

УДК 004.82

ОСОБЕННОСТИ РАЗРАБОТКИ ПОДСИСТЕМЫ ХРАНЕНИЯ ИНФОРМАЦИИ ДЛЯ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ОБЛАСТИ АНАЛИЗА ТЕЛЕКОММУНИКАЦИОННЫХ ДАННЫХ

Савельев О.О., Шевченко А.И.

*Государственный университет информатики
и искусственного интеллекта, г. Донецк*

В работе рассмотрена задача проектирования подсистемы хранения данных для системы поддержки принятия решений в области анализа телекоммуникационных данных. Полученные результаты анализа форм представления и типов исходных данных определяют необходимость применения концепции хранилища данных. Определены требования к хранилищу и его функциям в системе. Проведен синтез реляционной модели хранилища данных.

Анализ телекоммуникационных данных применяется не только в области управления телекоммуникационными предприятиями, но и во многих других смежных областях. Одной из сфер применения подобного анализа является решение специальных задач:

- установление особенностей возникновения и протекания различных происшествий;
- восстановление хронологии совершения лицами неправомερных действий;
- расследование преступлений.

Настоящее состояние наличия соответствующего доступного информационного обеспечения не удовлетворяет существующий спрос. Поэтому разработка подобных информационных систем является актуальной задачей.

Объектом исследования данной работы является процесс хранения информации в системах поддержки принятия решений

в области анализа телекоммуникационных данных, применяемых при решении специальных задач.

Предметом исследования являются модели хранилищ данных, аспекты их разработки и реализации.

Цель данной работы – увеличение качества функционирования систем поддержки принятия решений (СППР) при анализе телекоммуникационных данных за счет применения разрабатываемой модели хранилища данных. Для достижения поставленной цели необходимо решение следующих задач:

- анализ форм представления и типов исходных данных;
- определение требований к хранилищу данных и его функций;
- синтез реляционной модели хранилища данных.

Научная новизна данной работы заключается в том, что впервые разработана реляционная модель хранилища данных, позволяющая производить информационно-поисковый, оперативный (OLAP) и интеллектуальный анализ (Data Mining) телекоммуникационных данных, полученных из различных источников.

Практическая значимость состоит в возможности использования разработанной модели в прикладных СППР.

В работе [1] рассмотрены концептуальные проблемы разработки интеллектуальных СППР в области анализа телекоммуникационных данных. Была предложена концепция, основанная на применении типовой схемы СППР, и рассмотрены особенности разработки систем на ее основе, среди которых выделены преимущества применения физического хранилища данных (Data Warehouse) в качестве подсистемы хранения информации.

Исходными данными для СППР является статистическая информация, предоставляемая операторами мобильной и стационарной телефонной связи. Данные материалы могут быть представлены в виде отчетов выборок по запросам к базам данных операторов. На основе анализа ранее предоставленных отчетов их можно классифицировать следующим образом.

По оператору связи: почти все операторы мобильной и

стационарной связи Украины: «МТС», «Киевстар», «Астелит», «Укртелеком».

По типу мониторинга:

- мониторинг всех видов событий, происходящих на одной либо нескольких конкретных базовых станциях («сотах»), находящихся в определенной географической местности, оператора мобильной связи за конкретный временной период;
- мониторинг всех событий, происходящих по конкретному номеру абонента мобильной связи за конкретный временной период;
- мониторинг всех событий, происходящих на определенном узле стационарной телефонной связи за конкретный временной период;
- мониторинг всех событий, происходящих по одному либо нескольким номерам абонента стационарной телефонной связи за конкретный временной период;
- личные данные об абонентах мобильной и телефонной связи.

По типу носителя:

- бумажный носитель – распечатка отчета;
- электронный носитель – файлы форматов TXT, XLS, PDF, HTM/HTML.

По типу структурирования информации:

- таблица;
- таблица с текстовыми комментариями;
- структурированный текст;
- слабоструктурированный текст.

Также возможно использование фрагментов баз данных операторов, но достаточно низка вероятность наличия правовой и технической возможности их получения и обработки. Разные операторы используют различные форматы баз данных и различные СУБД. Кроме того, каждый оператор предоставляет материалы в индивидуальном формате, т.е. количество столбцов в таблицах, форматы записи дат, количественных и строковых величин,

комментариев различны. Всего при анализе было выделено 14 различных форматов, не исключено, что в дальнейшем их количество может увеличиться. А также различны сами типы носителей, на которых предоставляется информация. Следовательно, необходима разработка подсистемы, которая бы позволяла:

- обеспечивать возможность интеграции данных из неоднородных источников в распределенной среде;
- эффективно хранить все необходимые данные большого объема в компактной и универсальной форме;
- обеспечивать целостность, непротиворечивость и безопасность данных;
- предоставлять быстрый доступ к данным для информационно-поискового, оперативного и интеллектуального анализа.

Всем предъявляемым требованиям отвечает концепция хранилищ данных. Хранилище данных – предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений [2]. В простейшем представлении хранилище данных – это база данных, спроектированная с учетом некоторых ограничений.

В результате анализа ранее предоставленных материалов был выделен перечень групп данных, вызывающих наибольший интерес с точки зрения анализа:

1. информация о номерах мобильной и фиксированной связи, как-то: непосредственно номер, оператор либо город, IMSI SIM-карты;
2. информация о телефонах – IMEI аппарата;
3. информация обо всех событиях, происходящих между абонентами с фиксированием даты и времени события, типа события, длительности, стоимости и баланса счета инициатора события;
4. информация о базовых станциях, обслуживающих мобильных абонентов.

Используя полученные группы данных, была разработана

ER-модель хранилища данных в нотации Питера Чена [3] (схема объект-отношение/модель сущность-связь), диаграмма которой представлена на рис. 1.

