数据预处理代码编写能力测试（mapreducer处理）

数据预处理是所有智能化应用必须做的前期准备工作，在语音识别训练中，我们需要将语音标注结果转成适合训练的格式。

请参考sample文件夹下的样例，根据如下数据预处理要求，通过编写数据预处理代码，将fake\_mark.txt中的内容输出到result.txt文件，编程语言不限。

1. 将fake\_mark.txt内容根据样例转换为指定格式，以utf-8编码保存。
2. 一行一个句子，句子长度大于等于三个字，小于50个字。若小于3个字及超过50个字需删去该句。
3. 去除所有标点，仅保留汉字和英文。若包含其他语言，需删去该句。
4. 阿拉伯数字需将其转换为中文，例：85层转为八十五层。