

РАЗРАБОТКА И ЗАПОЛНЕНИЕ ХРАНИЛИЩА ДАННЫХ ПАРАМЕТРОВ КАЧЕСТВА ВОДЫ В РЕЧНОЙ СИСТЕМЕ МЕГАПОЛИСА ДАККА НА ОСНОВЕ МНОГОМЕРНОЙ МОДЕЛИ

*Мд. Насим Акхтар, аспирант, К.Ю.Колыбанов, доцент
кафедры Информационных технологий, МИТХТ им. М.В.Ломоносова
e-mail: cyrk@mail.ru*

Создание на основе многомерной модели данных хранилища данных параметров качества воды позволит создать единую информационную основу для сбора, накопления и обработки информации о качестве воды во всей речной системе мегаполиса Дакка, а также обеспечит информационную поддержку для принятия решений по выбору технологий очистки воды, сооружению станций очистки сточных вод и разработке других водоохраных мероприятий.

Developing and populating of data warehouse for water quality parameters based on multidimensional data model allows to collect, store and process ecological information of Dhaka river system and provides information support for decision making, such as selecting of water treatment technology, constructing of sewage treatment plant etc.

Ключевые слова: многомерная модель данных, хранилище данных, параметры качества воды.

Key words: multidimensional data model, data warehouse, water quality parameters.

Дакка – столица, деловой, промышленный центр и крупнейший мегаполис государства Бангладеш. В настоящее время население города составляет более 12 миллионов человек, прирост населения оценивается около 3% в год. В непосредственной близости от территории мегаполиса расположена замкнутая речная система, включающая такие реки, как Балу и Тонги Хал в северо-восточной части мегаполиса; Тураг, Бурганга и частично Далешари в западной части, а также Лахиа к востоку от мегаполиса, слияние которой на юге с Далешари образует реку Мегна. Высокая степень загрязнения воды в речной системе мегаполиса обусловлена, в основном, поступлением необработанных муниципальных и промышленных сточных вод, помимо сельскохозяйственных стоков. Снабжение населения питьевой водой осуществляет Департамент водоснабжения и канализации Дакки, используя для этого преимущественно грунтовые воды. Однако уровень грунтовых вод снижается, по данным исследований, на 1–3 м/год, что в ближайшей перспективе может привести к оседанию почв, изменениям в растительном покрове и другим экологическим последствиям, в том числе уменьшению поступления питьевых вод. Таким образом, проблемы изучения поверхностных вод и повышения их качества становятся все более актуальными.

Изучению качества воды в речной системе мегаполиса Дакка посвящены работы многих исследователей, однако большинство таких исследований охватывает лишь отдельные небольшие части речной системы [1]. Это вызывает ряд проблем при попытке сопоставления данных, полученных разными исследователями. Например, наиболее распространенным подходом является сбор данных о параметрах качества воды на отдельных участках одной из рек, без учета

влияния соседних притоков. Кроме того, каждое исследование обычно использует свою географическую привязку координат точек отбора проб, для которых используется последовательная нумерация либо указывается расстояние в километрах от первой точки. Поэтому даже если пробы были отобраны в одной и той же точке, она может получить различные обозначения и иметь различное расстояние от условных начальных точек в различных отчетах. Таким образом, создание на основе графовой модели речной системы хранилища данных параметров качества воды позволит создать единую информационную основу для сбора, накопления и обработки информации о качестве воды во всей речной системе в целом [2].

Согласно одному из классических определений, хранилище данных представляет собой «предметно-ориентированный, интегрированный, неизменяемый и поддерживающий хронологию набор данных, предназначенный для обеспечения принятия управляющих решений». Основной моделью, используемой как на этапе проектирования, так и при заполнении хранилища данных, является многомерная модель данных. Обычно выделяется одна главная таблица фактов, а остальные связанные с ней таблицы являются таблицами измерений. Главной таблицей фактов хранилища данных параметров качества воды (рис. 1) является таблица концентраций загрязняющих веществ, измеренных различными исследователями по результатам анализа проб, взятых в разное время в различных точках речной системы. Все таблицы являются нормализованными, связь между таблицами и поддержание ссылочной целостности данных осуществляется посредством первичных и внешних ключей (поля с названиями, начинающимися на «Код»).

