

УДК 681.3

Свойства аналитических измерений

Е. Б. Асламов, А. В. Висков, М. Б. Фомин

*Кафедра информационных технологий
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

В данной статье рассматривается формальное определение аналитического измерения, исходя из анализа которого сформулированы и доказаны некоторые свойства аналитических измерений

КЛЮЧЕВЫЕ СЛОВА: многомерные модели данных, аналитическое измерение, OLAP.

1. Введение

Анализ многомерных данных, как правило, производится в привязке к конкретному информационному контексту, определяемому предметной областью. Точка зрения на анализируемые данные задается аналитическими измерениями – сущностями предметной области, от которых зависит предмет анализа. Аналитическое измерение (далее просто измерение) состоит из множества объектов (называемых членами измерений) одного или нескольких типов, обеспечивающих информационный контекст анализируемых данных.

2. Определение измерения

Измерение включает объекты, представляющие различную степень детализации анализируемых данных. Совокупность элементов, обозначающая одну и ту же степень детализации, образует уровень измерения. Уровни показывают, как элементы измерения группируются для применения агрегационных функций. Связи, определенные между элементами на разных уровнях агрегации, являются композицией.

Таким образом, измерение можно определить как связный ориентированный граф. Каждая вершина в этом графе соответствует уровню агрегации, а дуга показывает, что каждый элемент уровня, определяющего конец дуги, декомпозируется на множество элементов нижележащего уровня, соответствующего началу дуги.

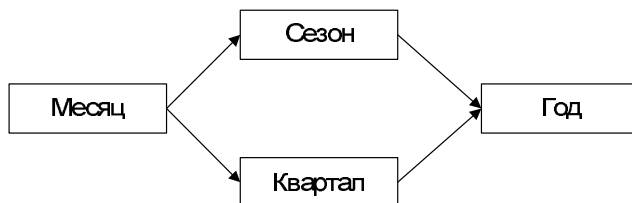


Рис. 1. Пример измерения

3. Свойства измерений

Как показано на рисунке 2, можно выделить три различных, независящих от области применения типа отношений «часть-целое», определяемых составной структурой целого. Выделяется три типа структуры целого:

- 1) масса – в случае отсутствия определенной структуры целого, считающегося гомогенным (например, кучка риса);
- 2) коллекция – в случае, когда рассматриваются различные элементы целого, имеющего однородную составную структуру (например, колонна грузовиков);
- 3) комплекс – в случае, когда различные части целого выполняют различные роли, целое является комплексом с гетерогенной составной структурой (например, части двигатель).

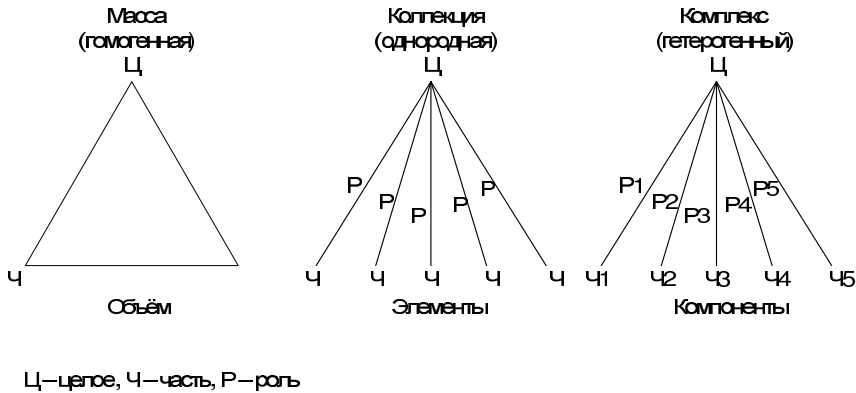


Рис. 2. Типы целого

Главная цель определения связей между различными сущностями измерения состоит в том, чтобы показать, каким образом могут применяться агрегационные функции (такие как сумма, минимум, максимум, среднее, и т.д.). Так как эти функции рассматривают элементы измерения как играющие в агрегации одну и ту же роль, то эти связи должны рассматриваться как коллекция. С этого момента, отношения «часть-целое» между уровнями агрегации в измерении должны пониматься как образующие коллекцию.

В случае рассмотрения коллекции, для описания и обоснования свойств целого, а также связей между составляющими элементами можно использовать аксиоматику классической экстенциональной мереологии [1, 2].

- 1) **Существование.** Если A является частью B , то и A и B существуют.
- 2) **Асимметричность.** Если A является частью B , то B не является частью A .
- 3) **Транзитивность.** Если A является частью B и B является частью C , то A является частью C .
- 4) **Дополнение.** Если A является некоторой частью B , тогда существует другой субъект C , который является оставшейся частью B .
- 5) **Экстенциональность.** A и B состоят из одних и тех же частей тогда и только тогда когда A и B совпадают.
- 6) **Сумма.** Для любых двух субъектов всегда найдется составленный из них третий субъект.

