病历数据提取说明

本文档说明的病历数据提取方式，是通过定时任务来将xml形式的病历数据按照提取字典解析，并最终将解析结果以二维表的形式保存在第三方中间表iemr\_mr\_para（通常是在cis库）。

# 准备工作

1、病历数据库（oracle）所需要的字典表，请保证md\_mr\_para 和md\_mr\_tp\_para表字段和下面文件(提取病历所需表结构.sql)一致。

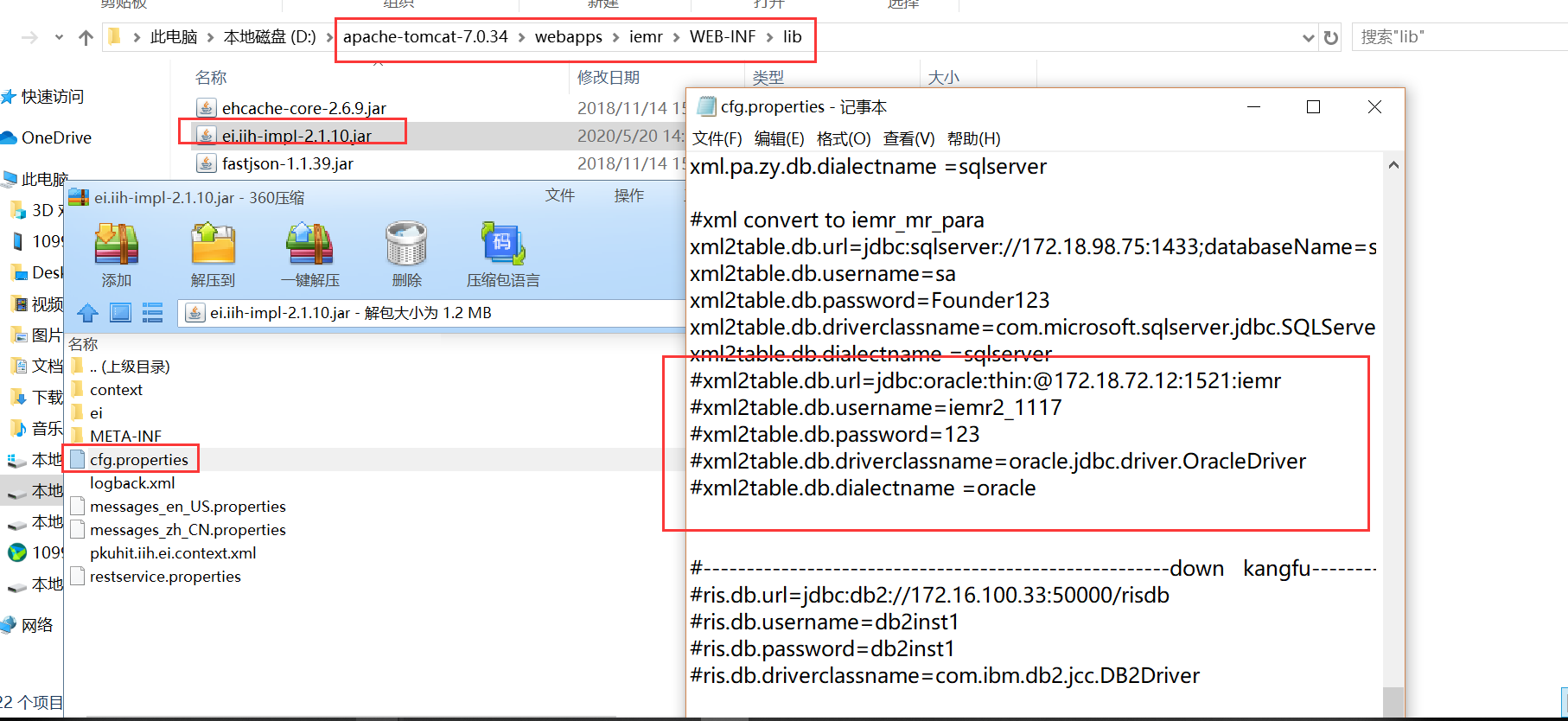
（双击打开）

2、确保cis中间表iemr\_mr\_para(文书的提取结果二维表)和下边文件一致。

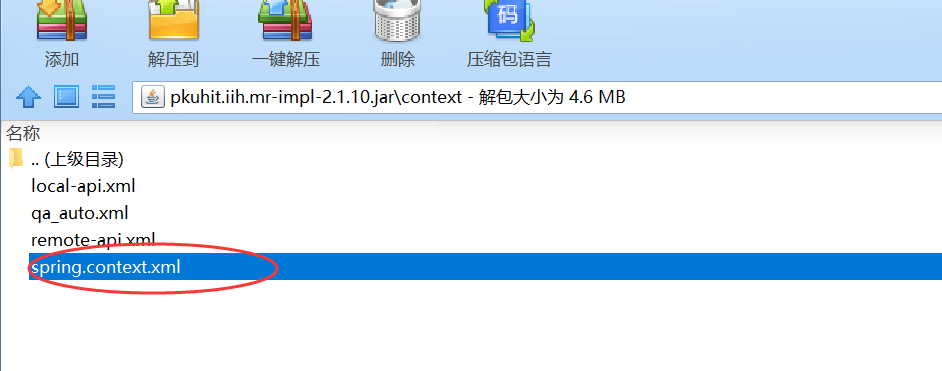
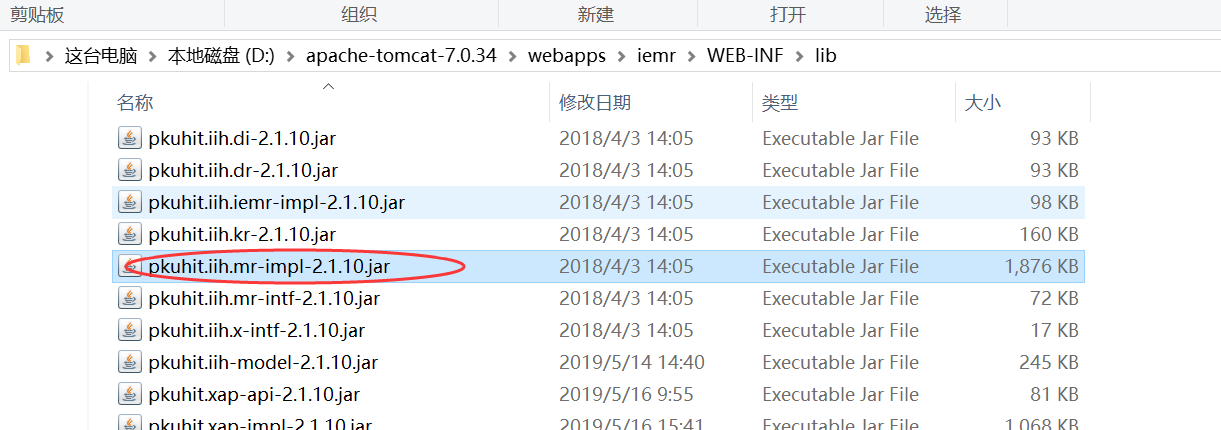


