# Thesis\_processing.py

1. find\_keywords\_indices(doc\_path,find\_key\_list):
   1. find\_keywords\_indices(doc\_path)：“关键词索引函数”输入参数doc\_path（原文件），根据关键词列表识别关键词，最后返回“段落索引列表"，用于后续分割文件
2. split\_doc\_by\_paragraphs(original\_doc\_path, save\_path):
   1. split\_doc\_by\_paragraphs(original\_doc\_path, save\_path):“分割段落函数”将原文件的每一段分割为一个单独的段落文件，段落文件存储在中间文件夹中，以便于后续操作输入参数为“原始文件”与需要存储的“中间文件夹”，无返回值
3. combine\_documents(start\_num, end\_num, base\_path, path\_new):
   1. combine\_documents(start\_num, end\_num, base\_path, path\_new)：“组合段落函数”输入参数分别为起始段、终止段、原段落文件目录、组合后文件名（注意是文件名，需要逐个设置）