



綜合型語言知識庫及其發展

俞士汶教授

北京大學資訊科學技術學院

摘要

任何一個自然語言處理系統都需要一個適配的語言知識庫。北京大學計算語言學研究所積 30 年之努力，建成以《現代漢語語法資訊詞典》為基礎的綜合型語言知識庫。本報告介紹綜合型語言知識庫的主要內容及其發展。當前語言計算研究已深入到面向語言大資料的語義計算。詞彙、詞法、句法層面的語言知識庫對語言淺層分析技術的發展起到了至關重要的作用，這個成功的經驗對當前的語義計算研究有啟示意義。一項新的語義知識資源建設的任務是確定句子中的謂詞並標注其論元的語義角色。本報告介紹這項任務的概況，論述對這項語義知識資源工程的認知並初步總結參與這項工程的實踐經驗。



簡 歷



俞士汶，男，1938 年 12 月出生。北京大學資訊科學技術學院教授。1964 年自北大計算數學專業畢業後一直在北大從事電腦學科的研究與教學工作。

1986 年起致力於計算語言學和自然語言處理的研究。作為第一完成人的主要研究成果有以《現代漢語語法資訊詞典》為基礎的“綜合型語言知識庫”，在獲得中國政府部門、中國全國性學術團體以及北京大學的多項獎勵後，於 2011 年獲中國國家科學技術進步獎二等獎。同年獲中國中文資訊學會成立 30 年來首次頒發的終身成就獎。著作 8 本（其中 2003 年出版的《計算語言學概論》2016 年獲評北京大學優秀教材），第一作者論文約 110 篇。作為博士生、碩士生的導師和博士後、訪問學者的合作教師，培養了一大批計算語言學領域的高端人才，於 2012 年獲北京大學訪問學者優秀導師榮譽。“綜合型語言知識庫”於 2013 年再次獲北京大學首屆產學研結合特別貢獻獎。當前參加中國國家重點基礎研究項目之 973 課題“融合三元空間的中文語言知識與世界知識獲取和組織”的研究工作。

地點： E11-4045

日期： 12/04/2017，星期三

時間： 11:00~12:00

Contacts: Dr. Derek F. Wong / Dr. Lidia S. Chao
Telephone: (853) 8822-8051
Website: <http://nlp2ct.cis.umac.mo/>

Address:
E11-2023, University of Macau,
Avenida da Universidade,
Taipa, Macau, China