**中文简介**

自然语言是我们平时每天交流所使用的语言，比如英语，汉语等等。不同于编程语言，自然语言更难进行分析和处理。同时，基于自然语言处理的技术正在得到广泛的应用。比如智能问答，机器翻译等等。

《自然语言处理导论》是一门面向本科生开设的任选课程。本课程属文理交叉性质，由北大计算机系和中文系联合开设，两方教师联合讲授。重视自然语言处理的基本思想、概念和常见算法的讲解，鼓励学生深入思考，在讨论、比较中获得对相关问题的理解。在程序设计课程的基础上，本门课程旨在向同学们介绍：自然语言处理的概率统计基础，自然语言处理的语言学基础以及自然语言处理的基础应用知识。同时，希望就一些热点课题，比如，机器翻译以及深度学习等进行入门介绍。本门课程除了对理论的介绍，还安排了实践任务。通过理论结合实践的教学方式，我们希望可以培养学生的创新精神和实践能力，促进学生的知识、能力、素质的综合提高。

**英文简介**

This is a book about Natural Language Processing. By “natural language” we mean a language that is used for everyday communication by humans; languages such as English, Hindi, or Portuguese. In contrast to artificial languages such as programming languages and mathematical notations, natural languages have evolved as they pass from generation to generation, and are hard to pin down with explicit rules. Technologies based on NLP are becoming increasingly widespread. For example, phones and handheld computers support handwriting recognition; machine translation allows us to retrieve texts written in Chinese and read them in Spanish. By providing more natural human-machine interfaces, and more sophisticated access to stored information, language processing has come to play a central role in the multilingual information society.

This course is an introduction for undergraduate students who are interested in natural language processing. This book provides a highly accessible introduction to the field of NLP. The book is intensely practical, containing hundreds of fully worked examples and graded exercises. It can be used for individual study or as the textbook for a course on natural language processing or computational linguistics, or as a supplement to courses in artificial intelligence, text mining, or corpus linguistics.

**参考书**

[1]Speech and Language Processing, Jurafsky, D. and Martin, J.H., 1st Edition Prentice Hall, 2000

中译本：自然语言处理综论，冯志伟等译，电子工业出版社，2005

[2]统计自然语言处理，宗成庆，清华大学出版社，2008

[3]统计学习方法, 李航, 清华大学出版社, 2012

[4]Pattern recognition and machine learning, Bishop, Springer, 2006

[5]Natural Language Processing with Python, Ewan Klein and Edward Loper, O’Reilly Media, 2009, 1st, 978-7-5641-2261-4

**课程大纲**

1：NLP的语言学基础

 语言学知识（I）—— 理论分析：构词、词类、句法、语义

 语言学知识（II）—— 实例分析：语料库与知识库

1.1：构词法与文本自动分词

1.2：词类与词性标注

1.3：句法规则与结构分析

1.4：语义分析

1.5：语料库与知识库

2：NLP的概率统计基础

2.1：NLP的总体介绍

2.2：概率论/信息论基础

2.3：Ngram统计语言模型

2.4：统计机器学习基础

3：NLP的方法和具体应用

3.1：序列标注问题

3.2：句法分析

3.2.1上下文句法分析

3.2.2依存句法分析

3.3：机器翻译概论

**考核方式：**

期中作业（以考核编程实现系统为主）+期末考试（以考核基础理论为主），各占50%比重。最后一堂课进行期末考试。