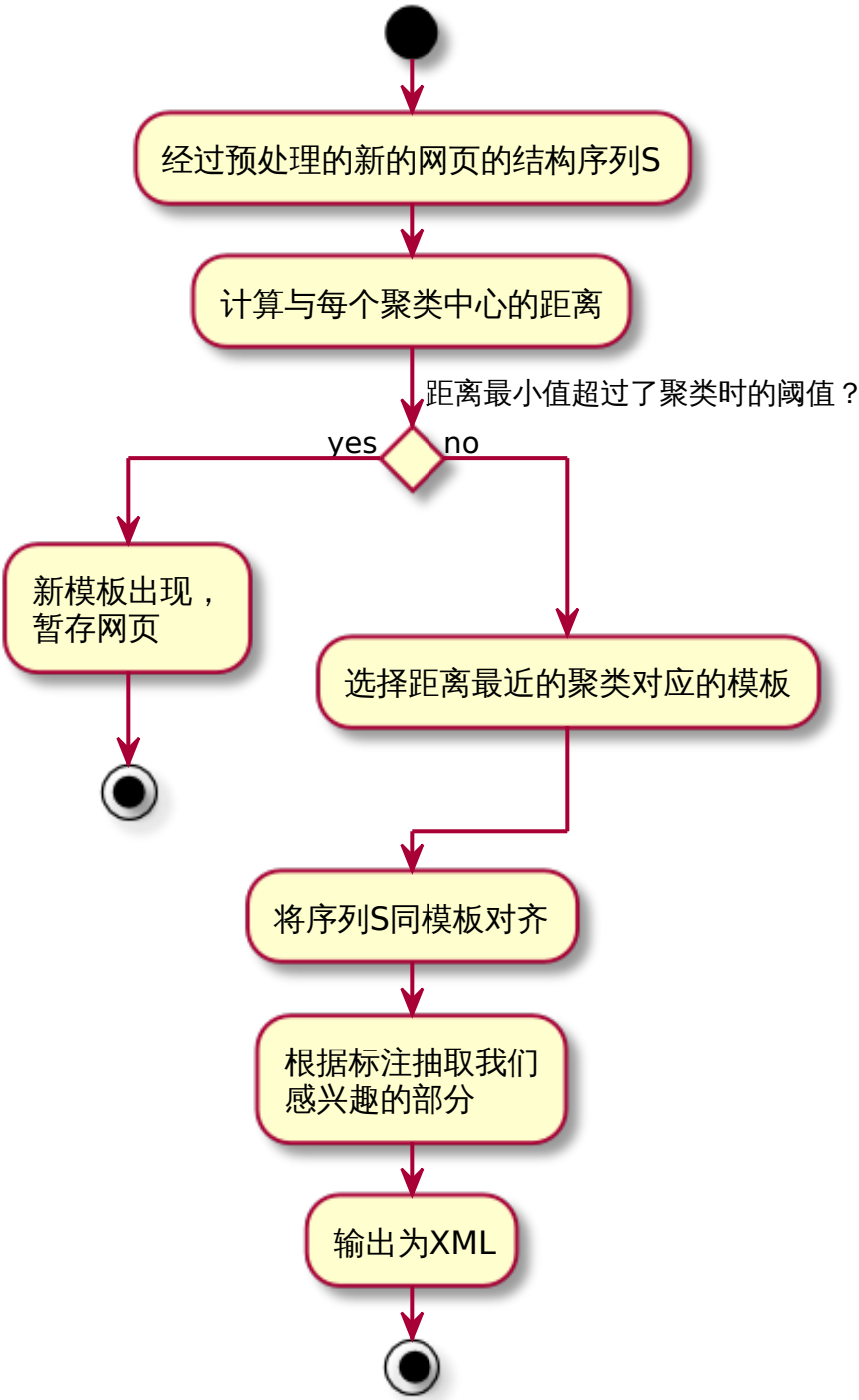
经过预处理的新的网页的结构序列S
计算与每个聚类中心的距离
距离最小值超过了聚类时的阈值？
yes
no
新模板出现，
暂存网页
选择距离最近的聚类对应的模板
将序列S同模板对齐
根据标注抽取我们
感兴趣的部分
输出为XML