S1：电子病历语料筛选

S11：对电子病历文本进行分词，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中的词汇包含于中，则该医疗题干文本舍弃。

S13：如权力要求S12和S13所述，若保留电子病历文本，将HSK中文水平等级的等级词汇“HSK”存储于数组strDict1中。

S14：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。设当前出题对应的MCT等级值为c（c取值为“一”，“二”，“三”），设置MCT等级一级对应HSK1-4级、MCT等级二级对应HSK5级、MCT等级三级对应HSK6-9级进行配置。若该词语对应的HSK等级超过其对应的MCT等级，则该电子病历文本舍弃。

S15：如权力要求S15所述，若保留电子病历文本，则将该电子病历文本确定为对话流生成语料文本EMRText。

S2：电子病历信息抽取

S21：对EMRText文本利用正则表达式，分别提取（“科室：”，“床号”）、（“性别：”，“年龄”）、（“年龄：”，“婚姻”）之间的字符串，并将其存储于数组backgroundWords中。

S22：将“CMeKG词库”中词典的词汇存储于数组strDict2[disease、symptom、expriement、medical]中，并将“电子病历触发词典”中的词汇分别存储于数组strDict3和strDict4中。

S23：对EMRText文本通过split(‘/n’)进行分句，将其存储与sentenceList数组中。

S24：依次sentenceList数组中的“主诉”、“现病史”、“既往史”、“辅助检查”、“初步诊断”进行遍历查找，若查找成功，则将该句进行分词，并依次判断其分词结果是否存在于strDict2，若存在则以键值对的形式（词汇类型-词汇）保存至chiefComplaint、presentIllness、pastHistory、experiment、diagnosis字典当中。(此处键值对如 {‘症状’:‘言语模糊’,‘疾病’:‘脑梗死’})

S25：将“电子病历触发词词典”中词典的词汇依次保存到trigger数组中。查找sentenceList数组中的“婚育史”、“家族史”，对查找到的句子进行分词，并依次判断其分词结果是否存在于strDict2，若存在以键值对的形式保存至marriage、family字典当中。

S26：依次判断chiefComplaint、presentIllness、pastHistory、experiment、diagnosis、marriage、family字典的长度，若数组大于3，则随机丢弃数组中的元素。

S3：对话流生成

S31：分别将presentIllness存储至MCTLevel1数组，将presentIllness、pastHistory、marriage、family、experiment存储至MCTLevel2数组，将presentIllness、experiment、diagnosis（未来还会添加诊疗）存储至MCTLevel3数组。

S32：将“模板问句”中的txt文件分别生成对应的数组并存储。

S33：根据MCT当前出题的分级，选择对应的MCTLevel数组。设当前需要生成MCT二级对话流，则对MCTLevel2进行遍历，依次随机选择MCTLevel2中数组的关键词，对所选关键词所在句利用split(‘;’)进行分句，判断句中是否存在“无”、“否认”等词。

S34：根据所选关键词的来源字典（如presentIllness、pastHistory、marriage、family、experiment）和关键词的类型（key-value中的key），按照规则选择对应的“模板问句”生成的数组。

S35：对关键词所在句使用JioNLP的时间实体抽取与解析工具，判断其句中是否包含时间短语信息，并根据时间信息是否存在选择数组中的模板问句。

S36：将关键词和时间信息与模板问句中的“\_\_\_\_”进行替换。

S4：对话流输出

S41：如权力要求S3所述，将获得一个模板问句长度（大部分是一问一答）的对话。

S42：根据当前需要生成的对话流的MCT分级，对S3生成的对话进行拼接后输出。MCT一级生成并拼接两个对话，MCT二级生成并拼接三个对话，MCT三级生成并拼接五个对话。