

FÓRUM DE BARCELONA - Actividades

# Las tecnologías de la lengua pueden facilitar el acceso a la numerosa información de Internet

El Fórum acoge la 42ª reunión anual de la Asociación de Lingüística Computacional

LAIA REVENTÓS - Barcelona - 25/07/2004

ATTemps es un sistema de información del tiempo por teléfono. Si usted llama, al otro lado responde una máquina: un sintetizador de voz que reconoce los nombres de los municipios de Cataluña. Cuando le preguntan por uno, da datos sobre temperatura, precipitación, velocidad y dirección del viento o altura de las olas, que obtiene de la red de estaciones meteorológicas. ATTemps funciona desde 2002 y es una de las múltiples aplicaciones de las tecnologías del lenguaje, que mediante voz o texto permiten programar ordenadores para que se comporten como si entendiesen la lengua humana.

Esta disciplina científica reúne las inquietudes de las humanidades con las de la ciencia y la tecnología. Cuatro son sus procesos básicos: reconocimiento de voz, procesamiento del lenguaje natural, traducción de texto a texto y síntesis de voz. Traductores automáticos, diccionarios electrónicos y tutores inteligentes para aprender idiomas son ejemplos de aplicaciones prácticas y cotidianas usadas por miles de personas. Desde el pasado miércoles, 700 investigadores y 200 estudiantes discuten sobre sus avances en el Fórum, recinto que acoge la 42ª edición del Congreso Internacional de la Asociación de Lingüística Computacional (ACL), integrado en el diálogo *Comunicación y lenguaje en la era digital*.

Uno de los objetivos de estas tecnologías es romper las barreras que imponen los idiomas; por ejemplo, con una agenda electrónica (PDA) con cámara de fotos incorporada que retrata símbolos y señales, y los traduce a otras lenguas. El prototipo ha sido diseñado para que los visitantes que acudan a los Juegos Olímpicos de Pekín (2008) sean capaces de comprender los símbolos chinos. La Mesa Aumentada, un sistema informático para obtener datos, ha compartido espacio con la agenda electrónica en la exposición de novedades del congreso. Este prototipo puede archivar charlas de seminario para después ofrecer un resumen con los contenidos más importantes o reproducir al completo la conversación.

Los beneficios de la lingüística computacional también llegan al campo de la enseñanza con sistemas para enseñar a los niños a leer. En la sesión inaugural, el profesor estadounidense Jack Mostow presentó el programa Listen de reconocimiento de voz, implantado en 2000 ordenadores de nueve colegios y probado por niños de 5 a 14 años durante este curso escolar. "Listen no pretende suplir a los maestros, sino contribuir a que en un aula de 30 niños cada alumno pueda tener, además del profesor, la atención personalizada de una máquina atenta a sus necesidades", concreta la escocesa Johanna Moore, presidenta de la ACL.

Los retos, en una disciplina relativamente nueva que ha reducido considerablemente sus tasas de error, son múltiples. Por ejemplo, mejorar las traducciones automáticas. En Internet, buscadores como Google "traducen con rapidez y gratuitamente cualquier texto, pero su calidad todavía está muy lejos de la que ofrecen los humanos", dice Mark Johnson, ex presidente de la ACL.

Otro desafío en el que se trabaja es mejorar los sistemas de búsqueda: reorganizar eficientemente la gran cantidad de información que circula por Internet -que puede provocar desinformación- para que el internauta obtenga una respuesta concreta a su pregunta. Es decir, si quiere saber quién inventó la imprenta, que el sistema responda Guttenberg, en lugar de ofrecer una lista infinita de documentos que contienen las palabras *invento* o *imprenta*, tal como sucede ahora.

## Transferir tecnología

España cuenta con una amplia comunidad de científicos que investigan en tecnologías de la lengua, un campo muy interdisciplinar integrado por traductores, filósofos, lingüistas, ingenieros informáticos y de telecomunicaciones... La Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), fundada en 1984, agrupa a más de 300, entre socios y empresas, y celebró su vigésimocongreso el día anterior a la cita en el Fórum. Manuel Palomar, catedrático de Lenguajes y Sistemas Informáticos de la Universidad de Alicante y presidente de la SEPLN, asegura: "Ahora nos encontramos en un buen momento para transferir

tecnologías de las que la sociedad española se beneficie". Algunas llevan tiempo en la calle. De la página del Gobierno del País Vasco se puede descargar un corrector ortográfico del euskera, desarrollado por la Universidad del País Vasco en 1994. "Xuxen está bastante extendido y ha contribuido activamente a la unificación de la lengua", asegura la investigadora Arantxa Díaz de Ilarraza.

En abril la empresa Thera, salida del grupo de investigación Clic de la Universidad de Barcelona, presentaba un programa que permite clasificar documentos procedentes de cualquier fuente electrónica a gran velocidad (más de 65.000 palabras por segundo). Todas las empresas y administraciones que manejan grandes volúmenes de información son sus clientas potenciales.

© **Diario EL PAÍS S.L.** - Miguel Yuste 40 - 28037 Madrid [España] - Tel. 91 337 8200  
© **Prisacom S.A.** - Ribera del Sena, S/N - Edificio APOT - Madrid [España] - Tel. 91 353 7900