

УДК 519.7

В. М. Левыкин, Т. М. Неофитная

РАЗРАБОТКА ПРАВИЛ ПОПОЛНЕНИЯ ИЕРАРХИЧЕСКОЙ БАЗЫ ЗНАНИЙ**1. Введение**

Модель концептуализации предметной области, используемая в базе знаний автоматизированной системы, может иметь слишком много релевантных деталей, что затрудняет решение отдельных функциональных задач пользователей, т. е. возникает проблема необозримости модели концептуализации (МК) предметной области (ПрО). При этом высокая степень детализации МК ПрО может оказаться несогласованной с необходимой точностью решения и необоснованными дополнительными затратами. В этом случае эффективной является декомпозиция знаний о ПрО в виде иерархии абстракций, которая позволяет распределить знания о ПрО по различным уровням детализации.

Представление семантики предметной области в виде иерархии абстракций является фундаментальным принципом объектно-ориентированной технологии разработки баз знаний [1]. При этом смысл абстракции состоит в извлечении существенных деталей, опуская при этом несущественные, и в раскрытии деталей на других уровнях детализации. На высших уровнях иерархии абстракций используются наименее детальные концептуальные представления, отражающие только самые общие знания о ПрО. На следующих уровнях степень подробности возрастает, причем рассматриваются только отдельные сущности ПрО, а не ПрО в целом.

Рассмотрим основные преимущества поддержки иерархии абстракций в базе знаний автоматизированной системы [1, 2].

1. Появляется возможность игнорировать многие релевантные детали знаний о ПрО на более высоком уровне абстракции, что часто облегчает поиск решения задачи.

2. В базе знаний могут быть аффективно интегрированы абстракции, имеющие отношение к различным пользователям системы.

3. Наблюдается повышенная стабильность базы знаний по отношению к различным эволюционным изменениям.

Выделяют две основных формы абстракции: агрегация и обобщение [2]. Агрегация — это такая абстракция, которая превращает связь между концептами в некоторый агрегированный (составной) концепт. Обобщение — это абстракция, превращающая класс концептов в родовой концепт. Таким образом, совокупность моделей концептуализации ПрО можно стратифицировать по различным уровням агрегации с помощью введения специального отношения агрегации, а также по различным уровням обобщенности с помощью введения специального отношения обобщения.

Существование отношения обобщения между МК ПрО означает, что одна модель является обобщением

другой модели. Существование отношения агрегации между МК ПрО означает, что одна модель является компонентом другой модели. Комбинация отношений обобщения и агрегации приводит к сложной иерархической структуре базы знаний о ПрО, которая имеет определенные семантические закономерности. Данные закономерности необходимо учитывать при разработке процедуры создания баз знаний в диалоге с пользователем. Поэтому целью данной работы является выявление и учет семантических закономерностей иерархических структур систем знаний о ПрО. Для этого в данной работе решаются следующие задачи: моделирование отношений обобщения и агрегации МК ПрО; выявление семантических закономерностей иерархических структур систем знаний о ПрО; разработка правил пополнения иерархической базы знаний в процессе ее разработки с учетом выявленных метазакономерностей иерархических структур систем знаний о ПрО. Данные правила позволяют поддержать семантическую корректность иерархической базы знаний о ПрО.

2. Семантические закономерности иерархических структур систем знаний

База знаний автоматизированной системы разрабатывается таким образом, чтобы иерархическая структура отношений функций классов объектов ПрО, описываемых МК ПрО, предопределяла иерархическую структуру отношений между МК ПрО. Пусть F — множество функций классов объектов, описываемых МК ПрО. На данном множестве естественно определить отношения обобщения и агрегации, которые обозначим «*is-a*» и «*part-of*». Отношения агрегации и обобщения, заданные на множестве функций, являются отношениями порядка, т. е. данные отношения рефлексивны, антисимметричны и транзитивны.

Определим отображение $Func$, которое каждой МК ПрО ставит в соответствие функциональное название описываемого данной моделью класса объектов ПрО. Запись $Func(B) \text{ is-a } Func(A)$ означает, что функция класса объектов, описываемого моделью концептуализации A , является обобщением функции класса объектов, описываемого моделью концептуализации B . Запись $Func(A) \text{ part-of } Func(X)$ означает, что функция класса объектов, описываемого моделью концептуализации A , является частью функции класса объектов, описываемого моделью концептуализации X .

