

El grupo Transducens, de la UA, coordina uno de los cinco proyectos europeos aprobados en tecnologías de la traducción - Información - 19/01/2021

El grupo Transducens, de la UA, coordina uno de los cinco proyectos europeos aprobados en tecnologías de la traducción

► La Agencia Ejecutiva de Innovación y Redes de la Comisión Europea invertirá más de 680.000 euros en el proyecto, el cual tendrá una duración de veinticuatro meses y será coordinado por el doctor Miquel Esplà Gomis

R.E.

■ El grupo de investigación Transducens, perteneciente al Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante, será el coordinador de la acción 2020-EU-IA-0078 «Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages», concedida en el marco de la convocatoria CEF-TC-2020-1 - eTranslation, en el que participarán como socios el Instituto Jozef Stefan de Liubliana de Eslovenia y la Universidad de Groningen de los Países Bajos, además de Prompsit Language Engineering, empresa surgida del grupo Transducens. El nuevo proyecto comenzará previsiblemente en junio de 2021, tendrá una duración de veinticuatro meses y será coordinado desde la UA por el doctor

Miquel Esplà Gomis.

Para el desarrollo de esta acción, la Agencia Ejecutiva de Innovación y Redes de la Comisión Europea invertirá más de 680.000 euros. Esta es una de las cinco propuestas aceptadas en el ámbito de las tecnologías de la traducción.

Con esta nueva acción la UA continúa, reenfoca, y pasa a liderar el trabajo realizado durante los últimos años como socio de las acciones 2016-EU-IA-0114 «Provision of web-scale parallel corpora for official European languages» (2017-2019); 2017-EU-IA-0178 «Broader provision of web-scale parallel corpora for official European languages», (2018-2020); y la acción 2018-EU-IA-0063 «Continued Web-Scale Provision of Parallel Corpora for European Languages», vigente en la actualidad.



Las personas de contacto del proyecto, Miquel Esplà Gomis, en calidad de coordinador; Antonio Toral, doctorado en el Departamento de Lenguajes y Sistemas Informáticos e investigador principal a Groningen; Gema

Ramírez, CEO e investigadora principal por parte de Prompsit; y Nikola Ljubešić, investigador principal en el Instituto Jozef Stefan, ya trabajaron conjuntamente en la acción Marie Curie «Abu-Matran, Automatic building of

Machine Translation» (2013-2016) de temática muy similar.

El componente Automated Translation del Mecanismo Conectar Europa (CEF) provee de servicios de traducción automática en las lenguas oficiales de Europa a un amplio conjunto de instituciones y empresas europeas. Los recursos más importantes para construir un servicio de traducción automática son los corpus paralelos, es decir, las colecciones de textos traducidos. El presente proyecto continúa el esfuerzo de los anteriores en dos frentes: mejorar la herramienta libre de código abierto Bitextor para recoger textos paralelos masivamente a partir de los dominios de Internet de nivel superior, en particular los de Islandia, Bulgaria, Malta, Croacia, Eslovenia y Turquía, y hacerlo seleccionando los que son relevantes para las infraestructuras de servicios digitales europeas. Los corpus resultantes serán automáticamente enriquecidos con información sobre su calidad, la variante lingüística, su procedencia (texto original o traducido), y con marcas para facilitar la protección de datos personales; y serán evaluados tanto automáticamente como por expertos. El software y los datos se diseminarán de forma efectiva para que sean adoptados por los actores económicos y sociales; en particular, se suministrarán al servicio eTranslation.