# 服务器数据提取定时任务配置

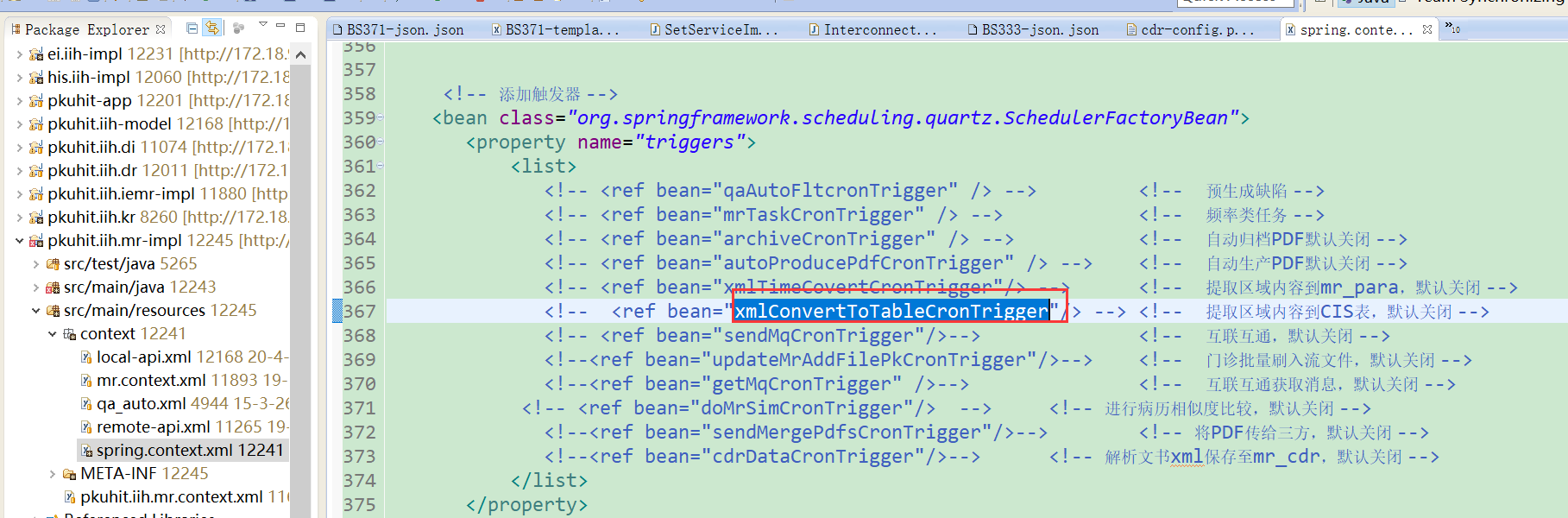
## 提取结果的数据库ip配置



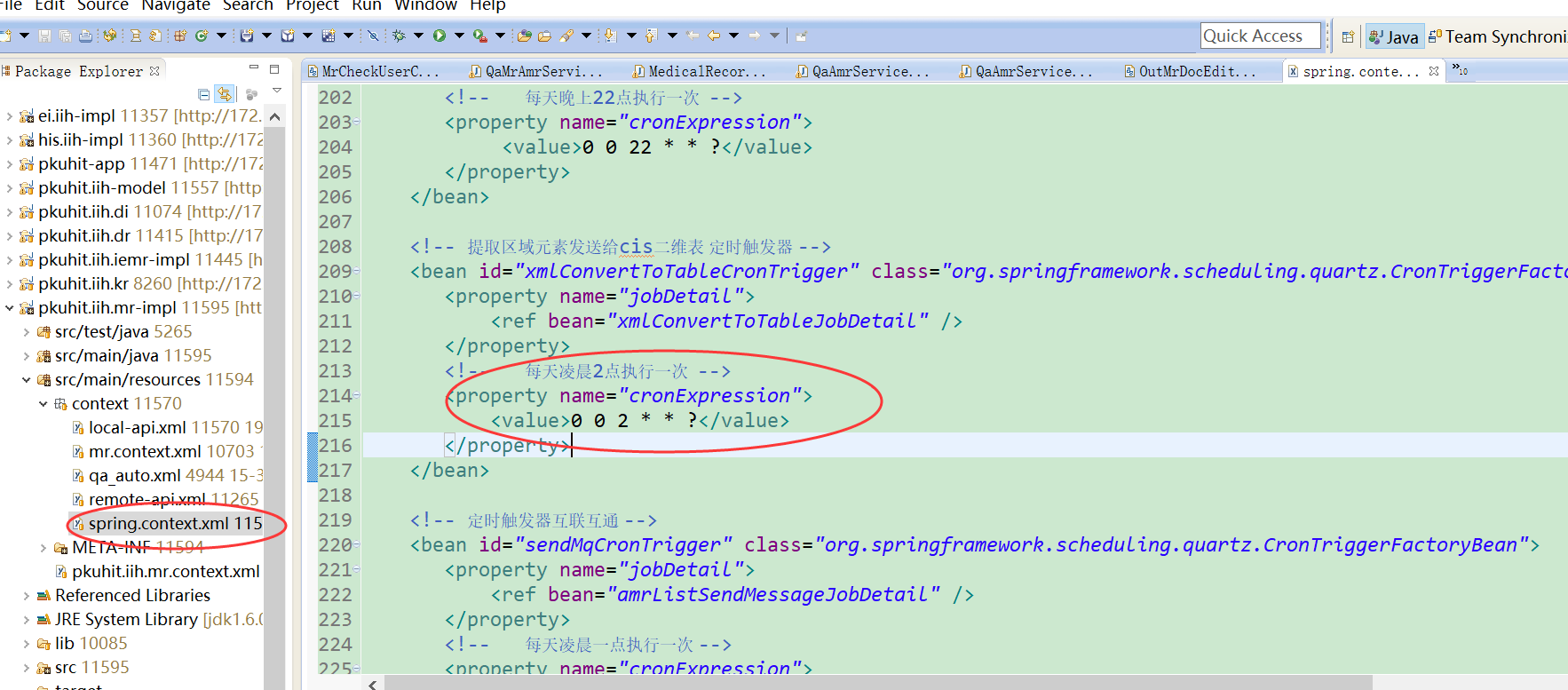
## 定时任务开启、频率等配置



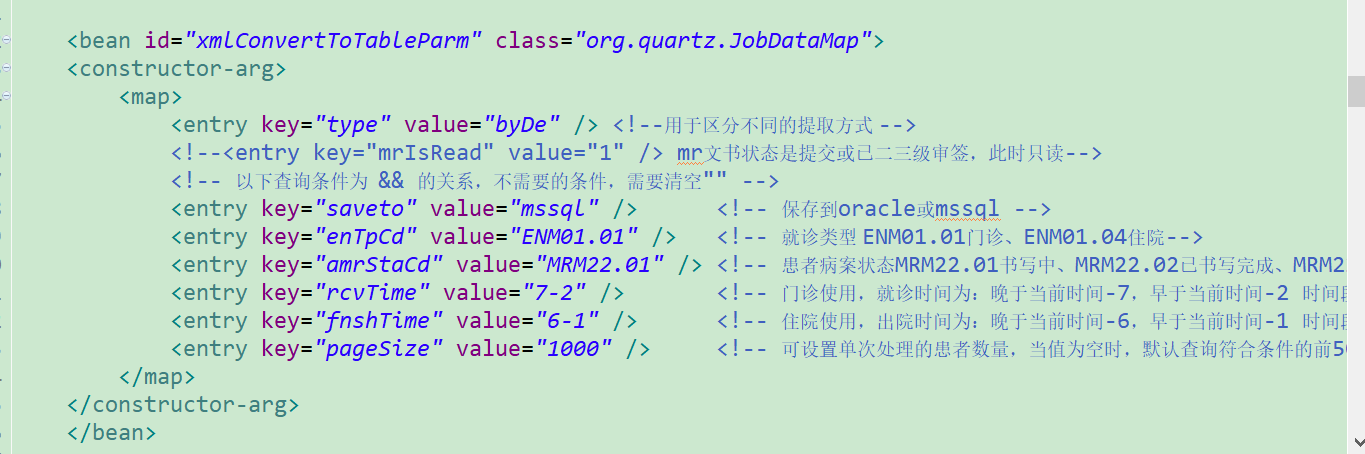
### 定时任务触发器



### 触发器频率



### 触发器参数配置



# 配置提取字典

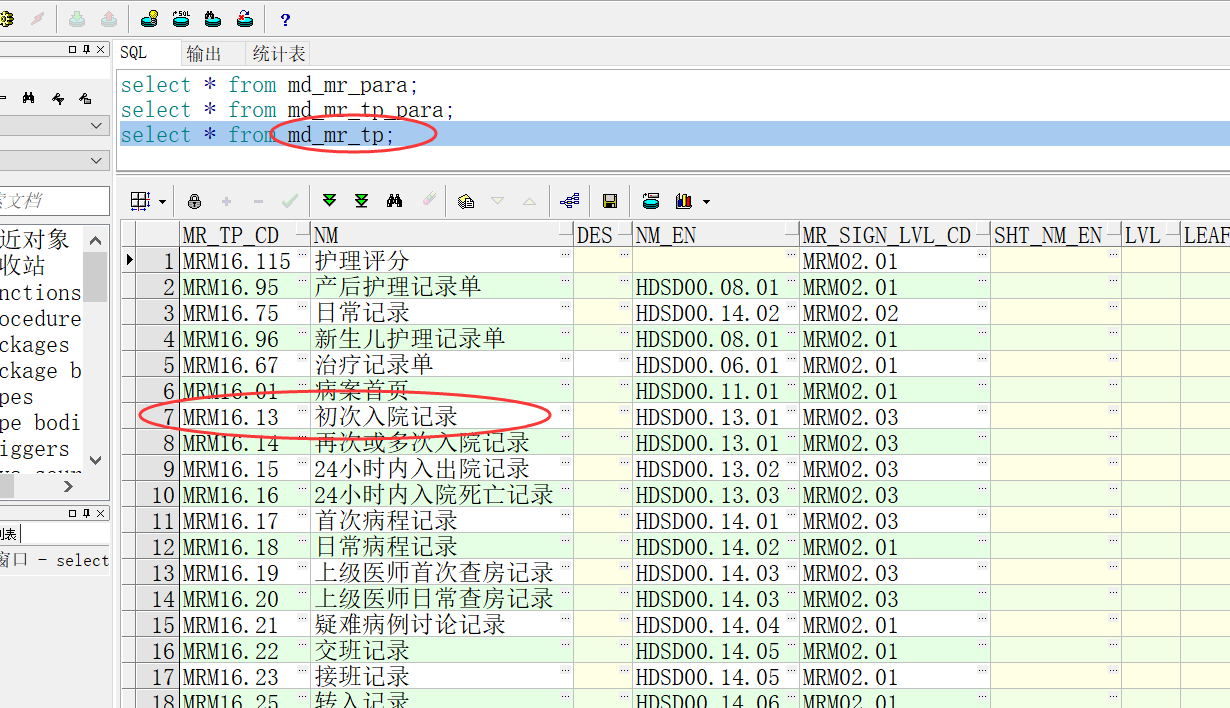
在进行提取字典配置之前需选择是否开启byDE模式（2.2的触发器参数配置-最后一张截图中的type参数是否开启），**推荐开启**。

## 查看文书类型编码

表md\_mr\_tp是文书类型表，在该表中根据nm字段检索要进行病历提取的文书类型的记录，获取主键mr\_tp\_cd字段值，下一步提取字典配置需要用，例：MRM16.13代表出错入院记录。

同时，如果**开启byDe模式**，要使该类型病历被提取，表md\_mr\_tp记录的xml\_cvet\_f配置为1。如果只是部分模板对应的病历需要提取， md\_mr\_tp表的xml\_cvet\_f字段不需要配置，只需要配置mr\_tpl表（模板表）的xml\_cvet\_f字段为1。

