1、需求分析

本题的主要需求根据用户自行输入的一段文字，是提取出该输入文本中的名词及其数量，并以符号化的形式进行结果显示，主要

约束是对于量词（大小写数字均可）范围为1~20，负数、零、大于20的数量不显示。

2、解决思路

本题考虑到时间上的因素，首先选用开源分词组件库Pangu分词库进行文本分词，实现上可以考虑二元分词、多元分词，限于时间，本题选用二元分词，将输入文本进行前期处理，得到输入文字的分词库；对于得到的分词库，对其进行处理，根据数字（中文数字）提取出本文中带数字的分词词组，然后根据中文语义来进行名词的判断，名词一般出现在两次之后，因此可从当前数字分词向后搜索，找出该两次之后的名词，在查找的过程中需要排除其他词性（在分词的前期处理过程中，可以利用词库判断出该分词的词性，如形容词，副词，动词等），在查找过程中需要排除一些特殊分词，对其进行特殊处理，如文本中出现的人名或地名中带量词的情况（如张三），最后根据得到的量词和名词，建立处理结果库；为了对分析结果进行显示，可以建立名词图片库，根据分析结果和名词图片库，输出分析结果。

3、算法设计

（1）使用Pangu分词对输入文本进行分词，建立分词库；

（2）提取出文本中带量词的分词；

（3）根据提取出的带量词的分词，向后搜索该量词之后的名词（数量为负数、零、大于20的数量不显示），由词法分析可知，名词前不可以是动词或介词等定语成分，得到分析结果；

（4）根据分析结果，结合图片库（可以来源于网络动态爬取，也可自行建立专题图片名词库），将结果以图形化的方式输出。

4、待完善的地方

（1）优化文本中的特殊人名、地名等特殊名词，如张三、三姐等词语；

（2）优化程序中名词的提取算法，结合中文文本语法和词法进行分析，提高识别精度；

（3）优化程序显示结果，建立图片库，结合图片库对结果进行显示；

（4）尝试结合NLB进行分析，通过相应的模型改善问题识别效果。