

*ДЕНИСЮК А.В.,
КАВИЦКАЯ В.С.,
ЛЮБЧЕНКО В.В.*

ОПРЕДЕЛЕНИЕ АДЕКВАТНОЙ МОДЕЛИ ДЛЯ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Обоснована ограниченность реляционной модели данных для использования современными информационными системами. Определены основные особенности знаний, которые присущи современным информационным системам. Проанализированы нереляционные модели данных, такие как фреймы, объектные модели, онтологические модели с точки зрения отображения особенностей знаний. Определена адекватная модель данных для задач, решаемых информационными системами.

The article summarizes limitations of relational database approach in modern informational systems. Analysis is made based on main distinctive features of knowledge that is processed by modern informational systems. Non-relational data models such as frames, object model and ontology are considered as alternatives to relational approach in regard to defined particularities. As result, rules are defined for choosing the adequate data model to be used in modern informational systems.

Введение

Любая информационная система (ИС) обязательно оперирует данными. Данные представляют собой совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки. Преобразование и обработка данных позволяет получить информацию. Например, в базах данных хранятся различные данные, а по определенному запросу система управления базой данных выдает требуемую информацию.

Усложнение ИС, а также решаемых ими задач и необходимой для этого информации, приводит к необходимости оперировать не данными, а знаниями. При этом знания рассматриваются как осознание и толкование определенной информации с учетом путей наилучшего ее использования для достижения конкретных целей. Очевидно, что переход к использованию знаний требует обратить внимание на модель, которой представляются знания в ИС, для простоты их понимания разработчиком и эффективного их использования в процессе работы ИС. В связи с этим актуальной является задача выбора адекватной модели данных для задач, решаемых ИС.

1. Анализ литературных данных и постановка проблемы

ИС оперирует различными данными в своей деятельности. Некоторые данные могут быть использованы непосредственно, без дополнительной обработки. Тем не менее, большинство

данных требуют дополнительной обработки в ИС, а именно, подготовки и форматирования данных, анализа данных и извлечения знаний из этих данных [1]. С ростом необходимости использования знаний в современных ИС, возросла сложность решаемых задач. Реляционные модели данных (РМД) уже не могли эффективно решать эти задачи [2]. Остро стала проблема выбора модели данных, которая бы позволила справиться с возросшей сложностью задач.

2. Цель и задачи исследования

Целью данной работы является обобщение результатов анализа существующих моделей представления данных для того, чтобы сформулировать правила выбора адекватной модели данных для задач, решаемых ИС.

Для достижения поставленной цели решались такие задачи:

- определить особенности знаний, используемых в современных ИС;
- проанализировать способности реляционной модели отображать особенности сложных знаний;
- проанализировать способности нереляционных моделей отображать особенности сложных знаний;
- определить адекватную модель для представления знаний в современных ИС.

3. Материалы и методы исследования

3.1. Анализ способности реляционной модели данных представлять знания ИС

Определение состава и структуры знаний, необходимых для функционирования ИС, осу-

ществляется на основе модели предметной области (ПрО). Модель ПрО – это совокупность структурированных моделей сущностей ПрО, их свойств и отношений между ними, а также ограничений, определяющих контекст ПрО [3].

В настоящее время наиболее распространенной формой представления модели ПрО в ИС являются данные – отдельные факты, характеризующие объекты, процессы и явления ПрО, а также их свойства [4].

Ключевым свойством данных является то, что почти всегда они являются результатами измерений и наблюдений, преобразованными для хранения и представления в ИС. Но стоит отметить, что данные, в том числе, могут быть высказаны человеком на основании его опыта.

Наиболее распространенной моделью представления данных в настоящее время является реляционная модель данных (РМД) – логическая модель данных, в которой данные представляют собой набор отношений (таблиц), отвечающих определенным условиям целостности, а также существуют операторы манипулирования отношениями. Кроме того, в состав РМД обычно включают теорию нормализации. На основе РМД строятся реляционные базы данных (РБД).

РМД является простой моделью ПрО, в которой не предполагается или предполагается небольшая изменчивость, а РБД хранит данные, собираемые в соответствии с моделью ПрО, зафиксированной на этапе разработки. РМД имеет ряд неоспоримых достоинств, среди которых следует отметить [5]:

1. Предсказуемость результатов работы с данными. В основе реляционной модели лежит математическая модель – любой корректный запрос к БД выдаст ответ, однозначно определяемый схемой БД и конкретными данными.

2. ПрО часто достаточно естественно описывается в терминах таблиц, кроме представлений иерархических структур.

Эволюция ИС, существенное усложнение решаемых ими задач, появление новых технологий реализации ИС привели к тому, что данные и РМД стали представлять «узкое место» в работе систем. В результате этого возрос интерес к использованию знаний в ИС.