Если функция класса объектов, описываемого моделью концептуализации A , является обобщением функции класса объектов, описываемого моделью концептуализации B , то модель концептуализации A является обобщением модели концептуализации B . Если функция класса объектов, описываемого моделью концептуализации A , является частью функ-

ции класса объектов, описываемого моделью концептуализации X , то модель концептуализации A является частью модели концептуализации X .

Отношения обобщения и агрегации между МК ПрО эффективно описываются с помощью морфизмов. Морфизм из подчиненной МК ПрО в данную МК ПрО назовем морфизмом агрегации. Наиболее строгим определенным морфизмом агрегации является его определение как обладающего свойствами инъективного гомоморфизма и корреспонденции одновременно. Морфизм из данной МК ПрО в ее обобщенную МК ПрО назовем морфизмом наследования. В общем случае морфизм наследования является сюръективной корреспонденцией. Математические определения гомоморфизма и корреспонденции, заданных между МК ПрО, даны в работе [3].

На основании данных о морфизмах агрегации и морфизмах наследования могут быть получены некоторые другие морфизмы между МК ПрО. Действительно, морфизмы наследования и морфизмы агрегации в самом общем случае являются корреспонденциями. Тогда существуют произведенные морфизмы наследования и морфизмы агрегации, называемые далее морфизмами связи, т. е. произведенные корреспонденции являются корреспонденциями. Для определения морфизмов связи воспользуемся следующими правилами.

Пусть о некоторой модели концептуализации A известно, что она является компонентом модели концептуализации X , т. е. существует морфизм агрегации $r: A \rightarrow X$. Кроме того, известно, что модель концептуализации A является обобщением модели концептуализации B , т. е. существует морфизм наследования $d: B \rightarrow A$. Тогда между моделями концептуализации B и X существует морфизм связи $k: B \rightarrow X$, при котором выполняется $k = dr$ (рис. 1, а).

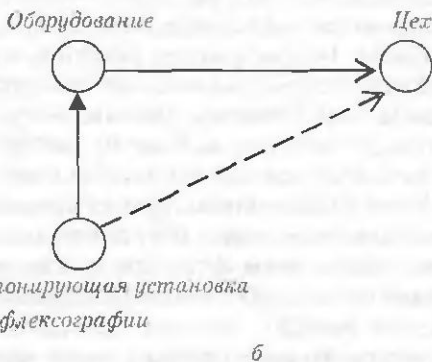
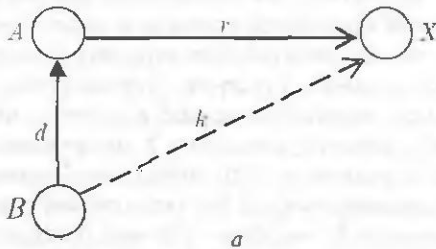


Рис. 1. Определение морфизма связи:

— морфизм наследования;
 — морфизм агрегации; \dashrightarrow — морфизм связи

Для примера рассмотрим фрагмент базы знаний, касающийся области производства флексографских печатных форм (ФПФ) (рис. 1, б). Пусть в базе знаний определены следующие МК ПрО: «Оборудование», «Экспонирующая установка для флексографии», «Цех». Известно, что МК ПрО «Оборудование» является компонентом МК ПрО «Цех», т. е. существует морфизм агрегации из МК ПрО «Оборудование» в МК ПрО «Цех». Кроме того, МК ПрО «Оборудование» является обобщением МК ПрО «Экспонирующая установка для флексографии», т. е. существует морфизм наследования из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Оборудование». Тогда существует морфизм связи из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Цех», свидетельствующий о существовании отношения структурного сходства между данными моделями. Семантика данного отношения заключается в том, что экспонирующая установка для флексографии находится в цехе, поскольку оборудование находится в цехе.

Пусть о некоторой модели концептуализации B известно, что она является компонентом модели концептуализации Y , т. е. существует морфизм агрегации $w: B \rightarrow Y$. Кроме того, известно, что модель концептуализации X является обобщением модели концептуализации Y , т. е. существует морфизм наследования $h: Y \rightarrow X$. Тогда между моделями концептуализации B и X существует морфизм связи $k: B \rightarrow X$, при котором выполняется $k = wh$ (рис. 2, а).

