

УДК 004.62:004.942

Е.В. Малахов, канд. техн. наук, доц., Одес. нац. политехн. ун-т

СТАБИЛЬНОСТЬ ИНФОРМАЦИИ И ФОРМИРОВАНИЕ ИЕРАРХИИ ИНФОРМАЦИОННЫХ ХРАНИЛИЩ

Е.В. Малахов. Стабільність інформації та формування ієрархії інформаційних сховищ. Розглянуто проблеми створення інформаційних сховищ на підставі незмінності (або стабільності) інформації в базах даних. Приведено схеми переносу інформації з баз даних до сховищ кількох рівнів та діаграми часу щодо моментів цього переносу.

E.V. Malakhov. The information stability and formation of the datawarehouses hierarchy. Problems of creating the datawarehouses based on invariability (stability) of the information at the databases are considered. The schemes of the information transferring from databases to multilevel datawarehouses and the time diagrams of the moments of this transferring are presented.

В ряде областей использования информации могут возникать задачи, связанные с обеспечением доступа пользователей к большим массивам данных только с целью их поиска и получения. Причем сами данные, однажды сформированные, уже не подвергаются модификации. Динамические базы данных в такой ситуации являются неэффективными, и в мировой практике на сегодняшний день для решения подобных задач осуществляется переход к информационным хранилищам.

В качестве наглядного примера можно рассмотреть сферу организационного управления учебным процессом в ВУЗе.

При поступлении абитуриента в базу данных деканата формируется запись, содержащая значения идентифицирующих атрибутов текущего экземпляра объекта: фамилия, имя, отчество, год и место рождения и пр. Затем добавляется еще ряд специфических атрибутов: факультет, номер группы и порядковый номер в группе.

В течение семестра в деканат поступает информация о текущей успеваемости студента по различным дисциплинам. Эти данные являются *динамическими* и в течение семестра *модифицируются*. После сессии данные об успеваемости по той или иной дисциплине фиксируются, приобретают *свойство стабильности* и используются в качестве критерия продолжения обучения. Более того, по завершении изучения дисциплины *степень стабильности* данных возрастает настолько, что эта информация не претерпевает изменений до конца обучения студента в ВУЗе [1]. И, наконец, при окончании студентом ВУЗа можно сказать, что экземпляр объекта “студент” обладает единственным набором значений, соответствующих результатам изучения дисциплин и защиты дипломного проекта, которые *никогда* не могут быть изменены.

Под *свойством стабильности* понимается возможность изменения данных при определенных условиях. Конечно, более корректно было бы говорить о том, что любые данные, даже динамические, обладают свойством стабильности, характеризуемым *степенью стабильности* S , под которой укрупненно можно понимать отношение интервала времени между потенциальными моментами модификации данных — *периода изменения данных* M к *периоду* (или длительности) *жизненного цикла* объекта T , который они описывают: $S = M/T$.

Если величина $S \rightarrow 0$, то речь идет о быстро и постоянно меняющейся информации типа результатов серии экспериментов, котировок акций, оценок студентов в течение семестра и т.п. При $S \rightarrow 1$ уже можно говорить о стабильных данных. Если же $S \geq 1$ или, еще нагляднее, $S \gg 1$, то это предельная степень стабильности или *абсолютно неизменяемые данные*.

Вообще понятия “свойство стабильности” и “степень стабильности” требуют отдельного рассмотрения, т.к. возможна их различная математическая трактовка. Например, S можно определять как произведение вероятности изменения данных P и интервала времени, в течение которого возможно такое изменение, т.е. $S = PM$.

Недостатком такого подхода является невозможность определения предельной степени стабильности. Достоинством же является возможность определения S в зависимости от вероят-

ности изменения данных P_i в различные периоды M_i жизненного цикла T объекта, который они описывают:

$$S = \sum_{i=1}^N S_i = \sum_{i=1}^N P_i M_i,$$

где N — количество рассматриваемых периодов при $\sum_{i=1}^N M_i = T$.

При такой интерпретации можно проследить динамику изменения S в течение жизненного цикла объекта.

Динамическая информация необходима деканату для контроля текущей успеваемости студентов. Информация о результатах сессии доводится до руководства ВУЗа для анализа и подведения итогов учебного процесса. Результаты обучения в ВУЗе после его окончания могут потребоваться не только в рамках учебного заведения. Как следствие, динамическая информация может храниться в обычных реляционных базах данных деканата и необходима только на этом административном уровне. Остальные же два вида данных, как уже было отмечено, являются стабильными и по этому критерию отвечают требованиям информационных хранилищ.