Знания – это закономерности ПрО (принципы, связи, законы). Их основным свойством является то, что почти всегда они – результат интеллектуальной деятельности, представляю-

щей либо обработку данных, либо обработку уже существующих знаний.

Для хранения знаний разрабатывают базы знаний (БЗ), которые содержат структурированную информацию, покрывающую некоторую ПрО, для использования ИС с конкретной целью. Полноценные БЗ содержат в себе не только фактическую информацию, но и правила вывода, допускающие автоматические умозаключения о вновь вводимых фактах и, как следствие, осмысленную обработку информации.

Сравнение РМД и модели знаний позволяет сделать вывод, что модель знаний – это упрощенная модель, описывающая часть понятий ПрО для работы со знаниями данной ПрО, в то время как РМД – это упрощенная модель, описывающая только те понятия ПрО, которые требуется сохранить.

Кроме этого можно утверждать, что РМД является простой моделью знаний, в которой предполагается нулевая или небольшая изменчивость, а РБД – это БЗ, построенная с использованием простой модели знаний, которая зафиксирована на этапе разработки.

Очевидно, что если знания можно описать простой моделью знаний, то для работы ИС с ними можно использовать РМД и РБД. Однако, применимость таких знаний существенно ограничена. В современных ИС все чаще возникает необходимость учитывать такие особенности знаний, как

- наличие исключений из правил;
- изменения в наборе атрибутов сущности, возможность появления новых либо отсутствия уже известных;
- изменения типа атрибутов в зависимости от контекста;
- необходимость отображения отношений;
- изменчивость структуры ПрО;
- начальная неопределенность в модели, так как при проектировании новой БЗ почти всегда не имеется полного объема входящей информации, из-за чего ее структура почти всегда некорректна;
- постоянные изменения и уточнения, как данных о сущностях ПрО, так и структуры и связей между ними;
- неоднородность данных;
- разное количество данных, так как разные классы сущностей могут быть представлены разным количеством данных, при этом количество может отличаться на несколько порядков.

В связи с этим возникает проблема выбора адекватной модели данных для представления знаний ИС.

3.2. Анализ способности нереляционных моделей данных представлять знания ИС

Известно, что модель данных следует рассматривать, как комбинацию трех компонент: коллекции типов объектов данных, коллекции общих правил целостности и коллекции применимых операций [6]. Фактически, способ определения каждой из компонент зависит от цели моделирования данных, то есть для нас представляет интерес адекватность модели данных – совпадение ее свойств и свойств моделируемых данных. Дадим определение.

Адекватная модель данных – это абстрактное, логическое определение объектов, моделирующих структуру данных, операторов, моделирующих поведение данных, и прочих элементов, в совокупности составляющих абстрактную машину доступа к данным, свойства которой соответствуют закономерностям ПрО.

Как было отмечено ранее, в случае достаточности простой модели знаний РМД является адекватной моделью. Рассмотрим свойства известных нереляционных моделей данных – фреймов, объектных моделей и онтологий – с точки зрения моделирования сложных знаний.

Фрейм – это структура данных, представляющая стереотипную ситуацию. Он состоит из нескольких слотов, каждый из которых имеет свое назначение и связанное с ним значение. Кроме того, со слотами могут быть связаны специальные процедуры, предназначенные для специфической обработки значений в слотах. Это единица представления знаний, детали которой при необходимости могут быть изменены согласно ситуации. Фреймовая модель является достаточно универсальной, так как обеспечивает не только фреймы для обозначения объектов и понятий, но и фреймы-роли, фреймы-ситуации и др. [7].

Проанализируем способность фреймовой модели обеспечивать отображение особенностей сложных знаний.

1. Фреймовая модель плохо учитывает наличие исключений из правил, из-за сложности реализации самой структуры фрейма, что на порядок увеличивает трудоемкость создания модели.

2. Фреймовая модель плохо учитывает изменения в наборе атрибутов сущности, так как возникает проблема хрупкости базового

класса — изменения в наборе атрибутов в корневом узле могут быть небезопасными для наследуемых классов (при этом критично удаление атрибутов в базовом классе).

3. Фреймовая модель плохо учитывает изменения типа атрибутов в зависимости от контекста, так как возникает проблема — изменения типа атрибутов в корневом узле могут быть критичными для наследуемых классов.

4. Фреймовая модель позволяет отображать отношения с помощью отношений «абстрактное – конкретное» и «целое – часть». Иерархическая структура фреймов основывается на отношениях «абстрактное – конкретное». Отношение «целое – часть» касается структурированных объектов и показывает, что объект нижнего уровня является частью объекта верхнего уровня.

5. Изменчивость структуры ПрО при описании фреймовой моделью приводит к снижению скорости работы механизма вывода и увеличивает трудоемкость процедур внесения изменений в родовидовую иерархию модели ПрО.

