**Сопоставление записей реляционных баз данных с использованием многоагентных систем.**

**Введение**

В работе рассматривается решение проблемы сопоставления записей реляционных баз данных, входящих в состав различных автоматизированных систем, имеющих одинаковое смысловое наполнение, но разные системы классификации и кодирования понятий. Решение задачи позволит свести к минимуму ручной труд оператора по сопоставлению понятий.

1. **Постановка задачи**

Каждую запись таблицы реляционной базы данных можно рассматривать как идентификатор (код, первичный ключ) и кортеж данных. Задача оператора при отождествлении понятий состоит в сравнении кортежей данных из двух таблиц и определении их совпадения.

Следующие рассуждения будут строиться на примере сведения объектов учета по ООСР.

Математически разность между двумя кортежами можно выразить как сумму разностей между каждой характеристикой в этом кортеже. Для объектов учета нескольких автоматизированных систем (АС) возможны два вида характеристик – числовые и текстовые.

Числовые характеристики сравниваются строго или нестрого. Строгое сравнение числовых характеристик подразумевает точное равенство (например, протяженность дороги в каждой из АС равняется 2456 км). Нестрогое сравнение числовых характеристик допускает попадание значения в определенный промежуток (координата Х объекта может лежать в пределах от 53,43 до 53,45).

Сравнение текстовых характеристик может быть только нестрогим и основанным на вычисление расстояния Левенштейна. Расстояние Левенштейна (также редакционное расстояние или дистанция редактирования) между двумя строками — это минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую. Поскольку расстояние Левенштейна обладает некоторыми недостатками, предлагается модифицировать его расчет, учитываю длину сравниваемых слов. Здесь написать что делать с расстоянием чтобы получить норм число и еще вычесть его из единицы.

Таким образом мы получим набор натуральных чисел – разницу между каждой характеристикой из сравниваемых кортежей. Просуммировав эти числа мы получим общее редакционное расстояние между кортежами.

В итоге, с математической точки зрения, мы получим граф с двумя наборами вершин и ребрами между вершинами. В получившемся графе необходимо наилучшим образом сопоставить точки из правого набора точкам левого набора, что приводит нас к решению задачи о назначениях.

Задача о назначениях – задача о наилучшем распределении некоторого числа работ между таким же числом исполнителей при условии взаимно однозначного соответствия между множествами работ и исполнителей. При отсутствии однозначного соответствия между множествами (одно множество больше другого) мы дополним меньшее множество, а веса ребер идущих к этим вершинам определим равным бесконечности. Для решения задачи необходимо найти оптимальное назначение из условия максимума общей производительности, которая равна сумме производительности исполнителей.

В математической модели задача представляется в виде двудольного графа, разбитого на два подмножества вершин *X* и *Y* одинаковой мощности *n* и множеством ребер *U,* соединяющих вершины из разных подмножеств. Информация о графе хранится в матрице чисел *Dij*, где *i*, *j* *∈ 1, 2,…, n*, представляющих собой стоимость выполнения *j*-й работы *i*-м исполнителем. Требуется найти перестановку *φ* из элементов множества *X*, такую, что значение ЦФ равно:

