procesar la información automáticamente. “De lo que se trata es de determinar los diferentes sentidos que puede tener una palabra en un conjunto de información contextual”, explica la propia Felisa Verdejo. Eurowordnet es fruto de una investigación a nivel europeo, que tiene su precedente en el estadounidense wordnet, una aplicación que únicamente funcionaba con la lengua inglesa. “La idea era plantear este recurso pero a nivel de la Unión Europea, de manera que fuera útil para todas las lenguas que entonces pertenecían a ella”, recuerda Verdejo y enumera los primeros idiomas con los que empezaron a trabajar: inglés, francés, alemán, holandés, italiano y español.

La extensión lógica de esta herramienta superó las posteriores ampliaciones de la UE, llegando a derivar en una especie de consorcio ‘globalnet’ que abarca, por ejemplo, la mayoría de lenguas de la India. Pero no hace falta irse tan lejos para hacerse una idea del alcance de este recurso léxico. En España, la UNED en colaboración con otros grupos de investigación de distintas universidades nacionales, como la Universidad Politécnica de Cataluña, extendieron esta aplicación a las diversas lenguas que configuran nuestra geografía. La verdadera importancia de Eurowordnet reside, en opinión de una de sus artífices, en “la ventaja de no tener que empezar de cero. Esto es, si alguien quiere construir, por ejemplo, un diccionario, puede extraer información de aquí, porque ésta ya está muy sistematizada y con formatos aceptables para su procesamiento informático. Hay que poner énfasis en los proyectos colaborativos que recojan las necesidades y especificaciones de cada lengua

y de cada comunidad investigadora, para que así estas aplicaciones se puedan reutilizar de la mejor manera posible”.

Con esta herramienta en las manos, el equipo de la profesora Verdejo se dispuso a dar su segundo paso en pro de la comunicación: el ITEM o Recuperación de Información Textual en un Entorno Multilingüe. “Fue la primera vez que intentamos aplicar Eurowordnet como recurso para la recuperación de información multilingüe, algo que además dio lugar a la primera colaboración en materia de transferencia de tecnología” recuerda Felisa Verdejo, “podimos ver cómo este motor de búsqueda se podía aplicar a una serie de servicios”.

Dejando a un lado los grandes hitos en la historia del procesamiento del lenguaje natural, a cuyo desarrollo contribuyeron la doctora Verdejo y su equipo de investigación, es el momento de volver la mirada al presente y enfocarla hacia Europa. Y es que es allí donde el grupo de la UNED tiene puestos gran parte de sus esfuerzos, a día de hoy, a través de su participación en dos proyectos comunitarios. Ambos tan necesarios como prácticos para el ciudadano de a pie. Para explicarlos podemos usar los sencillos ejemplos que nos brinda Felisa Verdejo con la sintaxis narrativa de quien, además de la investigación, ejerce la docencia.

El primero de ellos parte de la proliferación de sitios web relacionados con la medicina y la ingente información al respecto. “Existe una preocupación por certificar dónde se está dando información fundamentada y dónde no. Este proyecto trata de ver cómo se podría construir una herramienta de ayuda a estos certificadores de sitios web para automatizar así su tarea”, explica Verdejo.

En relación al segundo trabajo, la doctora parte de una hipótesis: “Supongamos que un profesor está estudiando el Siglo de Oro español con sus alumnos y quiere preparar una visita didáctica. Entonces se pregunta qué cosas interesantes puede encontrar en los museos de Madrid o a qué actividades puede acudir con ellos. Nuestra idea es desarrollar ‘metabuscadores’ especializados en un tipo de dominio, como por ejemplo, el patrimonio cultural, para que cuando este profesor haga su consulta, la respuesta sea una serie de resultados preordenados: libros, eventos... De este modo, el buscador ayudaría a explorar toda esa información de manera relacionada y en profundidad”.