6. Фреймовая модель плохо учитывает начальную неопределенность в модели, потому что обладает четко выраженной иерархией и процедурой наследования значений, что требует определения набора фреймов на начальном этапе проектирования.

7. Фреймовая модель плохо учитывает постоянные изменения и уточнения о данных и сущностях ПрО, так как фреймам присуща иерархическая структура, что ведет за собой проблему хрупкости базового класса.

8. Неоднородность данных описывается фреймовой моделью, так как модель позволяет отобразить все разнообразие знаний через фреймы-структуры, фреймы-роли, фреймы-сценарии, фреймы-ситуации.

9. Фреймовая модель позволяет учитывать разное количество данных.

Объектная модель описывает структуру объектов, составляющих ПрО, их атрибуты, операции, взаимосвязи с другими объектами. В объектной модели отображаются только те понятия и объекты реального мира, которые важны для проектируемой ПрО. Соответственно, в объектно-ориентированных БД хранятся не записи, а объекты. Цель разработки объектной модели — описать объекты, составляющие в совокупности проектируемую ПрО, а также выявить и ука-

затъ различные зависимости между объектами [8].

Проанализируем способность объектной модели обеспечивать отображение особенностей сложных знаний.

1. Объектная модель позволяет описывать исключения из правил, так как обладает возможностью выразить сложные зависимости.

2. Объектная модель плохо учитывает изменения в наборе атрибутов сущности, а также возможность появления новых либо отсутствия уже известных (удаление является критичным), так как изменения в наборе атрибутов базового класса могут оказаться разрушительными для производных классов.

3. Изменения типа атрибутов в зависимости от контекста объектной моделью описываются на низком уровне, так как изменения типа атрибутов базового класса могут быть небезопасными для производных классов.

4. Объектная модель на высоком уровне отображает отношения в БЗ, так как она реализует механизм ассоциативных связей и отношений наследования.

5. Объектная модель на среднем уровне учитывает изменчивость структуры ПрО, так как обладает проблемой уязвимости базового класса.

6. Объектная модель позволяет учитывать начальную неопределенность в модели, однако она обладает сложностью реализации. Использование объектно-ориентированной модели для проектирования новой БЗ, которая почти всегда не имеет полного объема входящей информации, требует больших временных затрат и приводит к построению сложной и требовательной к ресурсам модели.

7. Объектная модель позволяет учитывать постоянные изменения и уточнения, как данных о сущностях ПрО, так и структуры и связей между ними, однако усложняется методология проектирования — объектная модель требует введения дополнительных способов представления информации о ПрО и методов ее анализа.

8. Неоднородность данных при использовании объектной модели приводит к построению сложной и требовательной к ресурсам модели.

9. Объектная модель на высоком уровне позволяет описывать разное количество данных.

Онтология — формальная спецификация разделяемой концептуализации, которая имеет место в некотором контексте ПрО. При этом под концептуализацией понимают, кроме сбора понятий, также всю информацию, касающуюся понятий — свойства, отношения, ограничения, аксиомы и утверждения о понятиях, необходимые для описания и решения задач в избранной ПрО [9].

Онтология ПрО определяет формальное приближение концептуализации. В онтологии зафиксирована та часть концептуализации, которая зависит от взгляда на мир применительно к конкретной области интересов. На формальном уровне, онтология — это система, состоящая из набора понятий и набора утверждений об этих понятиях, на основе которых можно строить классы, объекты, отношения, функции и теории [10].

Проанализируем способность онтологической модели обеспечивать отображение особенностей сложных знаний.

1. Онтологическая модель позволяет учитывать исключения из правил, так как обладает возможностью описывать сложные знания, начиная с простых словарей до формальных логических структур, которые позволяют выполнять сложные логические выводы.

2. Онтологическая модель позволяет учитывать изменения в наборе атрибутов сущности, возможность появления новых либо отсутствия уже известных за счет использования свойств общего типа, с реализованной для них возможностью переопределения типа.

3. Онтологическая модель позволяет учитывать изменения типа атрибутов в зависимости от контекста за счет использования соответствующего математического аппарата (например, теории категорий), позволяющего учитывать эволюционные изменения.

4. Онтологическая модель позволяет отображать отношения за счет отношений таксономии («kind-of», «is-a»), партономии («has part»), генеалогии («parent-child»). Полноценные онтологии имеют больше возможностей для описания отношений, чем простые таксономические иерархии. Отноше-

ния (связи) между понятиями значительно увеличивают количество способов осуществлять навигацию и поиск по ПрО, выполнять анализ, классификацию и визуализацию знаний.

5. Изменчивость структуры ПрО описывается онтологией на высоком уровне. Онтологическая модель декларативным образом описывает ПрО, соответствующие словари типов, ограничения использования этих данных, в предположении понимания того, что может быть выведено из этого описания, а также добавлено в процессе проектирования ПрО.

