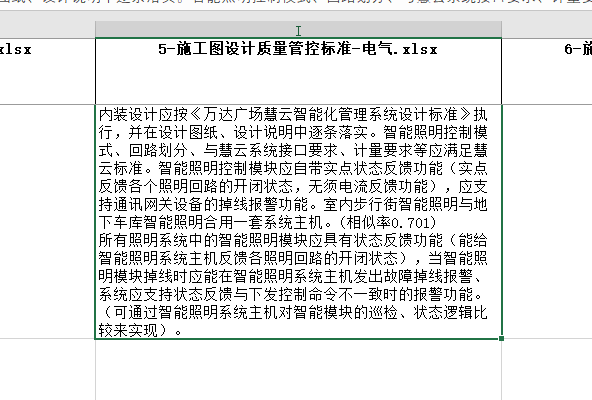
1. **给排水和幕墙，暖通和幕墙都有相似率很高的内容；但同样的内容，却没有出现在是给排水和暖通的交叉单元格里，这是为什么？**

**结果分析：给排水和暖通两个文件依然存在相似率很高的内容并出现在交叉单元格中，**

**查找方式为：行暖通-列给排水**

**（2）电气和内装，有一段话，相似度很高，你看下，我感觉好像他们俩谁说的不算是一回事呢，为什么会有这么高的相似度。**



结果分析：

因为两段话分词后的结果中相同词的出现次数较多

文本1去除停用词后的分词结果是：

['装', '设计', '应', '《', '万达', '广场', '慧云', '智能化', '管理系统', '设计标准', '执行', '并', '设计', '图纸', '设计 说明', '中', '逐条', '落实', '智能', '照明', '控制', '模式', '回路', '划分', '慧云', '系统', '接口', '、', '计量', '等', '应', '慧云', '标准', '智能', '照明', '控制', '模块', '应', '自带', '实点', '状态', '反馈', '功能', '实点', '反馈', '照明', '回路', '开闭', '状态', '无须', '电流', '反馈', '功能', '，', '应', '支持', '通讯网', '关', '设备', '掉线', '报警', ' 功能', '室内', '步行街', '智能', '照明', '与', '地下', '车库', '智能', '照明', '合用', '一套', '系统', '主机']

文本2去除停用词后的分词结果是：

['照明', '系统', '中', '智能', '照明', '模块', '应', '状态', '反馈', '功能', '智能', '照明', '系统', '主机', '反馈', '照 明', '回路', '开闭', '状态', '，', '智能', '照明', '模块', '掉线', '时应', '能', '在', '智能', '照明', '系统', '主机', ' 发出', '故障', '掉线', '报警', '系统', '应', '支持', '状态', '反馈', '下发', '控制', '命令', '一致', '时', '报警', '功能', '（', '通过', '智能', '照明', '系统', '主机', '智能', '模块', '巡检', '状态', '逻辑', '来', '）']

文本1和文本2共同的分词结果是：

{'控制', '系统', '照明', '回路', '状态', '应', '模块', '中', '报警', '反馈', '功能', '智能', '支持', '掉线', '主机', '开 闭', '，'}

文本1的词向量是：

[1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 3, 3, 0, 1, 0, 2, 1, 0, 1, 2, 1, 0, 1, 1, 4, 1, 3, 1, 1, 1, 1, 2, 1, 0, 1, 0, 0, 4, 1, 0, 1, 1, 1, 1, 5, 2, 1, 1, 1, 2, 0, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1]

文本1的词向量是：

[0, 0, 0, 1, 0, 1, 0, 1, 3, 0, 0, 2, 3, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 2, 1, 0, 0, 2, 2, 0, 1, 1, 1, 0, 1, 1, 6, 0, 1, 0, 3, 0, 0, 7, 4, 0, 0, 0, 5, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1]

两个文本的相似度是：0.7013