

汉语依存树编辑器的设计与实现

观察“哈工大信息检索研究室汉语依存树库”的数据表示与呈现方式，在此基础上，设计并实现一个汉语依存树编辑器。相关功能包括但不限于如下所述：

A. 单个汉语句子的依存树生成。假定汉语句子的分词、词性标注此前已经完成，设计并实现一个初始的依存树分析结果。基于对现有树库内容的考察，有兴趣的同学可以把它当作一个建模问题，设计、实现自己的结构标注方法。当然，做简单的处理，“掷骰子”的“决策”也可视为一种结构标注方法，用来产生初始的依存树分析结果；

B. 单个依存树的编辑操作和图形化呈现。这涉及数据格式、图形界面形式和相关事件动作的设计处理，希望综合考虑依存树的连线区域、连线方向、连线标签等操作因素，实现人机友好的一些基本编辑、呈现功能。若还能提供操作限制、操作记录与回撤、弧形选择、动画显示、gif 图形输出等扩展功能就更好了；

C. 网络环境下单个依存树的多盲编辑与决策。涉及多人标注修订中的通讯交流、任务分配与集成、超级仲裁等，这些外围功能不做统一要求；

D. 标注过程中的批量依存树管理，即树库管理。包括入库出库、批量替换、用户检索、信息统计等，这些外围功能也不做统一要求。

作业由 2~4 人自由结合完成，期末提交源码及小组报告。报告格式不限，一般包括需求分析、设计、实现、软件使用说明、作业评价与总结、成员分工与贡献等内容，以 15 页左右为宜。考核重在图形界面形式的设计和人机交互手段的实现，鼓励发掘并实现新的功能或性能。

这是 NLP 领域中的一个实用的综合性题目，涉及广泛的 Java 技术。具体做什么、怎么做、侧重点等并无统一要求，视各小组的实际情况决定。希望大家充分讨论交流、享受开发过程并确有收获。