

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»
(Университет ИТМО)

Факультет программной инженерии и компьютерной техники

Отчет по лабораторной работе №3

по дисциплине « База данных »

Тема: Data Mining на основе модели CRIPS-DM

Выполнила:

студентка гр.№Р33212 Ян Цзяфэн

Санкт-Петербург

2020

Оглавление

Задание	2
Описание предметной области	2
Список сущностей и их классификацию	3
Инфологическая модель	4
Даталогическая модель	5
Вывод.....	5

Задание

Для выполнения лабораторной работы №3 необходимо:

- Сформировать ER-модель и нарисовать ER-диаграмму предметной области, которая была описана в рамках лабораторной работы №1. ER-модель должна соответствовать описанию, представленному в лабораторной работе №1.
- На основе ER-модели построить даталогическую модель.

Отчёт по лабораторной работе должен содержать:

- ✓ титульный лист;
- ✓ текст задания;
- ✓ описание предметной области;
- ✓ инфологическая модель;
- ✓ даталогическая модель;
- ✓ выводы по работе;

Темы для подготовки к защите лабораторной работы:

- Инфологическая модель.
- Построение ER-диаграммы.
- Классификация сущностей.
- Даталогическая модель.

Описание предметной области

В этой предметной области будет анализироваться интеллектуальный анализ данных на основе модели CRIPS-DM. Интеллектуальный анализ данных (Data Mining), также известный как обнаружение знаний (Knowledge Discover in Database, KDD) в базах данных, в настоящее время является актуальной проблемой в области искусственного интеллекта и исследований баз данных. В проекте CRISP-DM (CRoss Industry Standard Process for Data Mining) была предложена комплексная модель процесса для выполнения проектов интеллектуального анализа данных. Модель процесса не зависит ни от отрасли, ни от используемой технологии.

Предметная область описывает типичные проблемы интеллектуального анализа данных, которые решаются с помощью процесса. У каждого процесса есть соответствующие подпроцессы, также называемые задачами. Выполнение каждой задачи требует от пользователя выполнения определенных действий для получения выходных данных. Результаты задачи могут быть сведены в несколько частей отчета. Технологии, необходимые для каждого типа проблем, различаются, и алгоритмы, которые необходимы для реализации технологий, также различны. Кроме того, данная база данных также включает связанные функции и атрибуты набора данных, которые можно использовать.

Список сущностей и их классификацию

№	Сущность	Классификация
1	Problem	стержень
2	Process	стержень
3	Task	стержень
4	Activity	стержень
5	Output	характеристика
6	Report_Section	стержень
7	Technique	стержень
8	Algorithm	стержень
9	Function	стержень
10	Dataset	характеристика
11	Next	ассоциация
12	Realize	ассоциация
13	Employ	ассоциация