总之，在byDe模式下，表md\_mr\_tp和表mr\_tpl只要有一个xml\_cvet\_f配置为1，用相应模板写的文书将被提取。

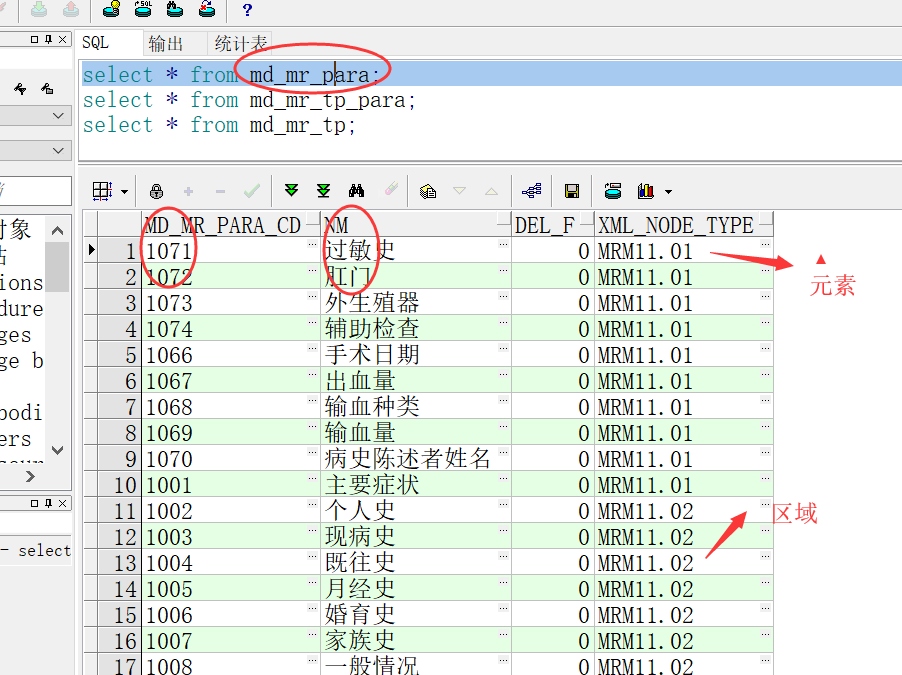


## 配置该文书类型的病历需要提取的内容

### 提取内容为元素或者区域

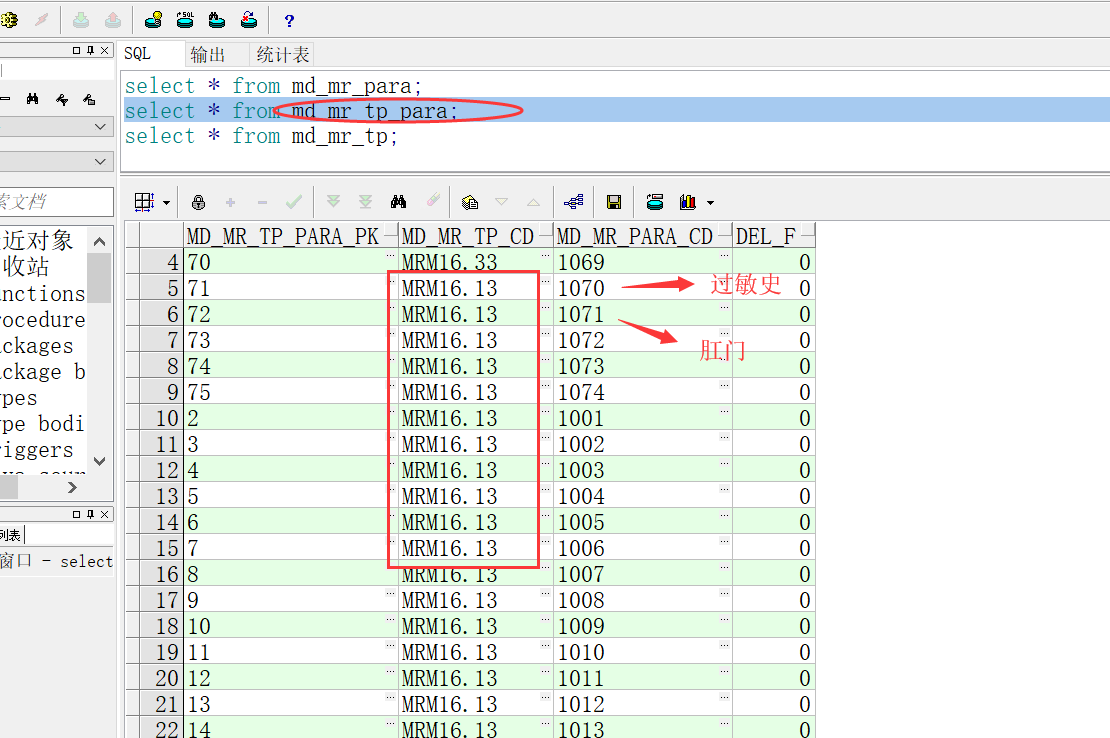
比如：入院记录需要提取过敏史

1、表md\_mr\_para要有能包含过敏史的一项，**不是byDe模式**需要配置文书中该内容是元素还是区域（xml\_node\_type），记住该过敏史记录的主键，下一步需要。



