

计算语言学概论

Computational Linguistics: An Introduction

詹卫东

北京大学中文系

<http://ccl.pku.edu.cn/doubtfire/>



课程安排

- 第一章 计算语言学概述 (2学时)
- 第二章 语言知识的形式化 (3学时)
- 第三章 汉语的形式句法体系 (3学时)
- 第四章 语义知识的形式表示 (4学时)
- 第五章 语篇知识的形式表示 (2学时)
- 第六章 语料库 (2学时)
- 第七章 词法分析 (6学时)
- 第八章 句法分析 (6学时)
- 第九章 机器翻译 (2学时)
- 第十章 信息检索 (2学时)

} 基础篇(16h)

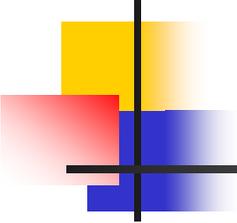
} 算法篇(12h)

} 应用篇(4h)



授课对象与教学目标

- 授课对象：
中文系语言学专业研究生和高年级本科生。
- 教学目标：
通过本课程的学习，希望达到两个目标。
 1. 具备以计算机为背景思考语言学问题，对语言知识进行形式化描述的能力；
 2. 了解计算语言学的基本理论和自然语言处理中的常用模型及算法，初步具备从事相关领域研究工作的能力。

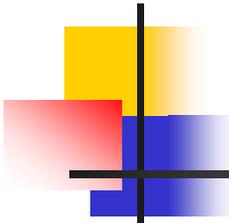


参考资料

1. 钱锋（1990）《计算语言学引论》，学林出版社1990年版。
2. 陆致极（1990）《计算语言学导论》，上海教育出版社 1990年版。
3. 刘开瑛、郭炳炎（1991）《自然语言处理》，科学出版社 1991年版。
4. 冯志伟（1991）《数学与语言》，湖南教育出版社1991年版。

5. 白硕（1995）《语言学知识的计算机辅助发现》，科学出版社1995年版。
6. 冯志伟（1995）《自然语言机器翻译新论》，语文出版社1995年版。
7. 姚天顺等（1995）《自然语言理解 —— 一种让机器懂得人类语言的研究》，清华大学出版社、广西科学技术出版社1995年版。
8. 冯志伟（1997）《自然语言的计算机处理》，上海外语教育出版社1997年版。
9. 吴立德 等（1997）《大规模中文文本处理》，复旦大学出版社1997年版。

10. 翁富良、王野翊（1998）《计算语言学导论》，中国社会科学出版社1998年版。
11. 俞士汶 等（1998）《现代汉语语法信息词典详解》，清华大学出版社、广西科学技术出版社1998年版。
12. 陈小荷（2000）《现代汉语自动分析》，北京语言文化大学出版社2000年版。
13. 赵铁军（2000）《机器翻译原理》，哈尔滨工业大学出版社2000年版。
14. 詹卫东（2000）《面向中文信息处理的现代汉语短语结构规则研究》，清华大学出版社、广西科学技术出版社2000年版。



参考资料（续）

15. James Allen, 1995, Natural Language Understanding, The Benjamin/Cummings Publishing Company, Inc.
16. Christopher D. Manning & Hinrich Schutze, 1999, Foundations of Statistical Natural Language Processing, The MIT Press.
17. Daniel Jurafsky & James H. Martin, 2000, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall Inc..
18. Michael P. Oakes, 1998, Statistics for Corpus Linguistics, Edinburgh University Press.
19. Anthony Woods, Paul Fletcher, Arthur Hughes, 1986, Statistics in Language Studies, Cambridge University Press.
20. Graeme Kennedy, 1998, An Introduction to Corpus Linguistics, Addison Wesley Longman Limited.

21. <http://ccl.pku.edu.cn/doubtfire/>
22. <http://www.nlp.org.cn/>

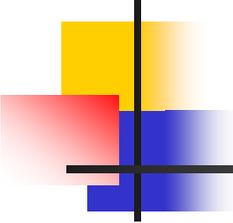
23. 俞士汶 常宝宝 詹卫东, 2003, 《计算语言学概论》, 商务印书馆。
24. 刘颖, 2002, 《计算语言学》, 清华大学出版社。

25. 冯志伟, 2004, 《机器翻译研究》, 中国对外翻译出版公司。
26. 冯志伟、孙乐译, 2005, 《自然语言处理综论》, 电子工业出版社（原著D.Jurafsky & J.H.Martin, 2000）
27. 苑春法 等译, 2005, 《统计自然语言处理基础》, 电子工业出版社（原著C.D.Manning & H.Schutze, 1999）
28. 刘群 等译, 2005, 《自然语言理解》, 电子工业出版社（原著James Allen, 1995）
29. 宗成庆, 2008, 《统计自然语言处理》, 清华大学出版社。
30. 冯志伟, 2010, 自然语言处理的形式模型, 中国科学技术大学出版社。



考核

- 两次作业 (2 × 20%)
- 期末报告 (60%)



致谢

- 在撰写本课程讲义和教学过程中，得到许多专家学者的帮助：

俞士汶、常宝宝、于江生、孙斌、张化瑞、胡俊峰（北京大学计算语言学研究所）

刘群、王斌、白硕（中科院计算所）

周强、刘颖（清华大学）

在此深表谢意