

Методика онтологического связывания объектов в автоматизированных системах с использованием классификаторов

Е. С. Джумайло, В. В. Баранюк

Аннотация – Статья посвящена вопросу разработки формальной спецификации концептуализации предметной области автоматизированных систем (АС). В настоящее время автоматизация различных видов деятельности приводит к внедрению различных информационных систем поддержки принятия решений, поисковых систем, а также других систем, основанных на базах знаний. Для разработки любой такой системы, как правило, требуется провести моделирование ее предметной области. Этот процесс является достаточно трудоемким и требует большого опыта и глубоких аналитических знаний в предметной области. В данной статье предлагается методика, позволяющая не только упростить процесс моделирования предметной области АС, но и существенно расширить разрабатываемую модель, благодаря использованию таких информационных ресурсов, как классификаторы, при формализации предметной области АС.

Особое место в информационном обеспечении занимают классификаторы. Они обеспечивают унифицированное представление данных в АС и содержат множество полезной структурированной информации об объектах, представленных в АС. В данной статье предлагается методика, позволяющая описывать определенные сферы предметных областей с помощью представленных в системе классификаторов. Такой подход может использоваться для построения более глубоких моделей предметных областей, что в свою очередь, обеспечит повышение качества информационного обеспечения разрабатываемой АС.

Ключевые слова – автоматизированные системы, информационные ресурсы, классификаторы, онтология, предметная область, моделирование предметной области.

I. ВВЕДЕНИЕ

В настоящее время развитие информационных систем и технологий приводит к масштабной автоматизации различных сфер деятельности человека. Внедрение систем поддержки принятия решений, поисковых систем и других информационных систем, основанных на базах знаний, упрощает работу сотрудников ряда организаций, повышает эффективность информационного взаимодействия систем, а также информационной поддержки определенных видов деятельности.

Разработка любой такой системы начинается, прежде всего, с описания предметной области, в рамках которой она будет функционировать. Процесс описания предметной области требует определенных экспертных навыков и опыта и является достаточно трудоемким.

Для разработки описания предметной области в основном используется два подхода: ER-моделирование и моделирование предметной области с помощью построения онтологического описания.

Метод, основанный на построении онтологии, является преимущественным, поскольку позволяет строить более глубокое формальное описание предметной области на концептуальном уровне.

Информационные ресурсы, циркулирующие в автоматизированной системе, обеспечивают информационную поддержку решаемых в ней задач. Одним из важных информационных ресурсов в АС являются классификаторы. Они обеспечивают единообразное представление данных и качественное информационное взаимодействие между объектами системы.

Поскольку классификаторы содержат множество полезной информации об объектах АС, они могут быть использованы при моделировании ее предметной области.

Для повышения качества информационного обеспечения АС предлагается подход, позволяющий расширить описание предметной области с помощью определенных информационных ресурсов, таких как классификаторы. Это, в свою очередь, также упрощает работу аналитиков по формализации и разработке описания определенной предметной области.

Данная статья описывает методику привязки информационных ресурсов на примере классификаторов к онтологии предметной области АС для формирования более глубокой и подробной модели концептуализации предметной области.

II. ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

В настоящее время для описания предметной области преимущественно используются два подхода: ER-моделирование и формирование онтологического описания как формальной спецификации концептуализации предметной области.

ER-моделирование (от англ. entity-relationship model, модель «сущность — связь») представляет собой процесс, позволяющий описывать концептуальные схемы предметной области на основе ER-модели.

Указанная модель используется при высокоуровневом (концептуальном) моделировании.

ER-модель в основном используется при проектировании баз данных и представляет предметную область в виде объектов и определенных отношений (связей) между ними, а также ограничений. И хотя ER-модель позволяет обеспечить независимость модели данных от средств реализации, создаваемые с ее помощью модели предметной области – это модели со «слабой» семантикой.

Онтология – это попытка всеобъемлющей и подробной формализации некоторой области знаний с помощью концептуальной схемы. Онтология определяется на языке представления знаний, который определяет формальную семантику, что обеспечивает возможность четкой интерпретации онтологии машиной. Онтология представляет предметную область на концептуальном уровне с помощью классов-концептов и их отношений. Эти концепты и отношения могут легко интерпретироваться человеком, поскольку они соответствуют элементам нашей мысленной модели.

Согласно [1] онтология предметной области определяется как кортеж следующего вида:

$$O^d = \langle C, A, R, D \rangle,$$

где C – множество классов предметной области, A – множество атрибутов этих классов, R – отношение частичного порядка на множестве классов, $R \subseteq C \times C$, D – множество областей допустимых значений атрибутов.

Классы определяются тройкой $\langle d_i, F(d_i), V(d_i) \rangle$, где d_i – имя класса, $F(d_i) = \{f_{ij}\}$ – множество признаков класса, $V(d_i) = \{v_{iq}\}$ – объем класса, v_{iq} – объекты, имеющие признаки $F(d_i)$.

Пусть $B(F(d_i)) = \{\emptyset, 2^{|F(d_i)|}\}$ – булеан содержания множества признаков $F(d_i)$ некоторого класса d_i , где $2^{|F(d_i)|}$ – множество всех подмножеств содержания $F(d_i)$. Элементы булеана образуют частично упорядоченное множество по включению его элементов – решетку $(B(F(d_i)) \leq, \wedge, \vee)$.

Любой элемент $F(d_i) \in B(F(d_i))$, $H \in [1, 2^{|F(d_i)|}]$, полученной решетки считается содержанием некоторого класса d_i^H , обобщающего понятие d_i , если $F(d_i^H) \subseteq F(d_i)$.