6. Онтологическая модель позволяет учитывать начальную неопределенность в модели, так как предоставляет совместно используемые понятия, которые могут использоваться для формулировки и записи утверждений об экземплярах ПрО.

7. Онтологическая модель позволяет учитывать постоянные изменения и уточнения, как данных о сущностях ПрО, так и структуры и связей между ними, так как онтологии являются более чем простыми словарями точно определенных понятий: их реальная сила заключается в описании связей между сущностями ПрО.

8. Онтологическая модель позволяет учитывать неоднородность данных, так как в каждой ПрО могут существовать различные понимания одних и тех же терминов. Онтология определяет соглашение о значении терминов и является посредником между человеком и машинно-ориентированным уровнем представления информации.

9. Онтологическая модель позволяет учитывать разное количество данных, так как разные классы сущностей могут быть представлены разным количеством данных, при этом количество может отличаться на несколько порядков.

4. Результаты исследований

Таким образом, можно утверждать, что для сложных знаний адекватной моделью следует считать онтологическую модель, так как она

способна наследовать некоторые характеристики описываемых понятий реального мира. К таким характеристикам можно отнести изменчивость структуры ПрО, изменения в наборах атрибутов сущностей и связях между ними, наличие исключений из общих правил и другие. Какие конкретно характеристики унаследует онтология, зависит от того, какую часть реального мира она описывает.

Выполненный анализ позволяет сделать следующее обобщение. Онтология по сути – это упрощенная модель, описывающая часть понятий реального мира для работы со знаниями данной области, в то время как РМД – это упрощенная модель, описывающая только те понятия реального мира, которые требуется сохранить.

Онтологии можно классифицировать следующим образом:

1. Простые – онтологии с постоянной структурой, однородностью, небольшой изменчивостью и связностью. Такие онтологии обычно описывают либо небольшую часть понятий реального мира, либо описывают понятия ПрО, которая была создана человеком (а значит, сама является моделью – упрощенным представлением).

2. Сложные – онтологии с изменчивой структурой, обладающие большой степенью связности, изменчивости, либо неоднородности. Такие онтологии обычно описывают большую часть понятий реального мира, понятия на грани нескольких областей и др.

Выводы

Очевидно, РМД является простой онтологией, в которой предполагается нулевая или небольшая изменчивость, а РБД – это БЗ, построенная с использованием простой онтологии зафиксированной на этапе разработки.

Из этого следует простой вывод: знания, которые можно описать простой онтологией, возможно также представить в виде данных. Можно также предположить, что при достаточно мощном инструменте для работы с РБД возможно также сохранение знаний обладающих сложной онтологией.

Список литературы

1. Géczy P. Big data characteristics / P. Géczy. – A multidisciplinary journal of global macro trends, 2014. — V. 3 (6). — P. 94-104.
2. Bhat U. Moving Towards Non-Relational Databases / U. Bhat, S. Jadhav. — International Journal of Computer Applications, 2010 — V. 1 (13).

3. Любченко В. В. Методологічні основи розробки моделей предметних областей з метою, що інкапсульована : автореф. дис. ... д-а техн. наук : 05.13.06 / В. В. Любченко. — Одеський національний політехнічний університет, 2013. — 43 с.
4. Гаврилова Т. А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В. Ф. Хорошевский. — СПб.: Питер, 2001. — 384 с.
5. Когаловский М. Р. Энциклопедия технологий баз данных [Текст] / М. Р. Когаловский. — М.: Финансы и статистика, 2002. — 800 с.
6. Akerkar R. Knowledge-Based Systems / R. Akerkar, P. Sajja. — Jones & Bartlett Learning, 2009. — 354 p.
7. Minsky M. A framework for representing knowledge / M. Minsky. — B.; N. Y.: Gruyter, 1980. — 153 p.
8. Booch G. Object-oriented analysis and design with applications [Електронний ресурс] / G. Booch. — Santa Clara, California. — 1998. — Режим доступа: [www / URL:http://www.cvauni.edu.vn/imgupload_dinhkem/file/pttkht/object-oriented-analysis-and-design-with-applications-2nd-edition.pdf](http://www.cvauni.edu.vn/imgupload_dinhkem/file/pttkht/object-oriented-analysis-and-design-with-applications-2nd-edition.pdf). — Загл. с экрана.
9. Brunetti G. Feature ontologies for the explicit representation of shape semantics / G. Brunetti, G. Stephan. — International Journal of Computer Applications in Technology (IJCAT), 2005. — V. 25. — #2/3/4 — P. 192-202.
10. Grimm S. Knowledge Representation and Ontologies / S. Grimm, P. Hitzler, A. Abecker. — Semantic Web Services: Concepts, Technology and Applications. — Berlin: Springer, 2007. — P. 51-106.