Como si la investigación no exigiera ya de por sí una dedicación constante, Felisa Verdejo Maillo saca tiempo para invertirlo en otro tipo de trabajos. Así por ejemplo, Felisa y su equipo se encargan de evaluar algunas de las tareas de concurso CLEF (Cross Language Evaluation Forum), una competición a través de Internet que fomenta la investigación de los sistemas de recuperación multilingüe y multimodal, como explica la propia Verdejo: “Cuando usamos varios sistemas diferentes que dicen que hacen ‘x’ o ‘z’, tenemos que ver qué criterios hay que establecer para saber si estos sistemas son comparables o no. Por ejemplo: hacemos una serie de preguntas y estos sistemas tienen que contestarlas y encontrar la solución. Después nosotros comparamos las respuestas. Así en este concurso mi grupo prepara la competición definiendo las tareas que tienen que realizar los participantes, estableciendo el conjunto de preguntas con las que se deben entrenar los sistemas, tratando los resultados, comparándolos...etc.”

De la universidad a la empresa

La investigación en tecnologías de la lengua traspasa los centros de educación superior y los laboratorios para aterrizar en la sociedad de consumo mediante la formación de empresas. Sin embargo, en opinión de Felisa Verdejo, estas nuevas compañías aún tienen la vida un poco difícil porque, a pesar de las previsiones más optimistas que les auguran un futuro prometedor, son necesarias inversiones muy fuertes para obtener un buen rendimiento del producto o servicio que se

La riqueza lingüística de España es un valor añadido para todos los grupos que trabajamos aquí

comercializa. “Creo que cada vez va a haber más posibilidades en el mercado, pero no para desarrollar aplicaciones completas sino, por ejemplo, productos interactivos. Se trata de encontrar nichos donde tú puedas ofrecer una solución que, aunque no sea completamente automatizada, merezca la pena por la ayuda interactiva que ofrezca al usuario”, afirma la profesora Verdejo, desvelando la clave que podría abrir las puertas del éxito a las nuevas empresas especializadas en tecnologías de la lengua: “El reto es encontrar aquellos problemas donde, con el estado de la tecnología, tú puedas dar un valor añadido para contribuir, o bien a que la solución del problema mejore para la persona que lo va a ejecutar de forma interactiva, o bien porque el propio sistema admita cierto nivel de imperfección a pesar del cual pueda resolver una tarea determinada”.

“Estamos en el mapa”

Con esta afirmación, que no deja lugar a dudas, Felisa Verdejo resuelve la incógnita de la situación actual que atraviesa la investigación española en materia de tecnología lingüística. Y es que, como afirma la doctora Verdejo, uno de los referentes internacionales en esta disciplina, España ha experimentado una evolución significativa en los últimos 10 años, un salto cualitativo que sitúa a nuestro país en el mapa mundial de la investigación en lingüística computacional, tanto a nivel de equipos de trabajo como a aportaciones de investigadores particulares.

Sin embargo aún queda mucho camino por recorrer. El primer paso, a juicio de la profesora Verdejo, es ampliar la masa crítica, es decir, contar con más recursos humanos que trabajen en este ámbito científico. Pero no sólo. Esta mejora debe completarse combinando la forma de plantear la carrera investigadora con el modo de tener grupos consolidados y con suficiente masa crítica. Siempre sin olvidar la necesaria colaboración de los diversos equipos para construir recursos tecnológicos que estén a disposición de todo el mundo. Una fórmula de optimización de resultados que Felisa Verdejo Maillo presenta a través de una oportuna comparación con la llamada comunidad del ‘software libre’: “Imaginemos que hay un software que esta ahí porque alguien lo ha desarrollado y yo lo necesito. Entonces lo cojo, lo uso y añado algo nuevo. De este modo la aplicación progresa continuamente. Ésa es la idea que tendríamos que tener en mente cuando hacemos los planteamientos de todos los productos de tecnología lingüística. Y ése es el salto cualitativo que nos hace falta para avanzar con más peso en esta área”.



Hay que poner énfasis en la colaboración en materia de tecnologías de la lengua. Los proyectos deben ser muy cooperativos para que la aplicación final recoja las necesidades y especificaciones de cada lengua y de cada comunidad investigadora, de modo que todos podamos reutilizarla de la mejor manera posible

FICHA TÉCNICA

Centro: Departamento de Lenguajes y Sistemas Informáticos de la UNED

Investigador: Felisa Verdejo Maillo

Dirección: c/ Juan del Rosal, 16
28040 Madrid

Teléfono: 91 398 64 84

Email: felisa@lsi.uned.es

Página web: www.lsi.uned.es/

Líneas de investigación: Aplicación de la ingeniería lingüística a la recuperación de información multilingüe y multimodal, Inteligencia Artificial, Evaluación de sistemas de búsqueda y recuperación de información.