

О ПОДХОДАХ К ПОСТРОЕНИЮ СИСТЕМЫ НАКОПЛЕНИЯ И ОБРАБОТКИ ДАННЫХ ГОРОДСКОЙ СРЕДЫ

Л. В. Рудикова-Фронхёфер

*Учреждение образования «Гродненский государственный университет имени Янки Купалы»,
ул. Ожешко, 22, 230022, г. Гродно, Беларусь, rudikowa@gmail.com*

В статье излагаются общие подходы к построению системы накопления и обработки данных городской среды. Для определения составляющих городской среды, а также моделирования отдельных аспектов используется системный подход, который учитывает определенные ограничения и специфику предметной области взаимодействия городской среды и горожан. Кроме того, приводятся рекомендации по реализации универсальной системы данных городской среды на основе технологии складирования данных. Полученные результаты важны для методологии разработки универсальной системы городской среды, а также – расширения формальных описаний для теории урбанистики.

Ключевые слова: городская среда; модель; объекты городской среды; системный анализ; прагматическая модель; архитектура системы; технология складирования данных.

Развитие современных городов – нелинейная и сложная задача, для решения которой следует учитывать большое число средовых, социальных, экономических и экологических аспектов, а также необходимо согласовывать комплекс взаимосвязанных мероприятий, совместно обеспечивающих достижение поставленных целей развития. Сложность управления процессом формирования новых и модификации имеющихся городских территорий определяется количеством различных взаимосвязанных факторов и заинтересованных лиц, требующих учета всех составляющих при планировании мероприятий. Итоговой целью развития современных городов, несомненно, является улучшение качества жизни населения и развития стратегического потенциала города [1-4].

Основная концепция предлагаемой системы для хранения и анализа данных городской среды основана на технологии складирования данных. Разработка системы наряду с транзакционными системами предполагает использованием хранилища данных с учетом того, что, в конечном итоге, универсальная система будет предоставлять большой комплекс услуг соответствующим группам пользователей. Очевидно, что ресурсоемкость системы будет расти по мере того, как система будет наполняться данными и обслуживать все большее количество пользователей. Немаловажную роль в плане определения концепции построения системы также играет необходимость осуществления аналитической обработки поступающих данных, поиска информации, хранение документов, а также предоставление программного интерфейса для доступа к функциональности системы извне [5-8].

Отметим, что формирование архитектурной концепции системы, включающей учет требований к современным городам, основано на понятии прагматической модели города [3, 4].

Формирование прагматической модели городской среды направлено на учет потребительских свойств городской среды с точки зрения объективно выраженных условий жизни горожан. Прагматическая модель строится на концепциях городских инфраструктур-функций-услуг. На рис. 1 представлена основная схема городских инфраструктур, которые можно рассматривать как важные для городской среды. Одни и те же сервисы могут быть включены в разные городские инфраструктуры и одновременно выполнять разные функции. Например, парк может служить как туристическим центром, так и зоной отдыха. По этой причине, в настоящее время один из самых популярных способов описания городской среды является использование онтологического подхода [9]. Созданные онтологии городов предполагают

наиболее полное описание сущностей и компонентов городской среды в целом [10, 11] или в рамках отдельных предметных областей. Например, онтология транспорта [12], онтология Интернета вещей [13].