2、文书类型要提取得元素的字典表：

md\_mr\_tp\_para表配置MRM16.13这种类型需要提取的内容。

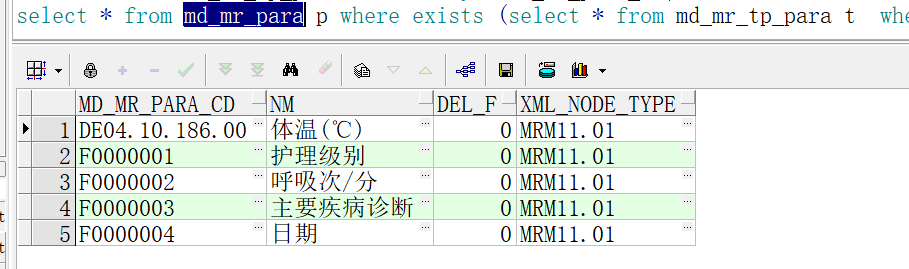


### 提取内容为表格的某列

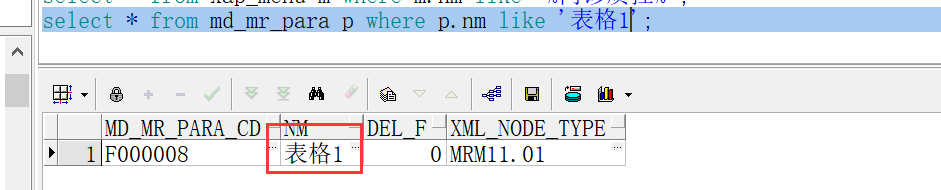
例：护理记录要提取体温、护理级别等（这些列都在一个叫表格1的表格中）

1、元素字典表md\_mr\_para：

**不是byDe模式：**需要配置提取列

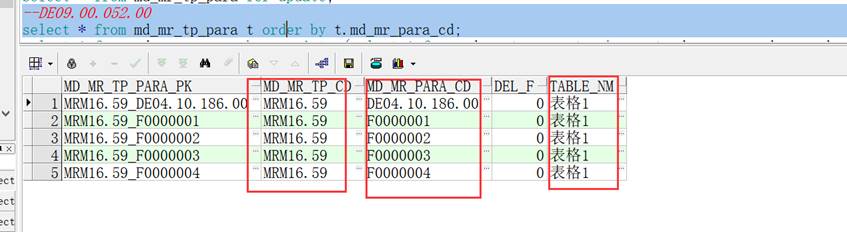


**是byDe模式：**只需要配置列所属表格名称即可

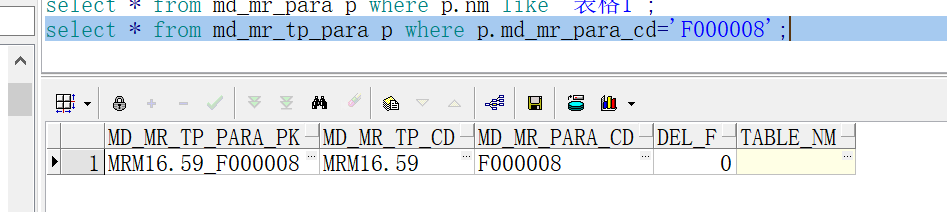