Таб.1 список сущностей

Инфологическая модель

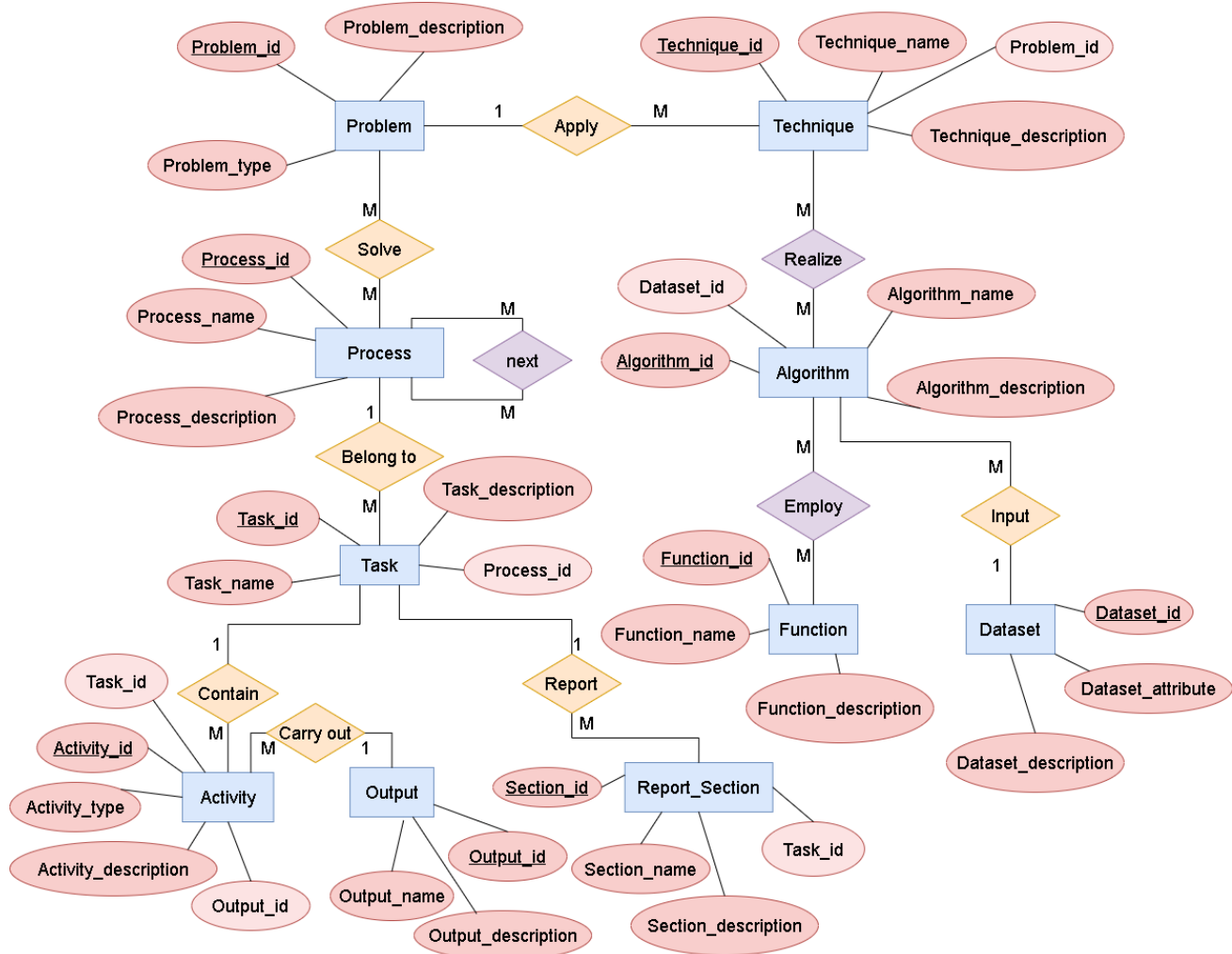


Рис.1 инфологическая модель(ER-диаграмма)

Даталогическая модель

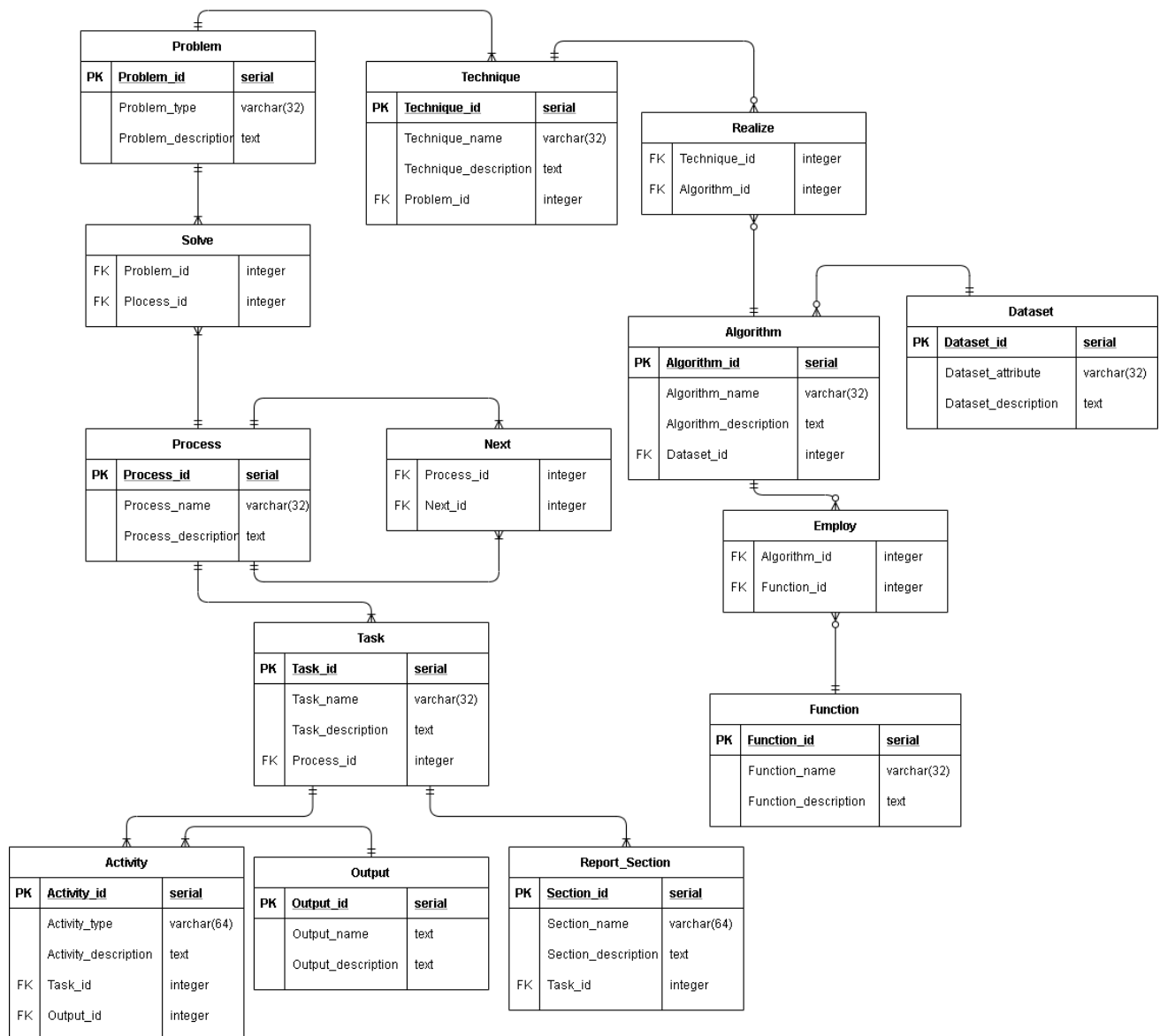


Рис.2 даталогическая модель

Вывод

Выполнив эту лабораторную работу, я научилась работать с DDL и DML, вспомнила типы переменных PostgreSQL, научилась выявлять сущности, составлять инфологическую и даталогическую модели.