



sica y no de política. Es decir, se puede calcular la probabilidad de su temática. El PLN abarca varias tareas, una de ellas es la traducción automática, otra es la capacidad de resumir. En la primera, el producto va a ser el mismo texto en dos o más idiomas y en la segunda es una síntesis de un documento extrayendo y concatenando las frases más importantes.

Procesar el lenguaje natural es muy difícil, por ello insisto en que no debemos dejarlo sólo en manos de lingüistas, pero al mismo tiempo los programas de PLN pueden ser tan complejos que no debemos dejarlos sólo en manos de informáticos. Se necesita entonces la participación de diversas especialidades, pues en cada una hay diferentes ideas que pueden usarse en PLN.

La variedad y complejidad del lenguaje humano son muy grandes y hacen su riqueza. La presencia cada vez mayor de páginas en la Web con temáticas tan diversas implica una masa de documentos enorme. La capacidad que tiene el ser humano para procesar esta información es limitada y no tenemos el tiempo suficiente, de ahí la necesidad que las máquinas nos ayuden. Las necesidades son evidentes pues la Web es como un organismo vivo.

Hay idiomas que son aglutinantes, donde las palabras no se separan, como el alemán. Para un ser humano (que comprenda el alemán) es fácil entenderlo, pero para una máquina es muy difícil. Hay otros idiomas como el árabe donde se eliminan las vocales y esto complica el procesamiento. Típicamente lo que nuestros métodos procesan son los lenguajes europeos donde las palabras se separan mediante espacios en blanco. Una vez que el sistema produce resultados, hay que evaluar si son correctos. Generalmente se acude a un evaluador humano que, por lectura directa evalúa los resultados.

Esto puede ser muy costoso en recursos, además de subjetivo. Sin embargo, estamos trabajando para que la evaluación de los sistemas que producen lenguaje natural sea también un proceso automático. Es decir, estamos escribiendo sistemas para evaluar automáticamente a los sistemas que procesan el lenguaje natural.

Pienso que los programas que nosotros desarrollamos deben estar a la disposición de todo el mundo, nosotros trabajamos en una universidad donde los fondos que soportan esas investigaciones son fondos públicos, entonces los productos que obtenemos de nuestras investigaciones los ponemos en la red, de forma gratuita, para beneficio de la sociedad. Investigaciones en PLN se hacen también en empresas, pero como son productos para vender, generan patentes para proteger la tecnología. En lo personal no me interesan las patentes porque una patente representa un conocimiento cerrado y a mí me interesa el código libre, abierto y gratuito, además de las publicaciones científicas en revistas y con-

gresos. Las publicaciones permiten difundir los avances científicos en la sociedad.

Mi estancia en el GIL —grupo reconocido internacionalmente— ha sido muy grata, en un principio estaba solo enfocada a realizar el proyecto de similitud textual, pero el Grupo de Ingeniería Lingüística es muy creativo y ya estamos involucrados también en la generación de evaluadores automáticos de los sistemas de resumen automático, en la clasificación de documentos y definiciones y en la comprensión (simplificación) de frases, entre otros proyectos.

Empecé mi sábado en el GIL en noviembre y estaré aquí algunos meses más. Durante este tiempo voy a impartir un curso sobre PLN que se va a centrar sobre las técnicas de generación de resúmenes y minería de textos y estará dirigido a estudiantes de ingeniería y de filosofía y letras. Es interesante juntar ambas poblaciones de estudiantes para que vean que en ambas disciplinas hay aportaciones interesantes al PLN. El modelo educativo ha hecho este divorcio entre disciplinas que no debe tener lugar.

También queremos crear una maestría en Lingüística Computacional en la que participen México, Francia, España, y Cuba. Esta maestría estará dirigida a estudiantes de licenciatura con buenas notas académicas y motivados del área de humanidades, ciencias exactas, ingenierías o ciencias sociales. Me relacioné con el GIL a partir de una conferencia que impartí en México sobre Resumen Automático en la UAM y asistieron varios integrantes del grupo, incluido el Dr. Gerardo Sierra, jefe del GIL. Entonces vimos la oportunidad de hacer una sinergia entre nuestros grupos, de forma que me invitó en una codirección de tesis de un alumno de maestría, Alejandro Molina, quien actualmente realiza su doctorado en Avignon bajo la cotutoría de ambos. El Dr. Sierra ha efectuado un par de estancias conmigo y en respuesta estoy aprovechando mi sábado aquí en este Instituto. Puedo afirmar que ha sido una relación muy provechosa —concluyó—. 🍷