1. 文书类型要提取的元素的字典表（MRM16.59是护理文书所属文书类型）

**不是byDe模式：**需要配置该列在病历中属于哪个表格的表格名称。



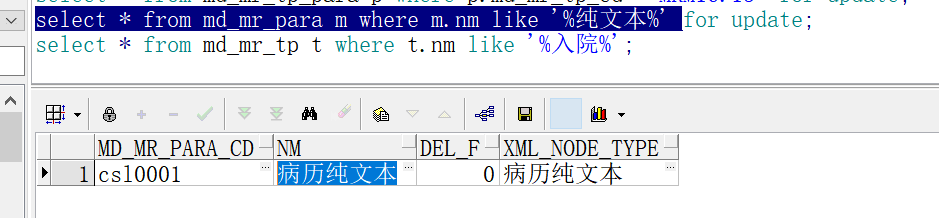
**是byDe模式**：只需要配置该类型文书要提取表格1就行。



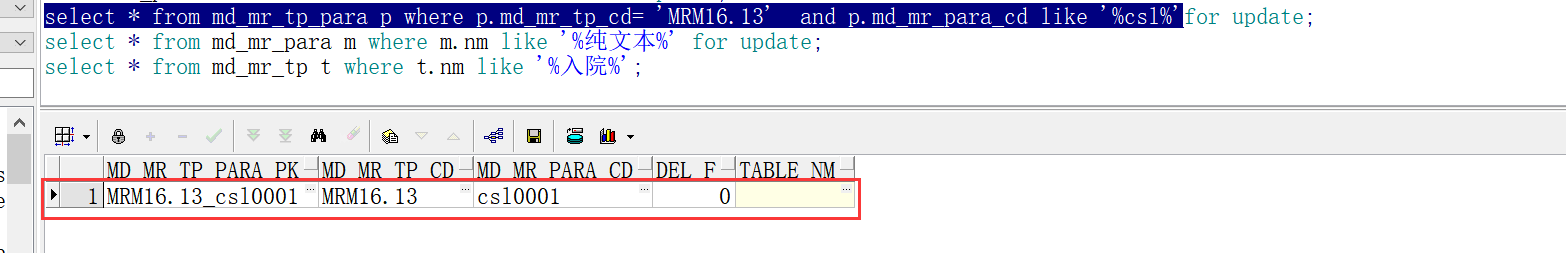
### 提取整篇文书纯文本

，**不是byDe模式可行。**

**1、**表md\_mr\_para添加“病历纯文本”记录



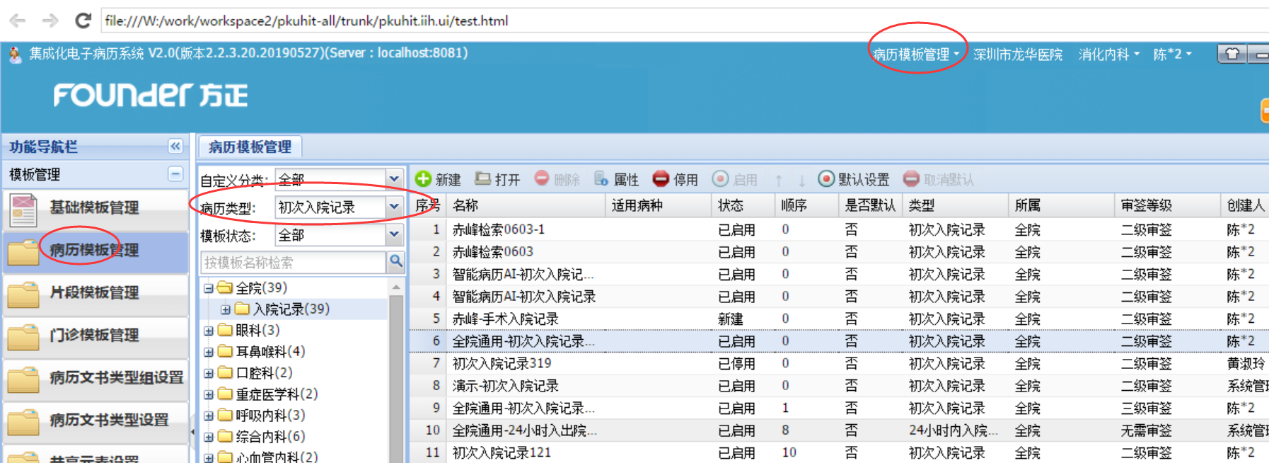
2、表md\_mr\_tp\_para表配置MRM16.13这种类型需要提取的内容。



