

《数字内容安全》期末考试试题

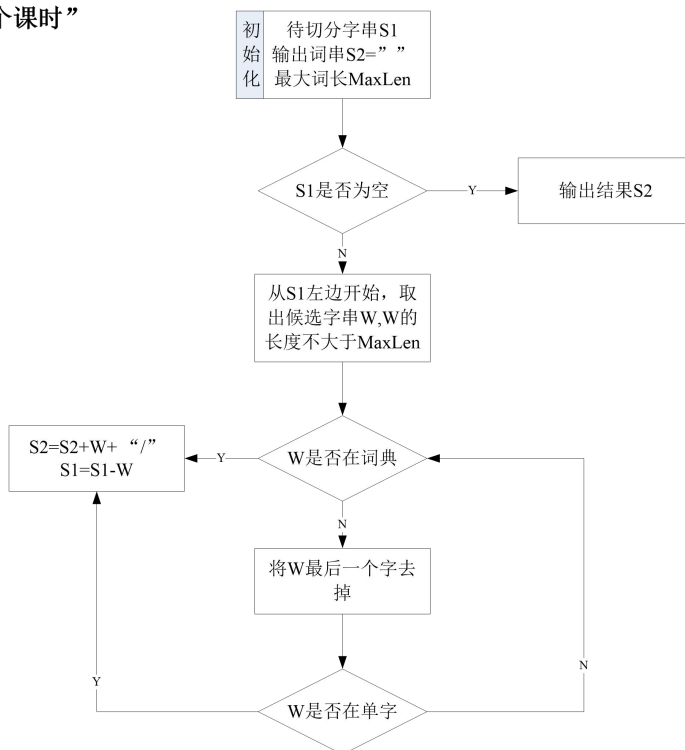
一、请利用正向最大匹配法给下列短语 S1 进行中文分词，输出 S2 作为分词结果，写出具体步骤。（20 分）

S1= “内容安全课程是三个课时”

设定最大词长

MaxLen=5

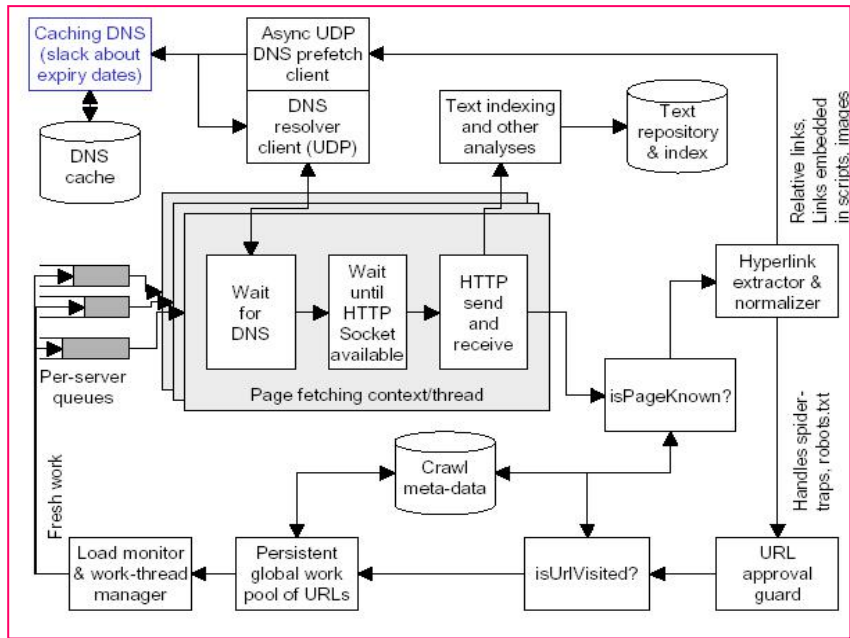
S2= “ ”



二、请根据下图展示的典型网络爬虫的各个功能模块，阐述下列功能模块的主要解决方案。（20 分）

块，阐述下列功能模块的主要解决方案。（20 分）

- （1）链接提取和规格化中，如何避免在重复网页上再提取链接？
- （2）页面采集中，可根据页面质量选择待采集页面。请列出 5 种页面质量计算方法。
- （3）动态页面有几种类型？分别是什么？怎样采集？



三、请列举三种针对社交网络的隐私攻击方式，并给出两种典型隐私保护算法：聚类 K 匿名算法和差分隐私保护算法的基本思想。（20 分）

四、数字化视频的压缩用到多种技术，请解释一下运动补偿、画面插补、关键帧（I 帧、P 帧和 B 帧）的概念。（20 分）

五、请从特征提取、数学模型、样本空间和判别取证四个步骤具体描述图像取证系统基本原理及实现过程。并尝试分析目前的检测技术针对 Deepfake 等换脸视频篡改的检测难点在哪里？（20 分）