1、一种MCT语料筛选工具，包括以下步骤：

S1：医疗文本出题语料筛选步骤

S11：对医疗题干文本进行分词，分词结果存储于数组listOfStr中；

S12：将“中医疾病与病征编码”词典中词汇存储于数组txt中，如果listOfStr中的词汇包含于词典中，则舍弃对应的医疗题干文本。【由此产生中医筛选API。】

S13：若保留医疗题干文本，则依次遍历listOfStr中的每个词语，判断listOfStr中的每个词语是否都存在于“科室-疾病”医学词典department中。若存在，则对字典time中的相同键次数＋1；

S14：遍历结束后，找出字典time中键次数最大的值，即出现次数最多的科室，将该科室名判断为题干文本的类别【由此产生科室判别API】；若类别符合MCT出题要求，则该医疗题干文本保留；

S15：针对保留的医疗题干文本，将HSK中文水平等级的等级词汇“HSK”存储于数组strDict1中；

S16：依次遍历listOfStr中的每个词语，在数组strDict1中寻找词语对应的HSK词汇中文水平等级，求出listOfStr中的每个词语所对应的HSK词汇等级。【由此输出HSK词汇等级标记API】。取最高出现的HSK词汇等级为该文本的语言词汇等级。

S16：依次遍历listOfStr中的每个词语，在数组strDict1中寻找词语对应的MCT词汇等级，求出listOfStr中的每个词语所对应的MCT等级。【由此输出MCT词汇等级标记API】。取最高出现的MCT词汇等级为该文本的MCT词汇等级。若该等级超过预设的MCT等级，则该舍弃对应的医疗题干文本；

S17：若医疗题干文本被保留，则将该医疗题干文本确定为出题语料文本TextQuestion。

具体的，如下将以一个具体的实施例进行说明，已知题干文本“demo.txt”的内容如下表1。

患者：胃会反酸偶尔会痛是什么病。

医生：你好很高兴能帮到你。

医生：胃炎胃溃疡都有这种表现。可以胃镜检查。

患者：那要吃什么药好。

医生：可以吃点奥美拉唑和铝碳酸镁片看看。

患者：两种一起药一起吃吗？

医生：是的。

患者：怎么吃法。

医生：你多大年纪啊。还有其他不适吗。

患者：79

表1

本发明方法对应的实施步骤如下：

S1：医疗文本出题语料筛选步骤

S11：对医疗题干文本进行分词，分词结果存储于数组listOfStr中。数组内容为：

['history', ' ', '患者', '胃会', '反酸', '偶尔', '会痛', '是', '什么', '病', '医生', '你好', '很', '高兴', '能', '帮到', '你', '胃炎', '胃溃疡', '都', '有', '这种', '表现', '可以', '胃镜', '检查', '那要', '吃', '药好', '吃点', '奥美拉唑', '和', '铝', '碳酸镁', '片', '看看', '两种', '一起', '药', '吗', '的', '怎么', '吃法', '多', '大', '年纪', '啊', '还有', '其他', '不适', '79']

S12:将“中医疾病与病征编码” （https://pan.hqu.edu.cn/share/1545f0798a2e625 cf10

f472fb8）中词汇存储于数组txt中。因listOfStr中的词汇不包含于其中，则该医疗题干文本保留。

S13：依次遍历listOfStr中的每个词语，是否存在于“科室-疾病”医学词典（https://pan.hqu.edu.cn/share/6394d8125e42c2d38b419a01e2）department中。若存在，则对字典time中的相同键次数＋1。

S14：遍历结束后，找出字典time中数量最大的值，即出现次数最多的为“内科”。则该题干文本标注类型为“内科”。该类型符合MCT出题要求，故医疗题干文本保留。

S15：将HSK中文水平等级的等级词汇“HSK”存储于数组strDict1（https://pan.hqu.edu.cn/share/a1a69fd1e5479dfde939ad2aac）中。

S16：依次遍历listOfStr中的每个词语，在数组strDict1中寻找该词语对应的HSK中文水平等级。获得listOfStr中每个词语的HSK词汇等级。因最高级为HSK5级，因此该文本的语言HSK词汇等级为5级。

S17：依次遍历listOfStr中的每个词语，在数组strDict1中寻找该词语对应的MCT词汇水平等级。获得listOfStr中每个词语的MCT词汇等级。因最高级为二级，因此该文本的语言MCT词汇等级为二级。

S17：设当前出题对应的MCT等级值为二，故该医疗题干文本保留。该医疗题干文本确定为出题语料文本TextQuestion。