



Distinguished Lecture Series

Semantics-based information access to large scale text bases



Monday, September 6th, 2010 10:00am
Auditorium 106 at new IIS Building

Dr. Jun'ichi Tsujii

Professor, Department of Computer Science, University of Tokyo, Japan
Professor, School of Computer Science, University of Manchester, UK

Abstract

Due to rapid development of corpus-based NLP and statistical modeling of language, NLP techniques such as shallow or deep parsing, named entity recognition, event or relation extraction, topic detection, machine translation, etc. have become ready for practical, large-scale application.

In particular, NLP techniques will provide practical solutions of resolving the difficulty of scaling up in semantic information access envisioned in Semantic Web. They not only provide (semi-)automatic means of meta-data extraction, large scale ontology building, etc. but also enable fine-grained information access to unstructured data (text) by associating semantic meta-data with fine-grained units (sentences, phrases) in text. In this talk, I will present a research plan toward fine-grained semantic information access via NLP and related research challenges.

Examples of fine grained access to information in text will be illustrated by a system (MEDIE) which we have constructed for text mining for biology, together with a workflow management of large scale NLP on large clusters of processors.

For more information: <http://www.iis.sinica.edu.tw/>

