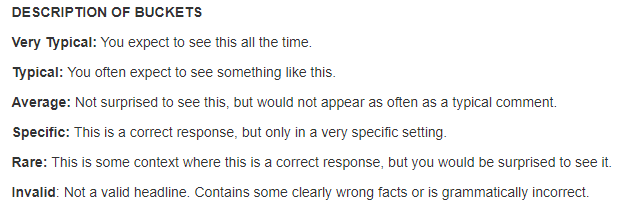
**智能写作模型质量评估**

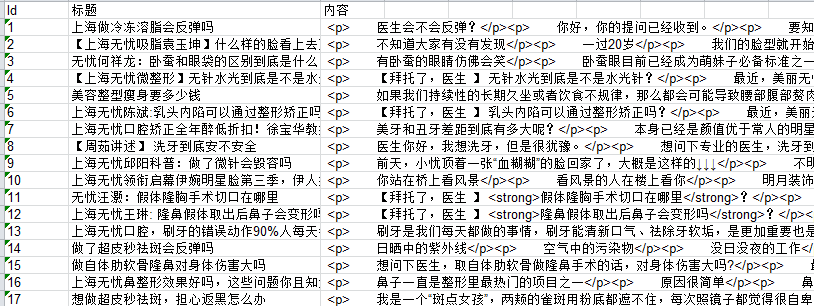
参考HUSE指标，结合人工和统计方法，设计一个百分制量化指标评判模型好坏。

1. 可读性：人工评判可读性，并转化为分数
2. 随机抽样句子
3. 人工打分标准



1. 切题度：对比该生成主题的基准语料，用统计手段得到切题度的分数
2. 主题基准语料的准备：

* 按语料标题作为主题，内容作为该主题语料，将现有四千多主题归至100主题以内



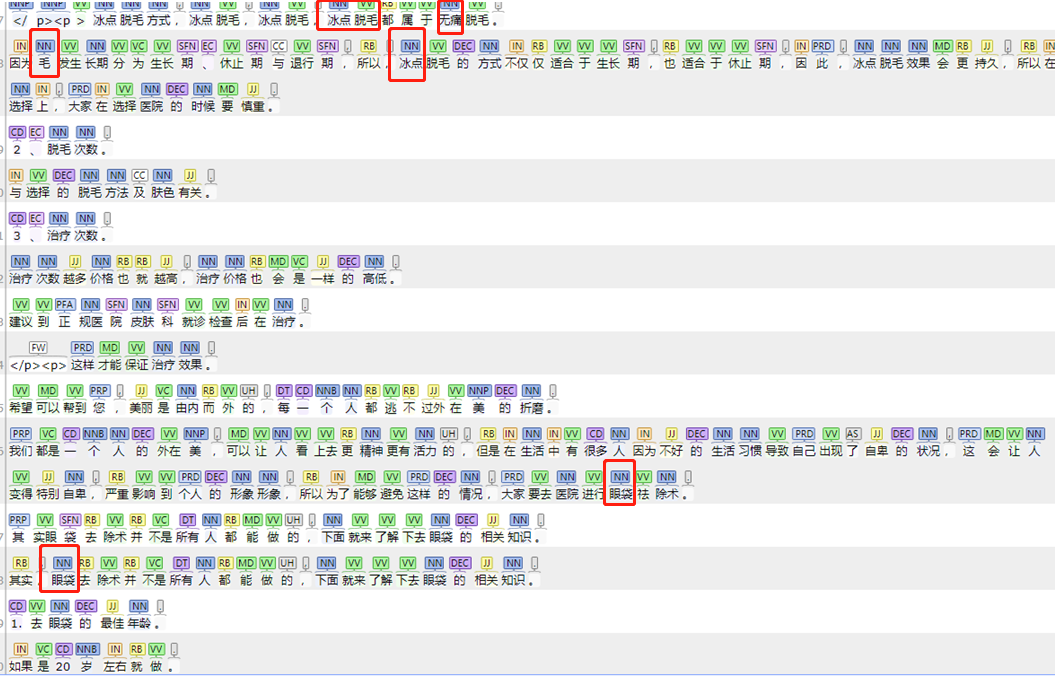
* 对每个主题的文章进行分词，得到含实义的词汇（动词、名词、形容词为主），如：自体、填充、丰满、玻尿酸等等



* 将得到的词汇作为特定主题的基准参考语料

1. 模型生成文章与基准语料对比：

* ­对生成文章判断类别，并进行分词，得到实义词汇表
* 检测生成文章实义词汇属于各类主题基准语料的比例，属于自身类别比例越高，切题度越高



1. 重复度：分为文章内重复度，同模型同主题生成文章间重复度，以及生成文章与语料文章重复度
2. 文章内重复度，N-gram去重处理后文章所剩字数与要求字数之比例
3. 同模型生成文章间N-gram, 最长公共序列LCS，越长扣分越多
4. 模型文章和语料间N-gram, LCS
5. 结合人工和统计的分数，调试各自权重参数，得到通用的评分标准