

“自然语言处理导论”课（语言知识部分） 知识要点
詹卫东 2016-11

第1讲 中文词法分析、汉语构词法

1. 词法分析任务——从字符串到词串：
(1) 中文词法分析面临的问题；(2) 基本方法；(3) 评测指标
知识点：交际型分词歧义，组合型分词歧义，未登录词，
最大匹配分词法，最大概率分词法，字位标注分词法
2. 汉语构词法：
知识点：词、语素、单纯词、合成词、复合词构词模式、词语变形形式

第2讲 中文词性标注、汉语词类体系

3. 汉语的基本句法结构类型有哪些，举例说明各类结构的主要特点是什么
4. 汉语词类划分的依据是什么，主要词类有哪些，主要特点是什么
5. 基于隐马尔可夫模型（HMM）和韦特比（Viterbi）算法的词性标注方法
6. 基于转换的错误驱动的词性标注方法的主要思想及其处理流程

第3讲 句法分析

7. 上下文无关文法，句法结构歧义的类型
8. 句法分析中的 Earley 算法
9. 句法分析中的 LR 算法、Tomita 算法，Tomita 算法对 LR 算法的主要改进是什么

第4讲 语义分析

10. 词汇语义知识的类型（聚合关系知识、组合关系知识）及知识表示的理论模型
知识点：论元结构理论
11. 将语义知识融入到句法结构分析中
知识点：特征结构、合一运算（Unification）
加入了合一运算的 Earley 算法
12. 自然语言语义现象的复杂性
知识点：名词指称义、代词指代义、羡余否定、语义指向