**Bin中：input为输入，output为输出，unit为量词库，noum为无量词时名词补充库。**

**解题思路：**

1. 将所有的中文汉字数字均转换成数字；
2. 进行初步切词，按照标点符号切割；
3. 二次切词，过滤每个用标点符号分割的词句中没有数字的数组项，再过滤同一个词句中有几个数字存在的情况，重序列化为数组；
4. 查错切词，主要是过滤类似汉语中“一两“、“三三两两”、“十几”、“二、三”这些情况（此步骤可以放在2之前做，要求颇高）；
5. 针对每个数组项中的切词结果进行匹配量词库；
6. 得出两种情况：有量词与无两次，有量词看7，无量词看8；
7. 有量词时按照数字紧跟量词原则降噪，提取数量，并得到量词后面的字符串，之后与名词库比对；
8. 五量词时直接与名词库比对，按照数字紧跟名词原则降噪，提取数量；
9. 按照数量\*名词的方式输出标准img的html语句，名词库中的名词事先在应用程序发布目录下建立imgs文件夹，并且图片名称与名词名称一致，后缀为.png

**存在问题：**

1. 汉语中的模糊数字与一些动词干扰结果项需建立算法用于提高降噪精度；
2. 切词最终可能存在数量与量词并非紧密连接情况，为提高降噪精度需另辟蹊径提前过滤，构想是讲量词用\*表示，用正则表达式将数字与\*之外的字符去掉来获取准去的数量值。