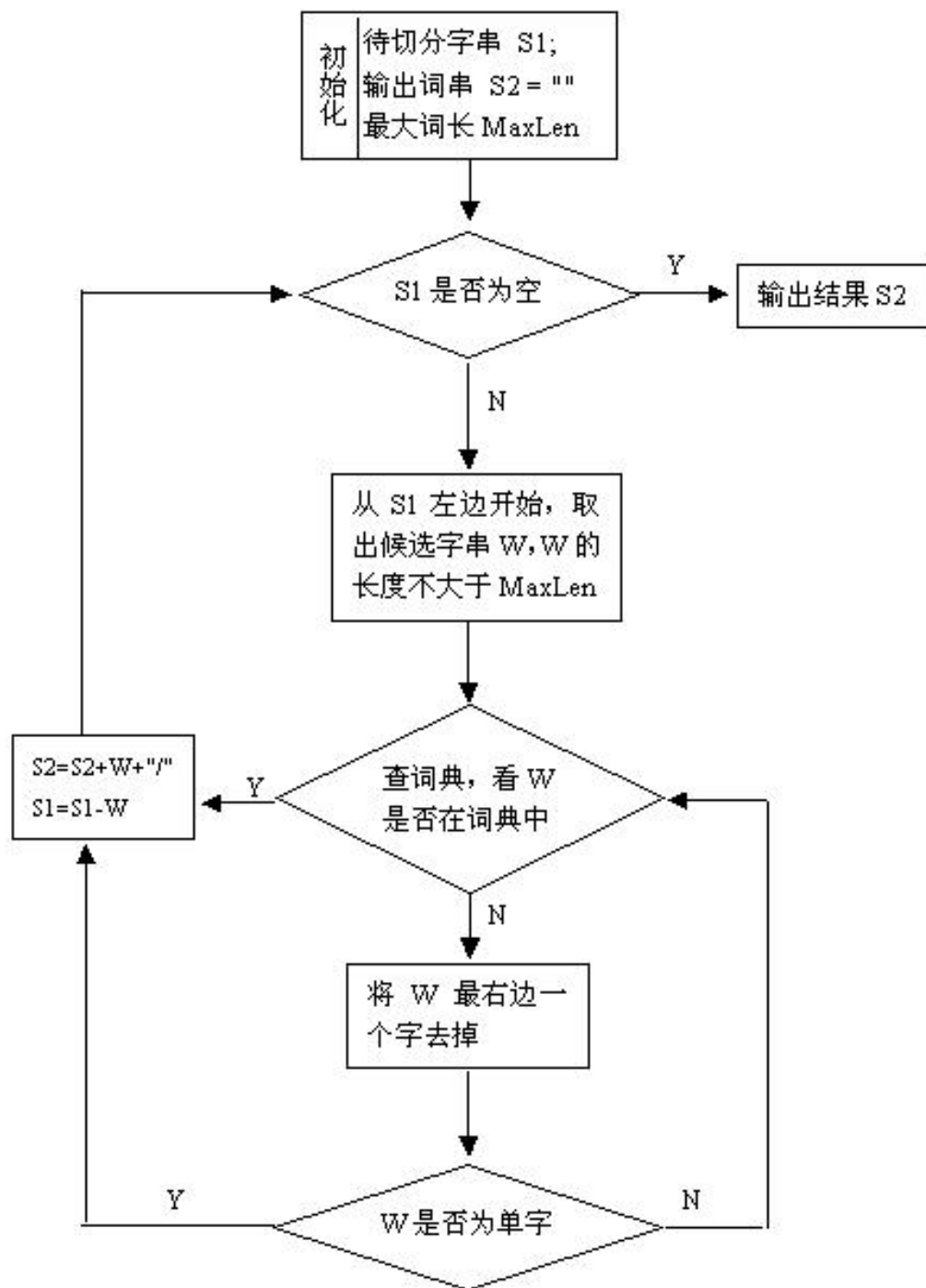
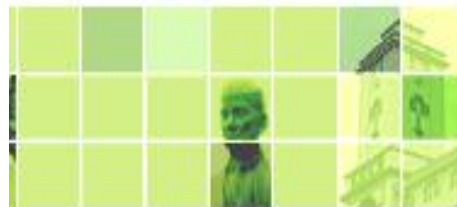




第5节 练习



1. 使用正向最大匹配算法，利用给定的数据：字典文件corpus.dict.txt，对语料corpus.sentence.txt进行分词，将分词的结果输出到文件corpus.out.txt中
2. 选做：计算分词结果的P/R/F，比较正向、逆向、双向匹配的分词效果
3. 使用jieba/ntlk等工具，查看词性标注结果





分词算法评价例子：

句子：我是中国人

分词：我是 中国人

答案：我 是 中国人

识别出的个体总数：2 正确识别的个体总数：1

测试集中存在的个体总数：3 正确识别的个体总数：1

正确率(Precision) = $1/2 = 50.00\%$

召回率(Recall) = $1/3 = 33.33\%$

F值 = $(1/2) * (1/3) * 2 / (1/2 + 1/3) = 40.00\%$



Thank you!