

Jakintza Sortzeko Estatu Azpiprograma 2015, 2.modalitatea: "Explora-Ciencia" eta "Explora-Tecnología" Proiektuak / Subprograma Estatal de Generación de Conocimiento 2015, modalidad 2: Proyectos "Explora-Ciencia" y "Explora Tecnología"

Kodea /Código	Izenburua / Título	Finantzaketa / Financiación			Laburpena / Resumen
		Ministerio de Economía y Competitividad	Anticipo reembolsable FEDER	Esleitutakoa / Total concedido	
TIN2015-72646-EXP / AEI	GENERACIÓN AUTOMÁTICA DE PROCESADORES SEMÁNTICOS	Bai/Si	Ez/No	44.770,00	<p>La ausencia de datos de entrenamiento manualmente anotados es un impedimento fundamental para el desarrollo de procesadores semánticos estadísticos para la mayoría de los idiomas, incluyendo los cuatro idiomas oficiales en España. Este proyecto tiene como objetivo generar automáticamente procesadores semánticos multilingües mediante el uso de herramientas de Procesamiento del Lenguaje Natural (PLN) y corpus paralelos existentes para otros idiomas para su proyección mediante alineamientos de palabras (tradicionalmente usadas en Traducción Automática) a un idioma para el cual no se dispone de datos manualmente anotados. Además, se investigarán técnicas de semi-supervisión para inducir procesadores semánticos robustos a partir de los datos obtenidos mediante las proyecciones de varios idiomas origen. Además de suponer una propuesta altamente novedosa dado que se trata de un problema escasamente investigado, los procesadores semánticos resultantes podrían ser transferidos para su uso tanto en la administración pública como en la industria. Mientras que tanto en entornos industriales como de la administración se necesita procesar textos multilingües de diferentes géneros y dominios, las mejores herramientas de PLN ha sido entrenadas sobre una cantidad limitada de textos cuidadosamente anotadas por lingüistas. Ello quiere decir que una parte muy significativa de recursos deberían ser destinados por la industria para acceder a tecnología de PLN para cada idioma que sea requerido, lo cual obstaculiza seriamente la expansión a otras oportunidades comerciales. Además, esto impide también la adopción del PLN para importantes aplicaciones dentro de la administración pública. Los resultados de CrossText contribuirían a la mejora de los servicios que dependen de PLN tanto en el ámbito empresarial como en la administración pública.</p>