Исходя из определения и аксиом мереологии, можно вывести некоторые свойства измерений.

Свойство 1: Измерение не содержит циклов.

Доказательство. Предположим, что в графе измерения есть цикл. В цикле участвуют хотя бы два уровня. По аксиоме 3 (транзитивности) элемент A одного уровня есть часть элемента B другого уровня и наоборот. Получаем противоречие с аксиомой 2. Значит предположение не верно. □

Свойство 2: Для каждого измерения существует единственный атомарный уровень агрегации, содержащий элементарные сущности, которые не могут быть разделены на более мелкие части.

Доказательство. По свойству 1, существует как минимум один уровень, элементы которого не содержат частей.

Если существует более одного атомарного уровня, то, так как граф измерения связный, и, учитывая аксиому 3, будет существовать элемент E , который является композицией элементарных сущностей каждого из атомарных уровней. По аксиоме 5, все коллекции элементарных сущностей, составляющих E , должны являться одной и той же коллекцией элементов. Таким образом, существует только один атомарный уровень. $\square$

Свойство 3: Для любого измерения, может существовать уровень «Все», содержащий элементы, составленные всеми элементарными сущностями измерения. Если такой уровень существует то,

- 1) его элементы не являются частями никакого другого уровня агрегации;
- 2) этот уровень агрегации содержит в точности один элемент;
- 3) этот уровень единственный в измерении.

Доказательство. Последовательно применяя аксиому 6, можно построить элемент E , состоящий из всех элементарных сущностей измерения.

- 1) Если E будет являться частью некоторого элемента E' то, по аксиоме 4 будет существовать элементарная сущность, не входящая в E , что противоречит предположению. Поэтому, E принадлежит уровню, элементы которого не являющегося частью никакого другого уровня в измерении.
- 2) Если этот уровень будет содержать два элемента, оба содержащие все элементарные сущности, то, по аксиоме 5, они будут являться одним и тем же элементом.
- 3) Этот уровень единственный, потому, что если бы существовал другой уровень, элементы которого включают все элементарные сущности, то они бы совпадали с элементом, имеющимся на уровне «Все» (по аксиоме 5). $\square$

Свойство 4: Те уровни, элементы которых не являются коллекцией элементов любого другого уровня (т.е. они не являются началом дуги в графе) могут быть связаны некоторым ребром с уровнем «Все».

Доказательство. Элемент уровня «Все» может быть разложен на сущности любого уровня, покрывающие атомарный уровень. Если существует уровень, не покрывающий атомарный уровень, то, по аксиоме 6, к его элементам могут быть добавлены новые элементы, составленные из коллекций недостающих элементов атомарного уровня. $\square$

Свойство 5: Каждый элемент уровня не являющегося атомарным содержит как минимум одну часть.

Доказательство. Сущности, не включающие частей, являются элементарными, а все элементарные сущности, по свойству 2, располагаются на атомарном уровне. $\square$

Свойство 6: Каждый элемент уровня, не являющегося атомарным может иметь более чем одну часть.

Доказательство. По аксиоме 4, если элемент включает часть, то найдется другой элемент, являющийся оставшейся частью исходного элемента. $\square$

Свойство 7: Элемент может являться частью нескольких коллекций в одно и тоже время.

Доказательство. Не существует мереологической аксиомы запрещающей, чтобы элемент являлся частью нескольких коллекций. В качестве доказательства можно рассмотреть пример. Как показано на рисунке 3, заданный товар (на уровне «товаров») может принадлежать двум различным видам товаров на уровне «видов товаров». Такое допущение не создает каких-либо концептуальных проблем, однако может являться причиной возникновения вычислительных сложностей. Так, например, некоторые вычисляемые атрибуты сущностей на уровне «семейство» (которые состоят из элементов уровня «виды товаров») должны рассчитываться на основании элементов на уровне «товар» (например, размерность множества разложения элемента подарки (уровень «семейство товаров») не равна сумме размерностей множеств разложений элементов сладости и игрушки (уровень «типы товаров»): $card(подарки) \neq card(сладости) + card(игрушки)$).

