1、从文本到GWT

从纯文本文件读入原始的GWT文本，从Feature关键字处切分文本，划分出每个GWT的文本， 使用正则表达式提取出GWT每个字段下的文本，将Feature文本记录到GWT对象的Feature属性，对其他字段文本，首先执行分句操作，获取每一句的原始文本，生成对应的Sentence对象，将每个Sentence对象保存到其归属的GWT属性中。

2、从GWT到TaggedGWT

对一组具有相同Feature值的GWT对象，每一个GWT对象创建一个对应的TaggedGWT对象，将GWT对象的属性值复制到TaggedGWT的对应属性中。

对每一个TaggedGWT对象，首先标记基础标签，包括GWT级的UseCaseName和FlowType的basic或bounded标签，句子级的wordlist、actor和action标签。分析所有TaggedGWT的Given属性中句子的数量，将其中句子最少的TaggedGWT的FlowType标记为basic，句子数量超过标记为basic的GWT 2句或以上（如果有）的TaggedGWT的FlowType标记为bounded。对一组中的每一个TaggedGWT，从TaggedGWT的Feature字段读出用例名记录到UseCaseName属性中，对When属性中的每一个Sentence对象，首先执行分词操作，将分词结果保存到句子的wordlist属性，使用句子的分词结果对句子做依存句法分析，根据依存句法分析的结果搜索句法树，将依存关系为主谓关系的词记录到Sentence对象的actor属性，将依存关系的核心词记录到Sentence对象的action属性。

完成基础标签标记后，进一步标记GWT级FlowType中的specific和global，refer标签，condition标签；句子级的normalContent，type和associated标签。

对TaggedGWT的When中记录的句子，根据其句法分析结果判断是否为简单句，若为简单句，将其type标记为normal，否则分析其wordlist中的关键词，若含有表示条件的关键词则标记type为conditional，若含有表示循环的关键词则标记type为circular，在此基础上将这些句子的文本规范化，对normal类型的句子去掉表示具体数据的宾语的定语，生成规范化句子内容normalContent，对类型为conditional和circular的句子将其关键词替换为RUCM关键字，生成规范化的句子内容normalContent。

对TaggedGWT的Then中记录的句子，分别进行分词和依存句法分析，填充标签wordlist，对句子的表述做规范化，填充标签normalContent.

对FlowType标记为basic的TaggeGWT，将其Given中句子的type标签标记为common，对FlowType标记为bounded的TaggedGWT，将其Given中的句子与basic TaggedGWT的Given中的句子做相似性比较，找到其中相近的句子标记type为common，其他句子的type标记为unique.将标记为unique的句子与basic TaggedGWT中When记录的句子相似性进行比较，对其中相似的句子，将unique句子的associated标记为该When中句子的索引。

对未标记FlowType的TaggedGWT，将其Given中的句子与basic TaggedGWT的Given属性记录的句子进行相似性比较，与type为common的句子相似的句子type标记为common，其他的句子type标记为unique，将unique的句子添加到TaggeGWT的condition属性中。在此基础上，找出basic TaggedGWT的When属性中与unique句子相似度最高的句子，若两个句子是相近的，则将此TaggedGWT的FlowType标记为specific、refer标记为相近句子的序号，并将句子的associated标签标记为预期相近的句子的序号；若两个句子不相近，则将此TaggedGWT的FlowType标记为global，将unique的句子记录到此TaggedGWT的condition属性。

3、从TaggedGWT到RUCM

根据UseCaseName相同的一组TaggedGWT生成一个RUCM对象，依次为

Use Case Name：使用TaggedGWT的UseCaseName

Brief Description：遍历一组TaggedGWT，将所有TaggedGWT的Scenario句子中的originContent组合成一个文本，使用TextRank算法对文本进行摘要生成，将生成的摘要文本作为RUCM的Brief Description.

Primary Actor , Secondary Actors:对所有TaggedGWT的When属性中句子的actor，统计所有actor的数量，去掉其中表示系统本身的actor，选择剩下的actor中数量最多的作为RUCM的Primary Actor，其余的actor组成RUCM的Secondary Actors.

Precondition: 使用FlowType标记为basic的TaggedGWT，将此TaggedGWT的Given属性中句子的originContent属性值连接成RUCM的Precondition.

Basic Flow：使用FlowType标记为basic的TaggedGWT，将此TaggedGWT的When属性中句子的normalContent依次作为Basic Flow中的action，将此TaggedGWT的Then属性中句子的normalContent连接成Basic Flow的postcondition。

Specific Alternative Flow：使用FlowType标记为specific的TaggedGWT（如果有），将此TaggedGWT的Refer属性值作为Specific Alternative Flow的RFS，将此TaggedGWT的When属性中句子的normalContent依次作为Specific Alternative Flow中的action，将此TaggedGWT的Then属性中句子的normalContent连接成Specific Alternative Flow的postcondition。

Bounded Alternative Flow：使用FlowType标记为bounded的TaggedGWT（如果有），将此TaggedGWT的Refer属性值作为Bounded Alternative Flow的RFS，将此TaggedGWT的When属性中句子的normalContent依次作为Bounded Alternative Flow中的action，将此TaggedGWT的Then属性中句子的normalContent连接成Bounded Alternative Flow的postcondition。

Global Alternative Flow：使用FlowType标记为global的TaggedGWT（如果有），将此TaggedGWT的condition属性值作为Global Alternative Flow的发生条件，将此TaggedGWT的When属性中句子的normalContent依次作为Global Alternative Flow中的action，将此TaggedGWT的Then属性中句子的normalContent连接成Global Alternative Flow的postcondition。