



CIS Distinguished Lecture Series

Information Extraction: Knowledge Discovery from Text

Ralph Grishman

New York University

November 5, 2008

Abstract: Much of the information on the Web is encoded as text, in a form which is easy for people to use but hard for computers to manipulate. The role of information extraction is to make the structure of this information explicit, by creating data base entries capturing specified types of entities, relations, and events in the text. We consider some of the challenges of information extraction and how they have been addressed. In particular, we consider what knowledge is required and how the means for creating this knowledge has developed over the past decade, shifting from hand-coded rules to supervised learning methods and now to semi-supervised and unsupervised techniques.

Bio: Ralph Grishman is Professor of Computer Science at New York University. He has been involved in research in natural language processing since 1969, and since 1985 has directed the Proteus Project, with funding from DARPA, NSF, and other Government agencies. The Proteus Project has conducted research in natural language text analysis, including information extraction, and has been involved in the creation of a number of major lexical and syntactic resources, including Complex, Nomlex, and NomBank. He is a past president of the Association for Computational Linguistics and the author of the text *Computational Linguistics: An Introduction*.

When: 3-4pm, Wednesday, November 5

Where: 447 Wachman Hall

Refreshments will be served!