Для примера рассмотрим фрагмент базы знаний, касающийся области производства ФПФ (рис. 2, б). Пусть в базе знаний определены следующие МК ПрО: «Экспонирующая установка для флексографии», «Цех», «Цех производства ФПФ». Известно, что МК ПрО «Экспонирующая установка для флексографии» является компонентом МК ПрО «Цех производства ФПФ».

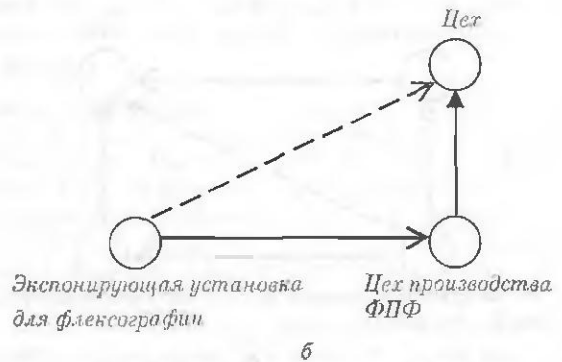
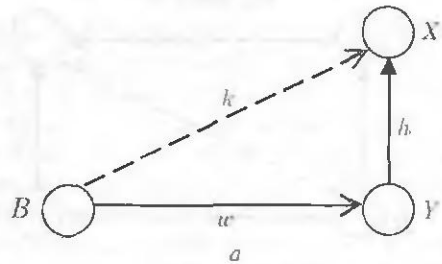


Рис. 2. Определение морфизма связи:

— морфизм наследования;
 — морфизм агрегации; \dashrightarrow — морфизм связи

ФПФ», т. е. существует морфизм агрегации из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Цех производства ФПФ». А МК ПрО «Цех» является обобщением МК ПрО «Цех производства ФПФ», т. е. существует морфизм наследования из МК ПрО «Цех производства ФПФ» в МК ПрО «Цех». Тогда существует морфизм связи из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Цех», который говорит о существовании отношения структурного сходства между данными моделями. Семантика данного отношения заключается в том, что экспонирующая установка для флексографии находится в цехе, поскольку она находится в цехе производства ФПФ.

Рассмотренные примеры проясняют семантику отношений структурного сходства, соответствующих морфизмам связи между МК ПрО. Из данных примеров следует, что, с одной стороны, экспонирующая установка для флексографии находится в цехе, поскольку оборудование находится в цехе, с другой стороны, экспонирующая установка для флексографии находится в цехе, поскольку она находится в цехе производства ФПФ (рис. 3, б). Таким образом, получаем коммутативную диаграмму $dr = k = wh$, где $d: B \rightarrow A$ — морфизм наследования, $r: A \rightarrow X$ — морфизм агрегации, $k: B \rightarrow X$ — морфизм связи, $w: B \rightarrow Y$ — морфизм агрегации, $h: Y \rightarrow X$ — морфизм наследования (рис. 3, а).

Для поддержки коммутативных диаграмм вида $dr = wh$ требуется в рамках итеративной процедуры создания баз знаний в диалоге с пользователем выполнять специальные правила пополнения иерархической базы знаний ПрО. При этом необходимо рассмотреть две стратегии проектирования базы знаний — нисходящую с последовательным уточнением МК ПрО сверху вниз и восходящую с последовательным обобщением МК ПрО снизу вверх.

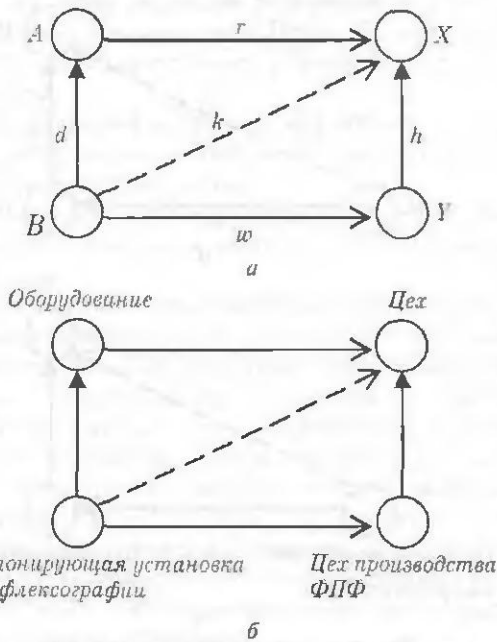


Рис. 3. Фрагмент иерархической структуры знаний

→ — морфизм наследования;