Таким образом, всю рассмотренную информацию целесообразно разделить на различные уровни, как по времени существования, так и по доступу, путем переноса информации в информационные хранилища в соответствии со схемой рис. 1.

Моменты времени, в которые выполняется формирование записей с результатами оценки знаний студента по какой-либо одной дисциплине можно показать на диаграмме (рис. 2). В базе данных деканата в начале изучения дисциплины создается единственная запись, которая модифицируется в течение всего интервала времени изучения дисциплины. По окончании этого процесса данная запись уничтожается.

В момент времени, соответствующий окончанию сессионного контроля, значения оценки знаний из базы данных деканата переносятся в информационное хранилище первого уровня (информационное хранилище 1). При переносе к этому атрибуту добавляется атрибут времени со значением, соответствующим указанному моменту.

Если же указанный момент времени соответствует завершению изучения дисциплины, то значение оценки знаний является итоговым и добавляется еще и в запись информационного хранилища второго уровня (информационное хранилище 2). Атрибут времени при этом добавляется в информационное хранилище не в этот момент, а только при окончании или прекращении обучения в ВУЗе — совместно с результатом защиты дипломного проекта, принимая значение даты этой защиты.

В дальнейшем выпускник ВУЗа может быть принят на работу в какую-либо организацию. Следовательно, начиная с даты выдачи диплома данный экземпляр объекта “студент” прекращает существование, и в базах и хранилищах



Рис. 1. Схема переноса информации

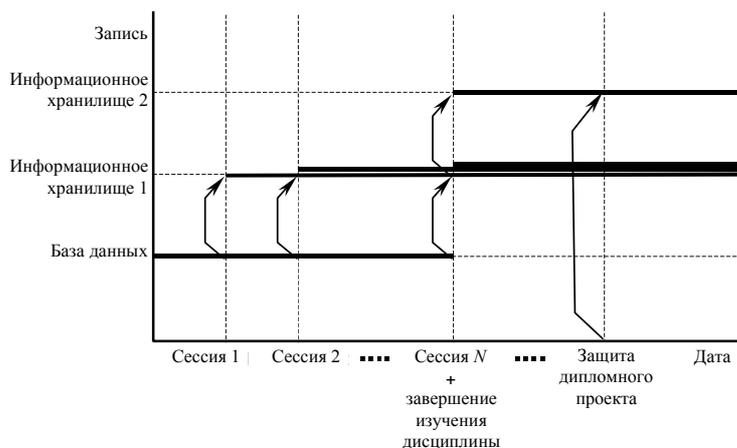


Рис. 2. Моменты формирования и время существования информации об успеваемости по конкретной дисциплине

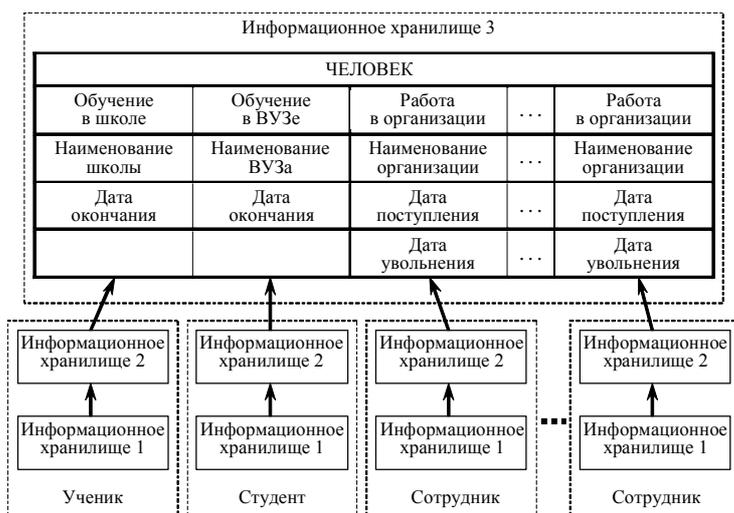


Рис. 3. Формирование информационного хранилища верхнего уровня

го уровня будет сформирована при увольнении сотрудника, т.е. перед уничтожением соответствующего экземпляра объекта.