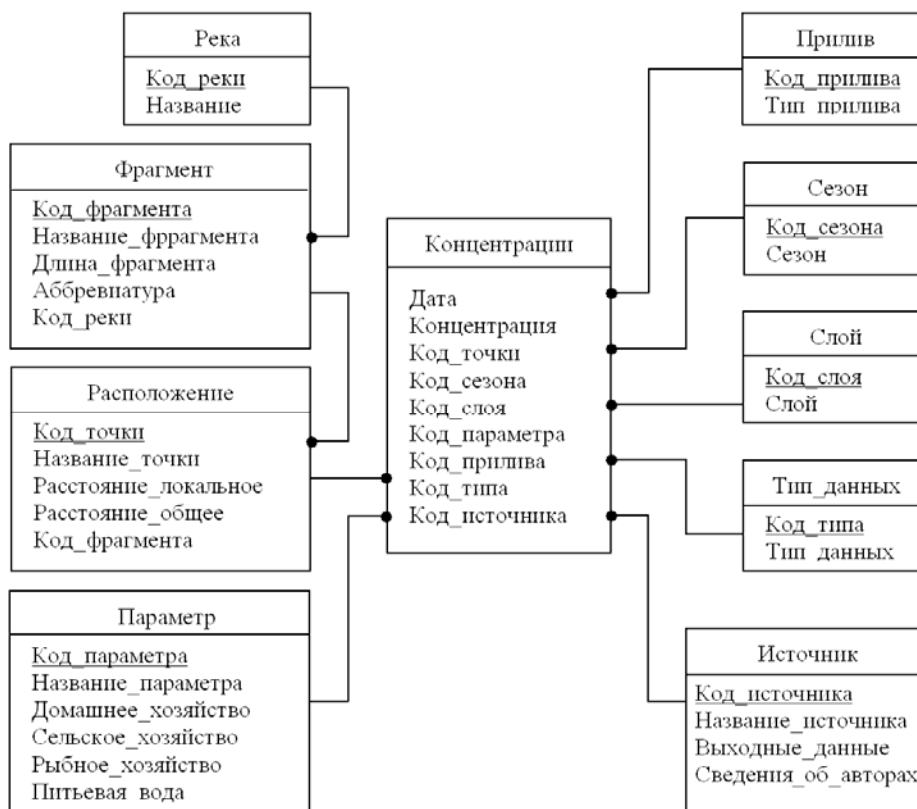


Рис. 1. Многомерная модель данных типа «снежинка».

К основным параметрам, определяющим качество воды в речной системе, относятся концентрация растворенного кислорода, химическая и биохимическая потребность в кислороде, содержание твердых частиц, а также концентрации химических загрязнителей (анионы кислот, ионы металлов, аммиак и т.д.). Параметры качества воды хранятся в отдельной справочной таблице «Параметры». В ней же приведены сведения о стандартах качества воды для различных целей: для использования в качестве питьевой воды, для домашнего хозяйства, для сельского и рыбного хозяйства.

Специфической особенностью речной системы мегаполиса Дакка является ярко выраженная сезонная зависимость качества воды. В течение влажного сезона (июнь-август) за счет дополнительного притока дождевых вод качество воды несколько улучшается по сравнению с сухим (декабрь-февраль) и переходными сезонами. Отдельная таблица позволяет анализировать накопленные данные по измерению «Сезон». Аналогично, значение концентрации загрязняющих веществ не может быть адекватно интерпретировано без указания на глубину, с которой взята проба – на дне или у поверхности воды, что отражено в измерении «Слой», а также необходимо указание на общий уровень воды в реке, подверженный приливам и отливам. Отдельное измерение «Тип_данных» позволяет различить данные непосредственных измерений концентраций загрязняющих веществ и результаты усреднения, моделирования

и прогнозирования.

Измерение, указывающее на расположение точки отбора проб, является иерархическим и включает в себя три связанные таблицы – «Река», «Фрагмент» и «Расположение». Для описания структуры речной системы была разработана графовая модель, вершинами которой являются условные границы и точки слияния отдельных рек, а дугами – фрагменты рек. Это позволяет объединять отдельные фрагменты рек в речные подсистемы независимо от принадлежности фрагментов к исходным рекам в географическом смысле. В таблице расположения точек отбора проб каждое расположение имеет две координаты – локальное неизменяемое расстояние от начала соответствующего фрагмента, и полное расстояние от начала речной подсистемы, пересчитываемое при добавлении в эту подсистему новых фрагментов. Таким образом, иерархическая организация этого измерения обеспечивает решение задачи совместного анализа данных, полученных разными исследователями на отдельных участках речной системы.

Ссылка на литературный источник, из которого были получены данные о качестве воды, хранится в отдельной таблице «Источники». Литературные данные для заполнения как справочных таблиц измерений, так и основной таблицы фактов, были получены из научно-исследовательских отчетов таких организаций, как Департамент Защиты Окружающей Среды; Институт моделирования качества воды; Депар-

тамент водоснабжения и канализации Дакки; Всемирный банк; Азиатский институт технологий. Исходные данные хранились как в бумажной форме, так и в электронном виде – в неструктурированных текстовых файлах и плоских электронных таблицах Microsoft Excel.

При заполнении хранилища данных были реализованы все три основные стадии информационного процесса – извлечение, преобразование и загрузка данных. На стадии извлечения данных было выполнено сканирование и распознавание текста бумажных документов, преобразование данных включало структурирование данных, извлечение отдельных атрибутов из составных полей, нормализацию данных и их приведение к третьей нормальной форме для заполнения реляционных таблиц, а также простановку внешних ключей для связи таблицы фактов и справочных таблиц измерений данных.

Кроме того, к стадии преобразования данных относятся такие операции, как проверка и преобразование типов и форматов данных в

столбцах, изменение названий полей и смена кодировки текста (при необходимости). На последней стадии преобразованные структурированные данные были загружены в таблицы хранилища данных. К настоящему моменту хранилище данных включает более 11000 записей, собранных различными исследователями и организациями в различные периоды времени в различных участках речной сети за период с 1980 по 2009 годы.

Таким образом, хранилище данных на основе интеграции экологической информации в рамках единой информационной платформы обеспечивает возможность оценки текущей экологической ситуации, выявления тенденций, моделирования и прогнозирования изменения качества воды, а также обеспечивает информационную поддержку для принятия решений по выбору технологий очистки воды, сооружению станций очистки сточных вод и разработке других водоохраных мероприятий.

ЛИТЕРАТУРА:

1. Джебуннахар. Моделирование переноса загрязнений в реке Buriganga (Dhaka city, Bangladesh) на основе интегральных критериев качества воды / Джебуннахар, К. Ю. Колыбанов, О. А. Морозова // Ученые записки МИТХТ. – 1999. – № 1. – С. 93–94.
2. Мд. Насим Акхтар. Графовая модель речной системы мегаполиса Дакка, Бангладеш / Мд. Насим Акхтар, К. Ю. Колыбанов // Интеграл. – 2010. – № 5. – С. 30–31.