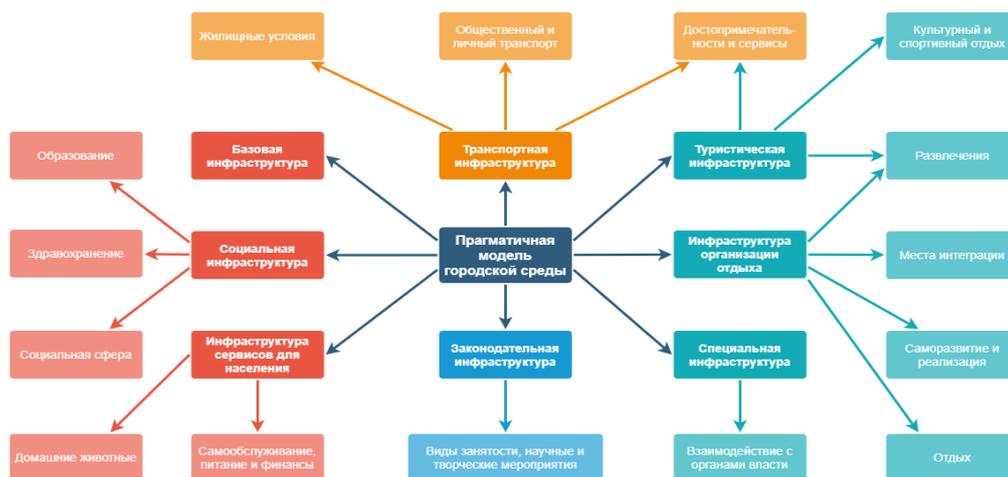


Рис. 1. Схема основных городских инфраструктур

Таким образом, может быть предложена концептуальная модель, где для каждой инфраструктуры $I \in \mathbf{I}$, где \mathbf{I} – конечный набор всех городских инфраструктур, определен состав функций $\{f_1, \dots, f_n\} \in I$, где каждая функция реализуется через конечный набор услуг $\{a_1, \dots, a_m\} \in f_i$.

При построении прагматической модели городской среды учитывается способность инфраструктур, сформированных городскими функциями и услугами, удовлетворять потребности горожан. Предлагается следующая модель, основанная на исследованиях в области оказания услуг горожанам.

Для каждой городской службы учитываются следующие параметры.

Территориальное расположение службы. Этот параметр задается координатами услуги, если она физически расположена в городской среде, или зоной покрытия, если есть возможность получить услугу на заказ.

Мощность услуги P , которая определяет легкость получения услуги для большого количества желающих, или объем потока клиентов, который услуга может обслуживать в единицу времени. В рамках данной статьи дискретное измерение мощности используется на трех уровнях: «1» – соответствует малому бизнесу; «2» – соответствует небольшим сетевым организациям; «3» – крупные центры обслуживания населения.

Сотрудничество $K_{a_1, a_2} = K(a_1, a_2)$ – характеристика, определяющая повышение потребительских свойств услуги при наличии других услуг в непосредственной близости. В этой статье дискретное измерение сотрудничества используется также на трех уровнях: «1» – службы не взаимодействуют друг с другом; «2» – услуги немного повышают привлекательность в сотрудничестве; «3» – значительный рост привлекательности в сотрудничестве.

Безопасность $S \in [0, \dots, 1]$ – характеристика, определяющая возможность получения услуги без вреда для клиента. Значение $S = 0$ соответствует абсолютно безопасному получению услуги, $S = 1$ самому опасному.

Физический комфорт $PC \in [0, \dots, 1]$ – характеристика, определяющая физическую легкость получения услуг клиентами любой социальной группы. $PC = 0$ соответствует худшему значению комфорта, лучшему $PC = 1$.

Эстетический комфорт $VC \in [0, \dots, 1]$ – это характеристика, которая определяет качество дизайна и предоставления услуг, включая окружающий контекст. $VC = 0$ соответствует худшему значению комфорта, $C = 1$ лучшему.

Следует отметить, что этот список не является полным и может быть дополнен или исправлен. Отметим также, что значение параметров P , K , которые определяются на основе свойств услуг, представленных в исходных данных, значение параметров S , PC , VC , определяются на основании данных жалоб обращения населения к конкретным объектам городской среды, реализующим соответствующие услуги.

На рис. 2 представлена диаграмма компонентов для универсальной системы накопления и анализа данных городской среды, которая базируется на технологии складирования данных.

