东北大学自然语言处理实验室
字-词还原

东北大学自然语言处理实验室
http://www.nlplab.com

NiuTrans 团队
niutrans@mail.neu.edu.cn

字-词还原

- **问题描述**

词语或短语被逐字切分后就失去了原来的意义，根据字符关系表才可以将其还原。

表 1 逐字切分

句子序号	短语	逐字切分的短语
1	物理实验室	物 理 实 验 室

假设有如下字符关系表:

表 1 字符关系表

字符关系
物 理
实 验
实验 室</w>

表中上下顺序表示关系的紧密程度。使用该表对“物 理 实 验 室”进行还原，由于字符关系中“物 理”处于第一位，因此一级还原结果是“物理 实 验 室”，二级还原结果“物理 实验 室”，三级还原结果“物理 实验室”。其中“实验 室</w>”中的“</w>”表示词语或短语的末尾，如有词为“实验 室 内”，此时的“室”并不是末尾，不能通过“实验 室</w>”这组字符关系还原。

现提供字符关系表，和待还原的切分后短语，请完成每个词的一级还原与多级还原。

- **数据**

当前文件夹 sample-data\目录下

祝好运！