**SQL查询语句自动生成的需求分析与设计**

# 总体需求

为优化HROSaaS的自定义报表功能，借鉴Business Objects软件的Universe概念，让用户选择需要用到的字段，设定需要的过滤条件，设置排序标准，自动生成所需的查询SQL，作为自定义报表的数据源。

该软件采用Java开发，为服务端的一个服务，提供相应的Restful Web Service接口。

输入：

* + Universe
  + 用户查询意图（显示的对象集，过滤条件，排序标准）

输出：SQL查询语句

# Universe的设计

Universe 的作用是为非技术用户提供一个易于使用和理解的接口，使他们能用Universe来表达数据库查询的意图。

## Universe的组成

Universe包含：

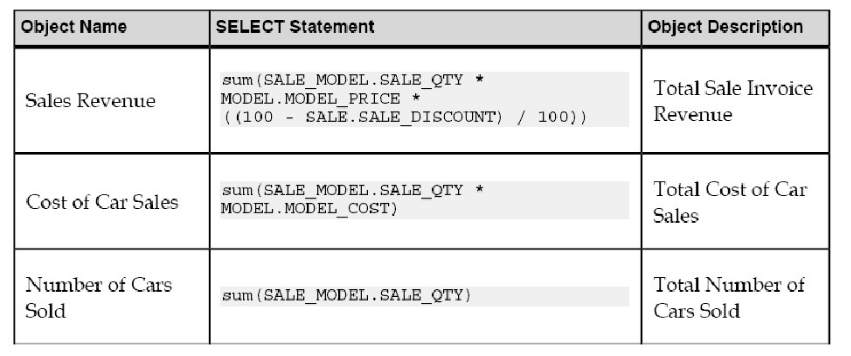
1. 数据库连接

包括连接数据源时使用的登陆凭据信息、数据库名称与位置等（这里只考虑本地的连接并只访问单一数据源）。

1. 数据库模式（Schema），又称为data foundation，

即在数据库中使用的表和联接的模式。

1. 业务层对象（Object）

* 属性对象 – 属性对象提供了有关特定维度的附加信息，指向表的某一字段。例如客户(customer)维度中，年龄、传真、电话号码、街道地址和备注等信息。
* 度量对象–传递用于确定对象数量的数值信息，是用户想要分析的数字。例如“Sales Revenue （销售收入）”度量总是以某种形式聚合，如求和、计数、平均值、最小值或最大值，需要定义度量的公式，公式中包含函数，需要计算的字段和基本计数方法。例如

1. 预过滤条件

专家用户可以提前设定好一些经常使用的过滤条件(包含对象，操作符（例如大于）和操作数)。普通用户可直接从中选取自己所需的过滤条件。

1. 文件夹

文件夹（也称类或类别），是用来组织度量对象、属性对象和预过滤条件的一种方式，即度量对象和属性对象按文件夹分组。

1. 查询语句

专家用户可以预定义一些常使用的查询语句，普通用户可直接从中选取作为过滤条件中的操作数。

## Universe的格式定义

我们用MySQL数据库来存储Universe，在此应定义Universe的表格格式。

### 数据逻辑视图



### 数据物理视图



# 用户查询意图

## 意图的组成

用户查询意图由三方面组成：显示的对象集，过滤条件，排序标准。

1. **选择对象集**

从Universe中选择查询要显示的一个或多个对象，按显示次序排。

1. **设置过滤条件**

在结果对象集中，选中某个对象，选择操作符（例如大于），然后选择操作数（例如类型的值或其它对象）

可以设置多个过滤条件，条件是“并且”关系。

表3-1 操作符

|  |  |  |
| --- | --- | --- |
| **得到数据** | **例如** | **操作符** |
| 和给定的一个值相同 | 某年 | 等于 |
| 和给定的值不同 | 国家里不包括法国 | 不同 |
| 比给定的值大 | 超过100000美元的收入 | 大于 |
| 大于或等于给定值 | 大于或等于60岁的客户 | 大于或等于 |
| 小于给定值 | 收入小于100000美元 | 小于 |
| 小于或等于给定值 | 60或60岁以下的客户 | 小于或等于 |
| 两个给定值之间 | 25周到36周之间 | 之间 |
| 两个给定值之外 | 不包括25周到36周的一年 | 不包括 |
| 和列表里的某些值相同  列表值的最大数目是256 | 两个地方的收入 | 在列表里 |
| 与给定值不同 | 非欧洲用户 | 不在表里 |
| 包含空行 | 没支付过的客户（例如，没有发票日期） | 空 |
| 不包括空行 | 支付过的客户（发票日期在数据库里） | 非空 |
| 全部包含相同的字母或词 | 以S开头的客户名 | 匹配模式 |
| 不包含相同的字母或词 | 不以S开头的客户名 | 不匹配模式 |
| 一个对象满足两个条件 | 在六月或七月确定定单的客户 | 满足两个 |
| 不包括没在法国居住的客户值 | 里维埃拉 | 不包括 |