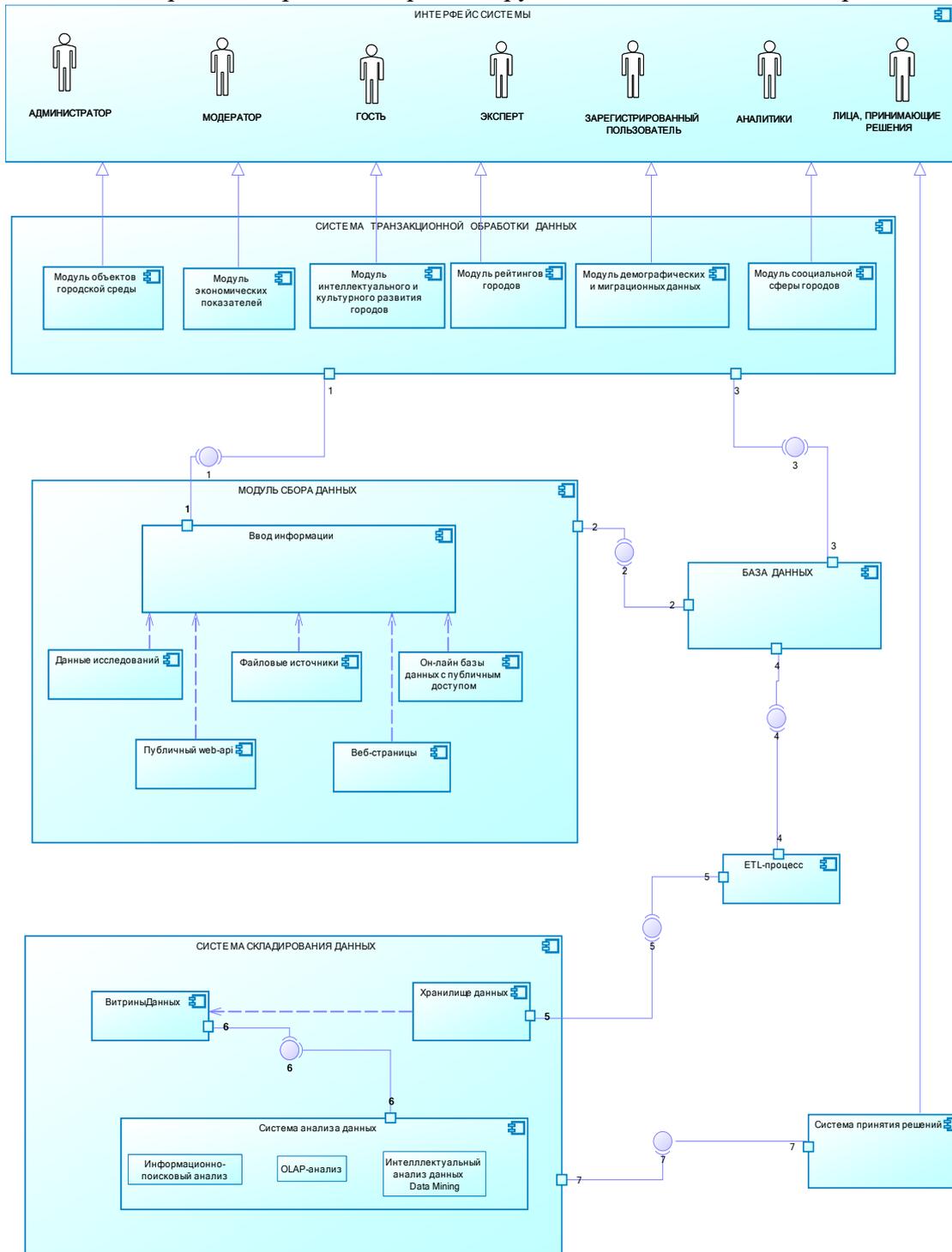


Рис. 2. Диаграмма компонентов для системы городской среды

Основными компонентами системы выступают *Система транзакционной обработки данных*, *Модуль сбора данных*, *База данных*, *Система складирования данных*. Кроме того, важные аспекты связаны также с ETL-процессом и Системой принятия решения.