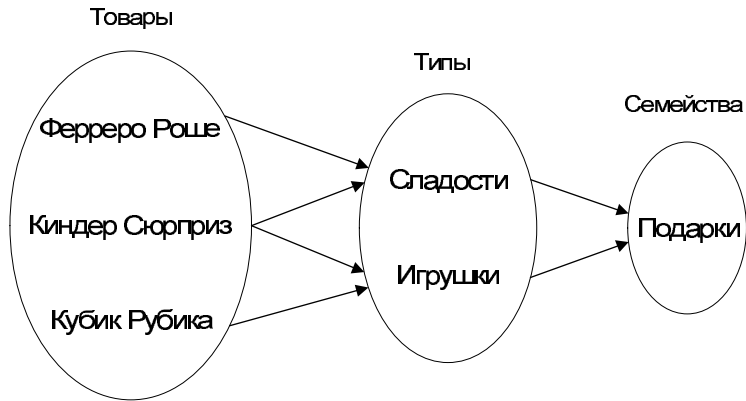


Рис. 3. Пример образования целого с пересечением

Исходя из определения измерения, и, по свойствам 1 и 2, можно сделать вывод, что в общем случае граф уровней агрегации измерения образует полусеть. Кроме того, свойства 3 и 4 показывают, что в графе измерений, для того, чтобы получить сеть, всегда можно определить уровень «Все». Свойства 5 и 6 показывают, что отношение между двумя уровнями будет включать 1..N частей для каждого целого. То есть для каждого элемента на вышележащем уровне всегда можно определить его разложение на нижележащем уровне в соответствующей иерархии. Свойство 7 поясняет, что часть может входить в состав как одного целого, так и нескольких. Следует отметить, что если элемент нижележащего уровня может являться частью только одного элемента вышележащего уровня, то в таком случае он обязательно входит в состав некоторого элемента вышележащего уровня, если же множества разложений элементов могут пересекаться, то, вообще говоря, элемент нижележащего уровня может и не входить в элемент вышележащего уровня. Однако, учитывая аксиому 6, которая свидетельствует о том, что для любой части всегда можно определить целое, получаем, что в случае необходимости всегда можно доопределить элемент вышележащего уровня, который будет включать рассматриваемый элемент нижележащего уровня. Это означает, что элемент нижележащего уровня всегда включается как минимум в один элемент вышележащего уровня в соответствующей иерархии.

На рисунке 4 показаны допустимые кардинальные отношения между элементами уровней в агрегационных иерархиях. Можно выделить два вида отношений, таких что, каждому целому будет принадлежать хотя бы одна часть. В наиболее распространенном случае оказывается, что существует в точности одно целое для каждой части. Однако, также возможно, что определенная часть принадлежит нескольким целым. В таком случае, если найдется часть, которая не принадлежит ни одному целому, всегда могут быть

построены целые, так что каждая часть будет принадлежать хотя бы одному целому.

Кроме отношений «часть-целое» между элементами измерения можно выделить также связи, являющиеся ассоциациями. Такой тип связи используется для того, чтобы показать, что одна сущность описывает некоторое свойство другой сущности. Все сущности в измерении будут связаны с некоторыми другими сущностями, так как они являются их частями, либо для того чтобы показать некоторые их свойства.

Отношение «часть-целое» показывает, как различные элементы группируются вместе в измерении, тогда как ассоциация показывает, какие характеристики доступны для построения выборок сущностей. Таким образом, операции «подъём» («roll-up») и «спуск» («drill-down») будут выполняться вдоль связей «часть-целое» (по иерархии), тогда как выбор элементов («slice-dice») будет выполняться в соответствии с отношениями ассоциации.

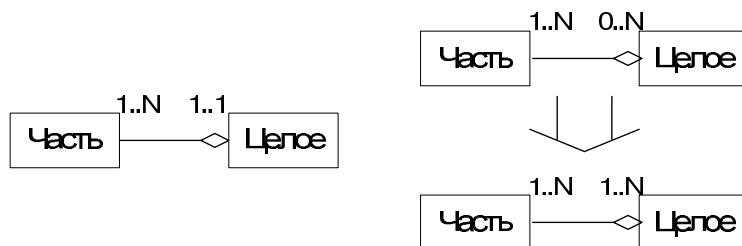


Рис. 4. Возможные кардинальные отношения между элементами уровней в иерархии измерения

Специализация – это одно из наиболее важных отношений, которое должно быть показано на многомерной схеме [3, 4]. Изолированные уровни агрегации не могут быть специализированы для отображения более специфичных аспектов, они должны рассматриваться в рамках измерения. □

Свойство 8: В общем случае, уровень и его специализация могут не принадлежать одному и тому же измерению.

Доказательство. Предположим, что и уровень L и его специализация L_s находятся в одном измерении. Для того чтобы определить полусеть с уровнем «Все», может потребоваться добавить на уровне L_s некоторые элементы, так как в таком случае L_s должен покрывать атомарный уровень. Эти элементы, могут не удовлетворять критерию специализации. Поэтому, не всегда возможно обеспечить принадлежность обоих уровней одному измерению.