→ — морфизм агрегации; - - -> — морфизм связи

3. Правила пополнения иерархической базы знаний ПрО в случае нисходящей стратегии проектирования

Пусть известно, что некоторая модель концептуализации A является компонентом модели концептуализации X , т. е. существует морфизм агрегации $r: A \rightarrow X$. Предположим, что требуется уточнить знания о ПрО, представляемые моделью концептуализации X . Для этого от концепции X была унаследована новая концепция Y , т. е. должен существовать морфизм наследования $h: Y \rightarrow X$. В этом случае в качестве компонента модели концептуализации Y вместо модели концептуализации A может быть использована только такая модель концептуализации B , при котором существует морфизм наследования $d: B \rightarrow A$, т. е. модель концептуализации B является унаследованной от модели концептуализации A .

Рассмотрим пример (рис. 3, б). Пусть в базе знаний определены следующие МК ПрО: «Оборудование», «Экспонирующая установка для флексографии», «Цех». Известно, что МК ПрО «Оборудование» является компонентом МК ПрО «Цех», а МК ПрО «Экспонирующая установка для флексографии» является унаследованной от МК ПрО «Оборудование». Предположим, что требуется уточнить знания о ПрО, представляемые моделью концептуализации «Цех». Для этого от МК ПрО «Цех» была унаследована новая МК ПрО «Цех производства ФПФ». Цех производства ФПФ содержит промышленное оборудование; точнее, оборудование, которое содержится в цехе производства ФПФ, в том числе является экспонирующей установкой для флексографии [4]. Поэтому вместо МК ПрО «Оборудование» в качестве компонента МК ПрО «Цех производства ФПФ» используется МК ПрО «Экспонирующая установка для флексографии».

Дополнительный анализ показал, что приведенное выше правило не отражает в полной мере семантические закономерности структуры системы знаний о предметной области. Оказывается, что выбор модели концептуализации B в качестве компонента модели концептуализации Y не является случайным. Осуществляя этот выбор, необходимо учитывать функциональное назначение классов объектов, описываемых моделями концептуализации ПрО. Точнее, модель концептуализации B может быть выбрана в качестве компонента модели концептуализации Y , если функция класса объектов, описываемого моделью концептуализации B , является частью функции класса объектов, описываемого моделью концептуализации Y , т. е. $Func(B) \text{ part-of } Func(Y)$. Кроме того, при определении модели концептуализации Y необходимо, чтобы функция класса объектов, описываемого моделью концептуализации X , являлась обобщением функции класса объектов, описываемого моделью концептуализации Y , т. е. $Func(Y) \text{ is-a } Func(X)$.

Сформулированное правило носит эвристический характер и может учитываться только в том случае, если выполняется предыдущее правило. Уже на ранее рассмотренном примере (рис. 3, б) видно, что

не случайно МК ПрО «Экспонирующая установка для флексографии» была выбрана в качестве компонента МК ПрО «Цех производства ФПФ». Функциональным назначением класса объектов, описываемого МК ПрО «Цех производства ФПФ», является производство ФПФ из флексографских фотополимеризующихся пластинок. Функциональным назначением класса объектов, описываемого МК ПрО «Экспонирующая установка для флексографии», является экспонирование флексографских фотополимеризующихся пластинок, что является составной частью производства ФПФ из флексографских фотополимеризующихся пластинок.

Таким образом, разработаны правила пополнения иерархической базы знаний ПрО в случае нисходящей стратегии проектирования базы знаний. Данные правила, по сути, позволяют предсказать морфизмы агрегации между МК ПрО, если определены функции классов объектов, описываемых данными моделями.