Каждая группа данных выделена в отдельную сущность. Также добавлена сущность «Владелец», включающая всю информацию об интересующих лицах, которая может быть предоставлена операторами связи (если эти лица – контрактные абоненты) или получена из других источников.

В модели учтена возможность владения одним лицом несколькими номерами и телефонными аппаратами, а также возможность перехода номеров и телефонов другим владельцам (вторичный рынок или использование в одной семье/организации).

Хранилище данных, построенное с использованием данной модели, позволит детализировать все связи между

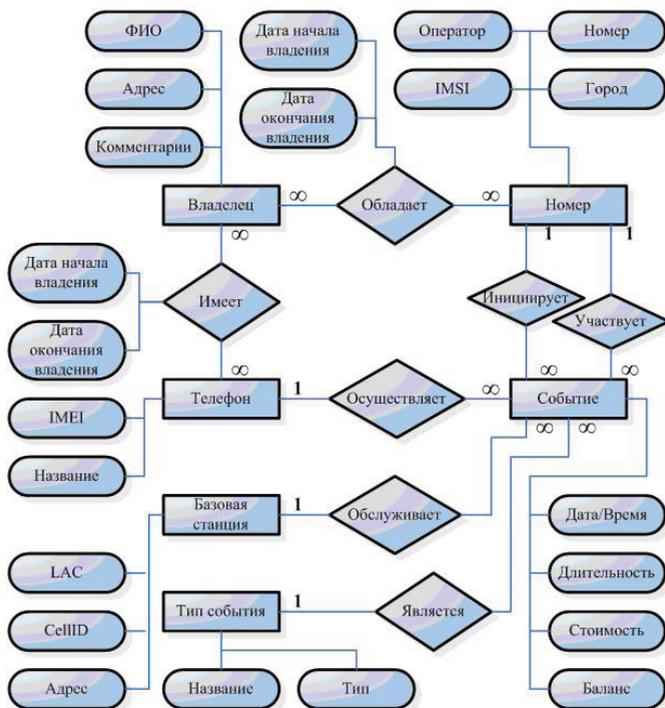


Рисунок 1 – ER-диаграмма хранилища данных

лицами интересующих пользователей СППР; устанавливать их местонахождение в определенные моменты времени, исследовать состояние их денежных счетов по услугам связи.

Согласно диаграмме на рисунке 1 была разработанная реляционная модель хранилища данных, которая показана на рис. 2.

Данная модель не достаточно сильно нормализована, так как концепция хранилища данных допускает контролируруемую денормализацию для эффективного извлечения данных при проведении анализа. С точки зрения анализа таблица «Событие» – это таблица фактов. А все остальные таблицы – таблицы измерений. Таблицы «Номер» и «Телефон» являются нижними уровнями, а таблица «Владелец» – верхним уровнем иерархии измерения

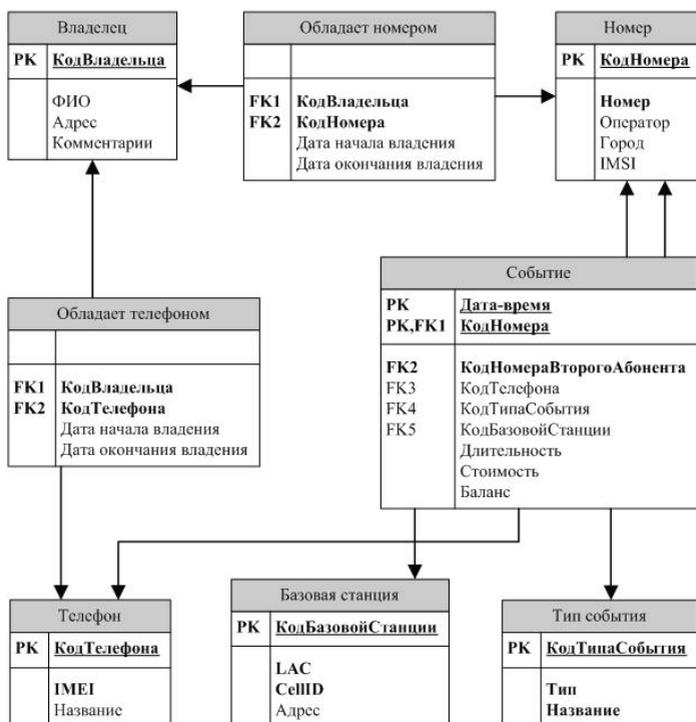


Рисунок 2 – Реляционная модель хранилища данных

«Абонент». Поля «Длительность», «Стоимость» и «Баланс» таблицы фактов являются мерами, также необходимо измерение количества событий. Данная модель отвечает схеме «снежинка» (SnowFlake Shema) [4].

Разработанная реляционная модель хранилища данных отвечает всем поставленным требованиям, обеспечивает выполнение необходимых функций и может использоваться как подсистема хранения информации в СППР в области информационно-поискового, оперативного и интеллектуального анализа телекоммуникационных данных для решения специальных задач.

Литература

- [1] Савельев О.О. О концепции создания информационной системы интеллектуального анализа данных телекоммуникационных компаний в рамках разработки интеллектуальной системы поддержки принятия решений / О.О. Савельев // Искусственный интеллект. – 2010. – № 3.
- [2] Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2007. – 384 с.
- [3] Chen P. The entity-relationship model – toward a unified view of data / Peter Pin-Shan Chen // ACM Transactions on Database Systems (TODS) – New York: ACM, 1976. – V. 1. – p. 9-36.
- [4] Федоров А. Введение в OLAP-технологии Microsoft / Федоров А., Елманова Н. – М.: Диалог-МИФИ, 2002. – 272 с.: ил.