Основная часть работы с системой приходится на транзакционную обработку данных, т.е. непосредственная работа с реальными данными охватывает обработку данных объектов городской среды, обработку экономических показателей городской среды, работу с расчетом индексов городов, проведение анализа данных текущих данных интеллектуального и культурного развития города, статистическую обработку демографических и миграционных данных, работу с данными социальной сферы.

Данный функционал доступен через единый веб-интерфейс пользователя и поддерживается соответствующим для каждой группы. Отдельно реализуется модуль сбора данных: в нем должен поддерживаться ввод данных через интерфейс, должны добавляться соответствующие данные из файлов исследований, данные из наследуемых файлов, файлов онлайн баз данных, производится сбор данных с веб-страниц и через публичный веб-аpi. Информация Модуля сбора данных доступна через интерфейс обмена данными Системе транзакционной обработки данных, а также Базе данных через соответствующие порты. Естественно, для дальнейшего накопления и обработки данных с использованием OLAP-средств и методов DataMining, предусматривается Система складирования данных, основу которой составляет Хранилище данных.

ETL-процесс поддерживает процесс извлечения информации из Базы данных (в системе возможна распределенная база данных или же совокупность OLTP-систем), ее преобразование в соответствии с требованиями к формату данных в хранилище, и непосредственную загрузку данных в хранилище данных. Для поддержки ETL-процесса применяются программы извлечения данных для чтения записей в исходной базе данных и для подготовки информации, хранящейся в этих записях, к процессу преобразования. Для извлечения данных из исходной базы данных можно поступить следующим образом: разработать собственные программы, выбрать готовое специализированное программное обеспечение для ETL-процесса или же использовать собственное и готовое ПО.

Архитектура области Хранилища данных, как правило, состоит из следующих областей.

Область временного хранения данных (Staging Area) – предназначена для временного хранения данных, извлеченных из систем-источников; является промежуточным слоем между операционными системами компании и хранилищем данных и включает: витрины данных (Data Marts) – тематические наборы данных, хранящиеся в виде пригодном для их анализа; витрины данных ориентированы на поддержку конкретных бизнес-процессов, приложений, подразделений компании, бизнес-целей; интерфейсы обмена данными с другими системами (Data Exchange Interface или Feedback Area) – таблицы базы данных, в которых хранятся подготовленные для передачи в другие информационные системы компании данные из области постоянного хранения данных.

Область постоянного хранения данных включает: детальные данные (System of records) – область хранения детальных данных, приведенных к структуре модели данных корпоративного хранилища, прошедших очистку и обогащение; агрегаты (Summary area) – сгруппированные по времени (чаще просуммированные) детальные данные; метаданные (Metadata) – являются важной частью архитектуры хранилища данных, которые описывают структуры для хранения данных; архивные данные (Archive data) – для длительного хранения данных.

В нашем случае Витрины данных вынесены в отдельный компонент в силу того, что через них будут поступать требуемые данные в Систему анализа данных, причем, нагрузка потоков данных будет достаточно высока.

Как правило, система анализа данных может быть расширяемой и иметь модульную структуру. Основные программно-системные модули, которые добавляются в нашу систему

анализа следующие: подсистема информационно-поискового анализа, включающая информационный поиск, визуальный поиск и семантический поиск; подсистема OLAP-анализа (англ. online analytical processing – интерактивная аналитическая обработка) – технология обработки данных, которая предоставляет итоговую (суммарную, агрегированную) информацию на основе больших массивов данных, структурированных по многомерному принципу; подсистема интеллектуального анализа данных Data Mining – инструмент для обнаружения в накопленных данных знаний, неизвестных ранее, которые могут помочь для принятия решений в аспекте развития городской среды.

Основная функция системы принятия решений (СПР, DSS – Decision Support System) – это вывод рекомендаций на основании изучения исторического и текущего состояния объекта исследования городской среды и сравнения их с информацией, хранящейся в базе знаний системы.