На рисунке 5 показан пример, когда измерение «Люди» специализировано в виде измерения «Продавцы» на уровне «Профессия» (сплошная непрерывная стрелка). Эта специализированное измерение содержит уровень, элементами которого являются все люди, являющиеся продавцами, и еще один уровень с единственным элементом (который является сущностью «Профессии»), представляющим собой множество всех продавцов. Разрывная линия показывает, что уровень является специализацией другого (элемент уровня «Все» измерения «Продавцы» является сущностью «Профессии» в случае, когда выполняется критерий специализации «Профессия = Продавец»). Уровень «Возрастная Группа» для измерения «Продавцы» не рассматривается. Следует отметить, что если бы такой уровень в измерении «Продавцы» рассматривался, то он бы не являлся специализацией соответствующего уровня в измерении «Люди», так как его элементы отличались бы от элементов на уровне «Возрастная Группа» в измерении «Люди» (они включали бы меньше людей).

Обобщая пример, если D_s это специализация измерения D на уровне L , то в этом случае D_s включает, по меньшей мере, уровень L_s (специализацию уровня L), и специализации каждого из уровней в D , содержащих часть элементов L_s . Эти специализированные уровни содержат в точности те элементы

соответствующих уровней из D , которые являются любой частью элемента L_s . Кроме этих обязательных уровней, D_s также возможно включаются и другие уровни (которые не являются специализацией каких либо уровней в D), элементы которых не содержатся в D .

Специализации измерения предоставляет возможность показать специфику отношения «часть-целое» для некоторого подмножества элементов измерения, а также выделить общие для этого подмножества атрибуты. Все сущности некоторого уровня имеют общие свойства, так как уровень определяет заданный класс объектов, выполняющих одну и ту же роль в коллекции. Отношение ассоциации, также как и агрегации наследует в рамках специализации. Таким образом, отношение специализации имеет большое значение в вопросе повторного использования измерений.

Рассматриваемые семантические особенности отношения специализации важны не только для конечных пользователей – аналитиков, они позволяют также оптимизировать запросы и повышать их производительность, позволяя, например, осуществлять параллельные вычисления по разным специализациям измерения.

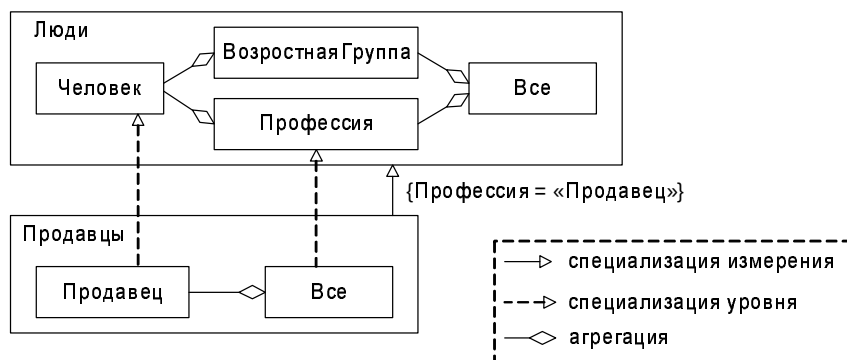


Рис. 5. Пример определения специализации

□

4. Заключение

В работе были рассмотрены и доказаны свойства аналитических измерений. Рассуждения основывались на мереологических аксиомах. Кроме описания связей между элементами одного измерения, также были разобраны возможные виды отношений между различными измерениями, как одной схемы данных, так и разных. Рассмотренный способ описания измерений и вытекающие из него свойства могут быть полезны для построения модели предметной области и позволяют разработчикам информационных систем более чётко проводить классификацию и группировку данных.

Литература

1. PartWhole relations in Object-centered systems: an overview / A. Artale, E. Franconi, N. Guarino, L. Pazzi // Data and Knowledge Engineering (DKE). — Vol. 20, No 3. — 1996. — Pp. 347–383.
2. Gerstl P., Pribbenow S. Midwinters, end games, and body parts: A classification of partwhole relations // International Journal of HumanComputer Studies. — Vol. 43(5,6). — 1995. — Pp. 865–889.
3. Trujillo J., Palomar M., Gomez J. Applying Object-Oriented Conceptual Modeling Techniques to the Design of Multidimensional Databases and OLAP applications //

In Proceedings of the 1st International Conference on WebAge Information Management – WAIM'2000, volume 1846 of LNCS. — Springer, 2000. — Pp. 83–94.

4. Tryfona N., Busborg F., Christiansen J. starER: A conceptual model for data warehouse design // In Proceedings of the 2nd International Workshop on Data Warehousing and OLAP – DOLAP 99. — ACM, 1999. — Pp. 3–8.

UDC 681.3

Analytical Dimensions Properties

E. B. Aslamov, A. V. Viskov, M. B. Fomin

In this paper, we propose a formal definition for an analytical dimension. We formulate and prove some properties of an analytical dimension within the proposed formal definition.