Необходимо отметить, что ряд атрибутов объектов “студент” и “сотрудник” полностью совпадают. Например, фамилия, имя, отчество, год и место рождения есть не только у всех этих объектов, но и имеют одинаковые значения у всех пар соответствующих экземпляров этих объектов. В то же время такие атрибуты, как место жительства и семейное положение существуют у обоих объектов, но их значения для конкретного экземпляра могут меняться. Это очевидно, т.к. речь идет об одном и том же человеке.

Следовательно, можно выделить объект “человек” как объект суперкласса (супертипа), а объекты “студент” и “сотрудник” рассматривать как объекты соответствующих подклассов (подтипов). В этом случае экземпляр объекта суперкласса осуществляет миграцию в рамках моделей состояний объектов подклассов, составляющих “жизненный цикл” объекта суперкласса.

Если сопоставить приведенную организацию сохранения данных с диаграммой моментов времени распределения информации по хранилищам, то очевидно, что момент уничтожения экземпляра объекта какого-либо подкласса можно зафиксировать в виде значения соответствующего атрибута экземпляра объекта суперкласса. С учетом еще большего повышения степени стабильности данных целесообразно вынести этот атрибут в информационное хранилище еще более высокого, например, государственного, уровня (информационное хранилище третьего уровня (информационное хранилище 3)) (рис. 3).

Если в записи информационного хранилища добавить ссылки на соответствующие хранилища более низкого уровня, то приведенную структуру можно использовать в “обратном” порядке: обратившись в хранилище данных государственного уровня, можно из хранилищ, расположенных ниже на дереве иерархии, получить информацию о каком-либо периоде жизни человека с “любой степенью детализации”.

В данном примере рассматривалась “биографическая” информация о каком-либо человеке. Но аналогично можно рассмотреть и дру-

данных соответствующей организации создается новый экземпляр объекта “сотрудник”. При этом схема переноса информации из базы данных в хранилище первого уровня и из информационного хранилища первого уровня в хранилище второго уровня аналогична рассмотренной выше. Схема претерпит изменения только в отношении атрибутов экземпляров “сотрудников”: они характерны для данной организации и их значений. Моменты времени переноса информации из базы данных в хранилище первого уровня соответствуют изменениям статуса сотрудника. Запись же в информационном хранилище второго

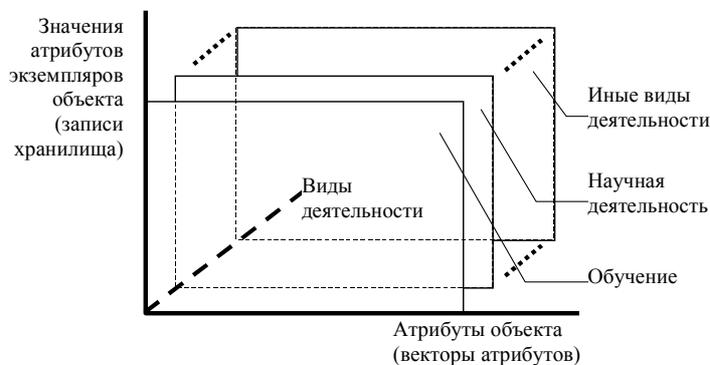


Рис. 4. Организация хранилища неоднородной информации

гие аспекты деятельности человека. Например, научную деятельность, политическую деятельность и проч. При этом речь будет идти о деятельности во вполне определенный период времени, например, в течение обучения в ВУЗе.

В этом случае уже на уровне одного и того же информационного хранилища приходится иметь дело с совершенно другим вектором атрибутов, естественно, кроме фамилии, имени и т.п.

Очевидно, что при таком подходе возникает необходимость в организации многомерных хранилищ данных (рис. 4), которые позволят хранить и получать разнородную информацию об одном и том же объекте.

Аналогичным образом можно хранить, обрабатывать и получать информацию о любой сфере: жизнедеятельности фирм и организаций, экологической обстановке, медицинском обслуживании, банковской деятельности и проч.

Литература

1. Малахов Е.В., Иванченко О.В. Базы данных как информационные модели предметных областей // Тр. Одес. политехн. ун-та. — Одесса, 1999. — Вып.1 (7). — С. 225 — 228.

Поступила в редакцию 6 апреля 2004 г.