表3-2 操作数

|  |  |
| --- | --- |
| **操作数** | **说明** |
| 类型的值 | 可以是1个或多个值，用逗句间隔。例如：年份=1994，1995 |
| 从对象值列表里选择的值 |  |
| 由其他查询返回的值 | 创建子查询（ANY） |
| 由其他查询返回的所有值。 | 创建一个子查询（ALL） |
| 计算的结果（总和，最小值，最大值，平均值或数） |  |

表3-3通配符

|  |  |
| --- | --- |
| **通配符** | **描述** |
| % | 代替一些符号，或对提示的反映。例如，N%返回所有以N开头的值（纽约，内华达等等） |
| \_ | 代替常量里的一个字符。例如，GR\_VE返回Grave, Grove, Greve,等等。 |

1. **设置排序标准**

在结果对象集中，选中某个对象，设置排序标准。

表3-3排序顺序

|  |  |  |  |
| --- | --- | --- | --- |
|  | **正文** | **数字** | **时间** |
| **上升顺序** | A-Z | 从低到高 | 从过去到现在 |
| **下降顺序** | Z-A | 从高到低 | 从现在到过去 |

可以设置多个排序标准，按主次排列。

1. **设置选项**

* 删除重复记录
* 定义要求查询返回的记录条数

我们不需要时间层次和值列表，也不需要组合查询（UNION，INTERSECT，MINUS）。

## 意图的格式定义

1. 接口名称：setObjects

接口说明：设置需要查询的对象集合

请求参数：

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **必选** | **类型和范围** | **说明** |
| objects | True | List<int> | 需要查看的对象ID集合 |

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| errorInfo | False | String | 错误信息 |

1. 接口名称：setFilter

接口说明：设置单个过滤条件，在结果对象集中，选中某个对象，选择操作符（例如大于），然后选择操作数（例如类型的值或其它对象）。

当需要设置多个过滤条件时，则继续选中该对象，选择操作符和操作数，构成新的过滤条件，多个条件间是“AND”关系。

请求参数：

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **必选** | **类型和范围** | **说明** |
| object | True | int | 过滤条件涉及的对象ID |
| operator | True | int | 过滤条件操作符 |
| operandType | True | int | 过滤条件操作数类型 |
| operand | True | String | 过滤条件操作数（数值，字符串，子查询等等） |

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| errorInfo | False | String | 错误信息 |

1. 接口名称：setPredefinedFilter

接口说明：设置创建好的单个预定义过滤条件

当需要设置多个预定义过滤条件时，则多次选中预定义的过滤条件，这些条件间是“AND”关系。

请求参数：

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **必选** | **类型和范围** | **说明** |
| filter | True | Int | 需要调用的预定义过滤条件ID |

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| errorInfo | False | String | 错误信息 |

1. 接口名称：setOrder

接口说明：设置单个排序标准，选择需排序的对象和排序标准。

当需要设置多个排序标准时，则继续选择排序的对象和排序标准，多个排序标准之间依据设置的先后进行排序。

请求参数：

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **必选** | **类型和范围** | **说明** |
| object | True | int | 需要排序的对象ID |
| order | True | int | 排序标准 |

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| errorInfo | False | String | 错误信息 |

1. 接口名称：setDistinctRecords

接口说明：设置返回结果中是否要去掉重复记录。

请求参数：

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **必选** | **类型和范围** | **说明** |
| distinct | True | Bool | 是否去掉返回结果中的重复记录（默认为False） |

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| errorInfo | False | String | 错误信息 |

1. 接口名称：setReturnNumber

接口说明：设置返回的记录数。

请求参数：

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **必选** | **类型和范围** | **说明** |
| returnNumber | True | int | 设置返回的记录数量（若returnNumber小于0，则返回所有记录） |

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| errorInfo | False | String | 错误信息 |

1. 接口名称：generateSQL

接口说明：根据之前设置的各项条件，生成SQL查询语句。

请求参数：无

返回格式：json

返回字段：

|  |  |  |  |
| --- | --- | --- | --- |
| **返回字段** | **必选** | **类型和范围** | **说明** |
| status | True | int | 状态码 |
| SQL | False | String | 生成的SQL语句 |
| errorInfo | False | String | 错误信息 |

排序规则：

|  |  |
| --- | --- |
| **顺序名称** | **对应值** |
| 上升顺序 | 1 |
| 下降顺序 | 2 |

过滤条件操作符：

|  |  |
| --- | --- |
| **操作符** | **对应值** |
| 等于 | 1 |
| 不同 | 2 |
| 大于 | 3 |
| 大于或等于 | 4 |
| 小于 | 5 |
| 小于或等于 | 6 |
| 之间 | 7 |
| 不包括 | 8 |
| 在列表里 | 9 |
| 不在表里 | 10 |
| 空 | 11 |
| 非空 | 12 |
| 匹配模式 | 13 |
| 不匹配模式 | 14 |
| 满足两个 | 15 |
| 不包括 | 16 |

操作数类型：

|  |  |
| --- | --- |
| **操作数** | **对应值** |
| 类型的值 | 1 |
| 从对象值列表里选择的值 | 2 |
| 由其他查询返回的值 | 3 |
| 由其他查询返回的所有值。 | 4 |
| 计算的结果（总和，最小值，最大值，平均值或数量） | 5 |

# SQL语句生成算法设计

# 程序文件清单