### 提取内容具体是元素还是区域、表格列名称

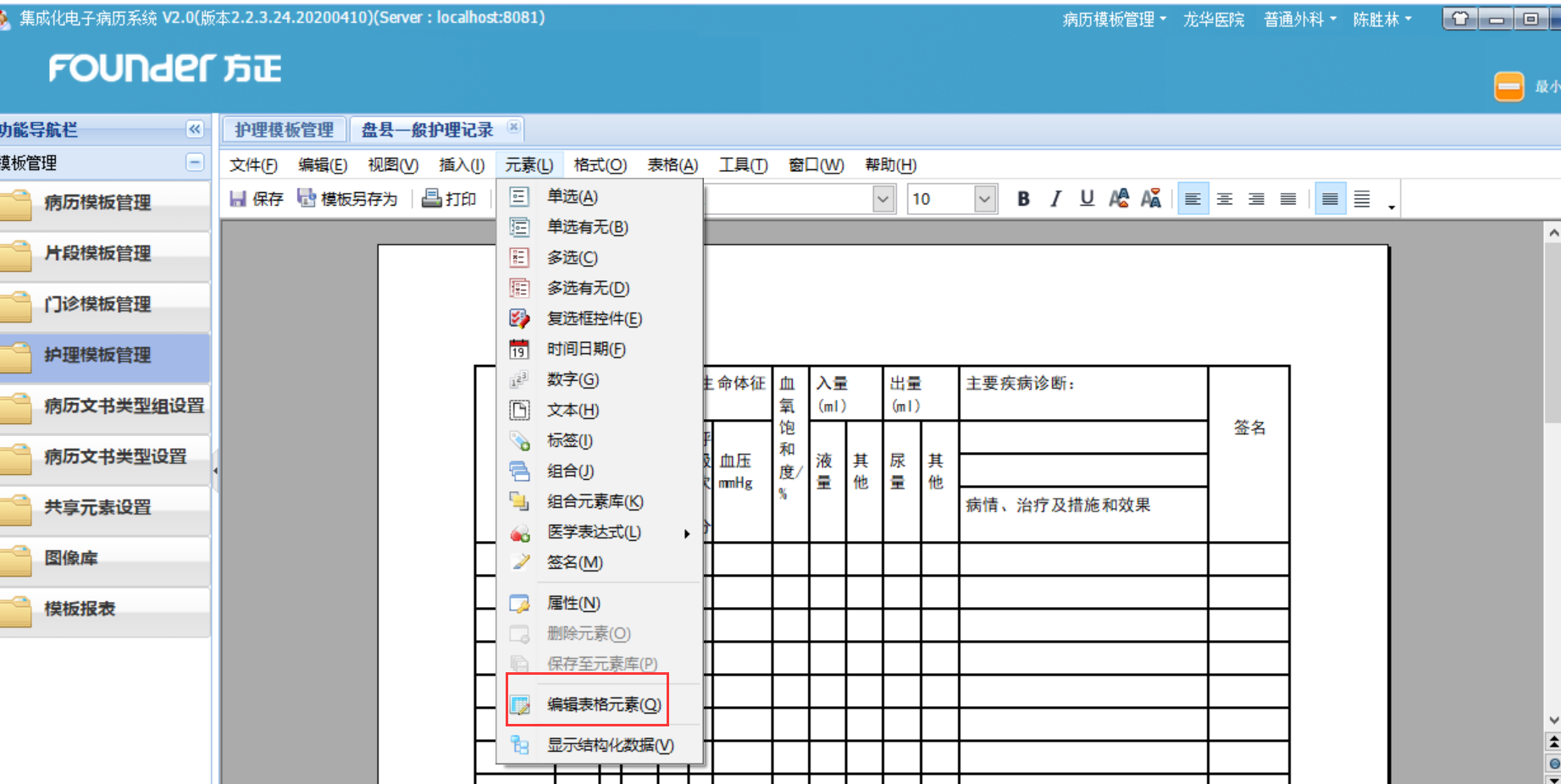
#### 确定元素或区域

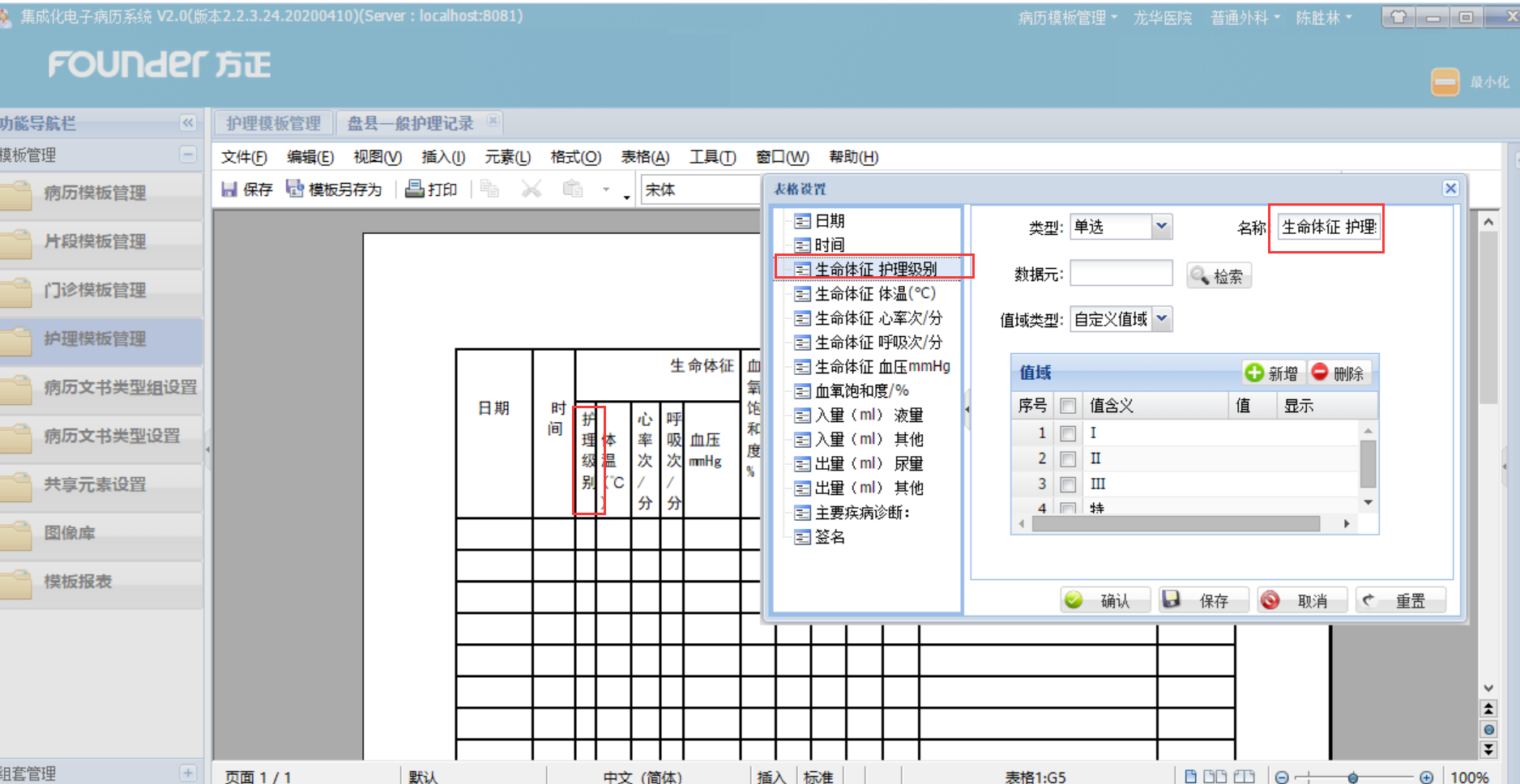
注意：每个文书类型要提取的内容是元素还是区域需要根据实际模板进行区分，方法如下截图：





#### 护理表格的列名称





# cis库病历提取结果表

关于文书提取结果表iemr\_mr\_para（在cis库），下图是结果示例：

|  |  |  |  |
| --- | --- | --- | --- |
| **字段** | **名称** | **类型** | **备注** |
| PARA\_PK | 主键 | int |  |
| PA\_ID | 患者id | varchar |  |
| EN\_PK | 就诊号 | varchar |  |
| MR\_PK | 文书主键 | varchar | 对应病历数据库的MR表主键 |
| MR\_TPL\_CD | 模板主键 | varchar | 对应病历数据库的MR\_TPL表主键 |
| MR\_TP\_CD | 文书类型编码 | varchar | 对应病历数据库的MD\_MR\_TP表主键 |
| PARA\_CD | 提取数据编码 | varchar |  |
| PARA\_NM | 提取数据名称 | varchar |  |
| PARA\_CONTENT | 提取数据内容 | varchar |  |
| MEMO | 备注 | varchar |  |
| SORT\_NO | 序号 | varchar | 对于护理记录，序号一样表示是同一行记录 |
| TABLE\_NM | 表名 | varchar | 提取的护理表格的名称 |
| DEL\_F | 删除标记 | varchar | 1删除 |
| ELEMENT\_VALUE | txt格式病历纯文本 | text |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

sort\_no相同表示是同一行的记录，mr\_pk是文书主键，MR\_TP\_CD是文书类型编码（对应md\_mr\_tp主键），PARA\_CD元素编码，PARA\_NM元素名称，PARA\_CONTENT元素内容。另外，结果中PARA\_CD等于MrCrtUserId的内容是起草医生工号，PARA\_CD等于MrCrtTime的内容是起草时间，PARA\_CD等于MrLastUpdTime的内容是病历最后更新时间