4. Правила пополнения иерархической базы знаний ПрО в случае восходящей стратегии проектирования

Пусть известно, что некоторая модель концептуализации B является компонентом модели концептуализации Y , т. е. существует морфизм агрегации $h: B \rightarrow Y$. Предположим, что требуется обобщить знания о ПрО, представляемые моделью концептуализации Y . Для этого была определена новая концепция X , такая, что должен существовать морфизм наследования $h: Y \rightarrow X$. В этом случае в качестве компонента модели концептуализации X вместо модели концептуализации B может быть использована только такая модель концептуализации A , для которой существует морфизм наследования $d: B \rightarrow A$, т. е. модель концептуализации A является обобщением модели концептуализации B .

Рассмотрим пример (рис. 3, б). Пусть в базе знаний определены следующие МК ПрО: «Экспонирующая установка для флексографии», «Цех производства ФПФ», «Оборудование». Известно, что МК ПрО «Экспонирующая установка для флексографии» является компонентом МК ПрО «Цех производства ФПФ». Кроме того, МК ПрО «Оборудование» является обобщением МК ПрО «Экспонирующая установка для флексографии». Предположим, что требуется обобщить знания о ПрО, представляемые моделью концептуализации «Цех производства ФПФ». Для этого была определена МК ПрО «Цех», которая является обобщением МК ПрО «Цех производства ФПФ». Цех не обязательно содержит экспонирующую установку для флексографии, но обязательно содержит промышленное оборудование. Поэтому вместо МК ПрО «Экспонирующая установка для флексографии» в качестве компонента МК ПрО «Цех» используется МК ПрО «Оборудование».

Однако приведенное выше правило не отражает в полной мере семантические закономерности структуры системы знаний о предметной области. Выбор модели концептуализации A в качестве ком-

понента модели концептуализации X не является случайным. Осуществляя этот выбор, необходимо учитывать функциональное назначение классов объектов, описываемых моделями концептуализации ПрО. Точнее, модель концептуализации A может быть выбрана в качестве компонента модели концептуализации X , если функция класса объектов, описываемого моделью концептуализации A , является частью функции класса объектов, описываемого моделью концептуализации X , т. е. $Func(A)$ *part-of* $Func(X)$. Кроме того, при определении модели концептуализации X необходимо, чтобы функция класса объектов, описываемого моделью концептуализации X , являлась обобщением функции класса объектов, описываемого моделью концептуализации Y , т. е. $Func(Y)$ *is-a* $Func(X)$.

Сформулированное правило является эвристическим и может учитываться только в том случае, если выполняется предыдущее правило. На ранее рассмотренном примере (рис. 3, б) видно, что не случайно МК ПрО «Оборудование» была выбрана в качестве компонента МК ПрО «Цех». Функциональным назначением класса объектов, описываемого МК ПрО «Цех», является производство промышленной продукции. Функциональным назначением класса объектов, описываемого МК ПрО «Оборудование», является выполнение технологических операций, что является составной частью производства промышленной продукции.

Таким образом, разработаны правила пополнения иерархической базы знаний ПрО в случае восходящей стратегии проектирования базы знаний. Данные правила, как и предыдущие, позволяют предсказать морфизмы агрегации между МК ПрО, если определены функции классов объектов, описываемых данными моделями.

5. Выводы

В данной работе рассмотрены семантические закономерности иерархических структур систем знаний ПрО и обоснована необходимость их учета для получения качественных баз знаний автоматизированных систем. На этой основе предложены правила пополнения иерархической базы знаний автоматизированной системы. Использование данных правил в рамках процедуры создания базы знаний в диалоге с пользователем позволит проводить проверку корректности базы знаний и упростит процесс ее уточнения.

Список литературы: 1. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. — СПб: Питер, 2000. — 384 с. 2. Smith J., Smith D. Database Abstractions: Aggregations and Generalization // ACM Transactions on Database Systems. — 1977. — Vol. 2. № 2. — P. 105–133. 3. Неофитная Т. М. Разработка категорной модели базы знаний интеллектуальной системы принятия решений // Проблемы библики. — 2003. — № 58. — С. 108–115. 4. Кипчан Г. Энциклопедия по печатным средствам информации. Технологии и способы производства: Пер. с нем. — М.: МГУП, 2003. — 1280 с.

Поступила в редакцию 14.08.2006