Отметим также, что со стороны Системы принятия решений можно выделить следующие группы пользователей: лица, принимающие решения, аналитики и другие пользователи.

Лица, принимающие решения (ЛПР) – пользователи, которым необходимо принимать решения в следующих случаях: при анализе объективной составляющей (понимание и оценка сложившейся ситуации и ограничений, накладываемых внешней средой); выявлении предпочтений ЛПР (выявление и ранжирование приоритетов), учете неопределенности в оценках ЛПР и формировании его предпочтений; для генерации возможных решений (формирование списка альтернатив); при оценке возможных альтернатив, исходя из предпочтений ЛПР и ограничений, накладываемых внешней средой; для анализа последствий принимаемых решений; для выбора лучшего варианта (с точки зрения ЛПР).

Аналитики – пользователи, которым необходимо сформулировать и проверить конкретные гипотезы, а также создавать требуемые запросы различного плана и подготовить отчеты по конкретным направлениям предметной области.

Другим пользователям, как правило, доступна общая информация, связанная с универсальной системой и для них ограничены возможности по обработке имеющихся данных предлагаемой системы. Однако, стоит помнить, что, в зависимости от целей использования системы, всегда можно определить конкретного пользователя и предоставить ему определенные права на доступ и обработку информации, а также – к отдельным модулям и функционалу системы.

Полученные результаты являются актуальными для специалистов в области ИТ-технологий и обработки данных, которые занимаются разработкой программных комплексов различной сложности, расширением структурной методологии и анализом данных различного профиля. Кроме того, результаты могут быть интересны для государственных и частных структур с целью выработки соответствующей стратегии регулирования развития городов в экономическом, территорияльном и социально-культурных аспектах.

Библиографические ссылки

1. Рудикова Л. В., Жавнерко Е. В. О моделировании данных предметных-областей практико-ориентированной направленности для универсальной системы складирования и обработки данных // Системный анализ и прикладная информатика. Мн.: БНТУ, 2017. №3. С. 12-19.

2. Рудикова, Л.В. Об общей архитектуре универсальной системы хранения и обработки данных практико-ориентированной направленности // Системный анализ и прикладная информатика. Мн.: БНТУ, 2017. № 2. С. 12-19.

3. City Information Modeling: The system Approach for Formation Requirement in Spatial Development / S. Mityagin [et al.] // Procedia Computer Science. 2020. Vol. 178. P. 134-144.

4. City Information Modeling: City Information Modeling: Designing a conceptual data model / S.Mityagin [et al.] // Communications in Computer and Information Science. Vol. 1349. EGOSE 2020. Pp. 219-231.

5. Wrembel R., Koncilia C. Data warehouses and OLAP: concepts, architectures, and solutions // IRM Press. 2007. PP. 1-26.
6. Carr R., Harrington M. Effective Communication Through Visual Design: Tables and Charts. Strategy Institute, 2011. 211 p.
7. Руководство Microsoft по проектированию архитектуры приложения / Patterns&Practices. 2-е издание. Microsoft, 2009. 529 с.
8. Codd E. F., Codd S. B., Salley C. T. Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate // Technical report. 1993. 342 p.
9. A Platform for Urban Analytics and Semantic Data Integration in City Planning / A. Psyllidis [et al.] // Springer, 2015. 1098 p.
10. Simonelli L., Amorim A.L. City Information Modeling: General Aspects and Conceptualization // American Journal of Engineering Research. Issue 10, 2018. Vol. 7. P. 319-324.
11. Malgundkar T., Rao M. Gis driven urban traffic analysis based on ontology, 2012. Т. 4. № 1. С. 15-23.
12. Espinoza-arias P., Poveda-villalón M. Applied sciences Ontological Representation of Smart City Data: From Devices to Cities, 2019. 321 p.
13. Kilov H., Ross J. Information modeling an object-oriented approach // Prentice Hall, 1994. 432 p.