Satoshi Sekine: Minimally Supervised Knowledge Discovery

14 noviembre 2007, ETSI Telecomunicación, UPM

Resumen del Seminario

For a decade, corpus has been taking the central role in Natural Language Processing studies. The supervised learning methods were most effective to solve tasks which can be translated into labeling problem of items with a handful number of classes. However, we are observing difficulty of applying supervised learning methods to most of the tasks which include semantic knowledge, such as Information Extraction or Question Answering as a whole. The semantic knowledge is more sparse than general supervised methods can learn from training corpus.

The crucial problem in creating systems in semantic domain is that a vast amount of knowledge is needed to create reasonable system. In other words, we have to overcome the sparseness problem and scalability problem. In order to solve it, we believe minimally supervised (i.e. un-supervised and/or semi-supervised) learning methods utilizing a large un-annotate corpus would be a possible solution. The corpus available in electric form at this moment is vast enough to make us believe that the knowledge to understand most of the world is written down consciously or unconsciously somewhere in the corpus. The new paradigm is aiming at a reformulation of the knowledge scattered in the corpus into the shape in which a system can use to solve a task.

In the talk, I would like to summarize the studies which have been conducted in the field. It would hopefully give the audience the overview of the field and possible research directions in the future. I will also descrive several own studies.

Satoshi Sekine

Satoshi Sekine is an Assistant Research Professor at New York University. He received his MSc at UMIST, UK in 1992 and his PhD in 1998 at NYU. He has been working on various topics, including parser, NE, Information Extraction and minimally supervised knowledge discovery. Recently, he organized Workshop on Textual Entailment and Paraphrasing (ACL-07), Web People Search task at SemEval-07 and served as guest editor on the special issue on Named Entity. He lead a middle size NSF project for 5 years on On-Demand Information Extraction.

Destinatarios

Este seminario está enfocado a un público especializado e interesado en los sistemas de Búsqueda de Respuestas, Extracción de Información y en aplicaciones de Procesamiento del Lenguaje Natural. Se ha organizado pensando concretamente en los investigadores y estudiantes de doctorado y máster de los grupos integrantes en MAVIR.

La asistencia es abierta y gratuita.

Lugar de celebración y horario

Miércoles 14 de noviembre de 2007. 16h00-17h30

Escuela Técnica Superior de Ingenieros de Telecomunicación

Salón de Grados, Edificio A.

Avenida Complutense s/n

Ciudad Universitaria

E-28040 Madrid

Cómo llegar y planos: http://www.etsit.upm.es/

¿Qué es MAVIR?

MAVIR es una red de investigación financiada por la Comunidad de Madrid y formada por un equipo multidisciplinar de científicos, técnicos, lingüistas y documentalistas para desarrollar un esfuerzo integrador en formación y en transferencia de tecnología, cubriendo las siguientes áreas de investigación:

- Tecnologías de la Lengua aplicadas al Acceso a la Información Multilingüe
- Comunicación científica a través de la Red

Contacto

Víctor Peinado, técnico de gestión de la Red MAVIR

email: <victor AT lsi.uned.es>

Tel: 91 398 8106