Решетка $K(d_i) = (B(F(d_i)) \leq, \wedge, \vee)$ всех подмножеств содержания начального класса d_i называется концептуальным «каркасом» онтологии предметной области. Концептуальный «каркас» онтологии предметной области является необходимой структурой, содержащей начальный класс d_i .

Онтология всегда разрабатывается для описания определенной предметной области, и чем она уже, тем более подробно ее можно описать, покрывая более широкий спектр относящихся к ней проблем. Это позволяет в свою очередь формализовать обширную предметную область с помощью нескольких онтологий, представляющих определенные части описываемой предметной области.

Концептуальное моделирование с помощью онтологий кажется очень похожим на разработку ER-

модели для базы данных. Тем не менее, между ними есть определенные различия. Во-первых, онтология обеспечивает более богатую формальную семантику, чем модель баз данных. Она позволяет отражать более сложную аксиоматическую информацию благодаря логическим нотациям. Следовательно, онтология обеспечивает семантически более глубокое представление предметной области. Во-вторых, в отличие от ER-модели, онтология разрабатывается для использования в решении различных задач одновременно. ER-модель используется в основном, для описания схемы базы данных, онтология же описывает предметную область как таковую, что позволяет строить на ее основе поисковые системы, системы поддержки принятия решений, экспертные системы.

Таким образом, можно сделать вывод о том, что метод описания предметной области с помощью онтологии позволяет формализовать исследуемую предметную область более точным ее концептуальным представлением. Одной из наиболее важных областей использования онтологического описания предметной области является применение его в АС для улучшения процессов принятия решений.

Построение онтологий – сложный, требующий достаточно большого времени процесс. Для упрощения данного процесса разработано множество программных средств, предоставляющих пользователю графический интерфейс построения онтологий. Наиболее популярным из таких средств является Protégé, поскольку имеет ряд ключевых особенностей:

- расширяемая модель знаний, позволяющая пользователям переопределять репрезентативные примитивы;
- настраиваемый формат выходного файла для адаптации под любой формальный онтологический язык;
- удобный настраиваемый пользовательский интерфейс;
- мощная архитектура, построенная на плагинах, обеспечивающая интеграцию с различными приложениями.

Поэтому при разработке онтологии в дальнейшем использовалось именно данное средство создания онтологических моделей.

Существуют различные подходы к разработке онтологий, но, как правило, создание онтологии предметной области проходит в несколько этапов:

- 1) определение масштаба и области онтологии;
- 2) определение основных терминов онтологии;
- 3) определение классов и их иерархии;
- 4) определение свойств классов и связей.

Рассмотрим поэтапно процесс создания онтологии на примере поликлиники. Для начала необходимо определить область онтологии и ее масштаб. Пусть необходимо разработать систему поддержки принятия решений для помощи врачам в постановке диагноза. Тогда разрабатываемая онтология должна охватывать лечебную деятельность поликлиники и содержать информацию о врачах, пациентах, болезнях и др.

На следующем шаге необходимо определить основные термины онтологии. Проанализировав

рассматриваемую область, можно определить следующий список терминов, представленный в таблице 1.

На следующем этапе необходимо определить иерархию классов. Вполне очевидно, что в рамках выбранной предметной области, корневым классом будет являться класс «Поликлиника». Далее из описанного ранее списка терминов предметной области можно выделить основные классы будущей онтологии, такие как: медицинский персонал, показатель, пациент, средство лечения, схема лечения. Окончательное распределение классов по иерархии приведено на рисунке 1.

Таблица 1 – Основные термины рассматриваемой онтологии

Поликлиника
Врач
Пациент
Медицинский персонал
Медсестра
Анализ
Процедура
Показатель
Средство лечения
Продукт
Ощущение
Фактор
Пол
Возрастная группа
Прием медикаментов
Состояние
Схема лечения
Симптом
Заболевание
Генетическое нарушение
Лекарство
Образ жизни



Рис. 1 – Иерархия классов онтологии

На следующем шаге, помимо уже определенных связей между классами внутри одной и той же ветки иерархии классов, необходимо описать все остальные связи между концепт-классами. На рисунке 2 представлена полученная онтология лечебной деятельности поликлиники, а все связи описаны в таблице 2.

Таблица 2 – Описание связей онтологии лечебной деятельности поликлиники

№ связи	Название связи
1	Имеет подкласс
2	Выявлено
3	Затрагивает
4	Имеет
5	Использует
6	Лечится
7	Назначает
8	Определяет
9	Отсутствует
10	Присутствует
11	Проводит
12	Рекомендуется

III. КЛАССИФИКАТОРЫ

Описание предметной области представляет собой трудоемкий процесс, требующий длительного времени и глубоких экспертных знаний. Для облегчения этого процесса предлагается подход, заключающийся в том, что отдельные сферы предметной области могут быть описаны с использованием существующих информационных ресурсов, таких, как классификаторы. Другими словами, онтология предметной области может быть расширена путем онтологического связывания классов онтологии с классификаторами.

Согласно [2] под классификатором понимается информационный ресурс, распределяющий информацию в соответствии с ее классификацией (классами, группами, видами и другими признаками).

В зависимости от уровня утверждения существуют классификаторы следующих категорий:

- международные (региональные) классификации;
- общероссийские классификаторы;
- межведомственные классификаторы;
- ведомственные классификаторы;

– системные классификаторы.

Международная (региональная) классификация – классификация, принятая международной (региональной) организацией [3].

Общероссийский классификатор – информационный ресурс, признанный общероссийским классификатором в соответствии с законодательством Российской Федерации в области стандартизации [2].

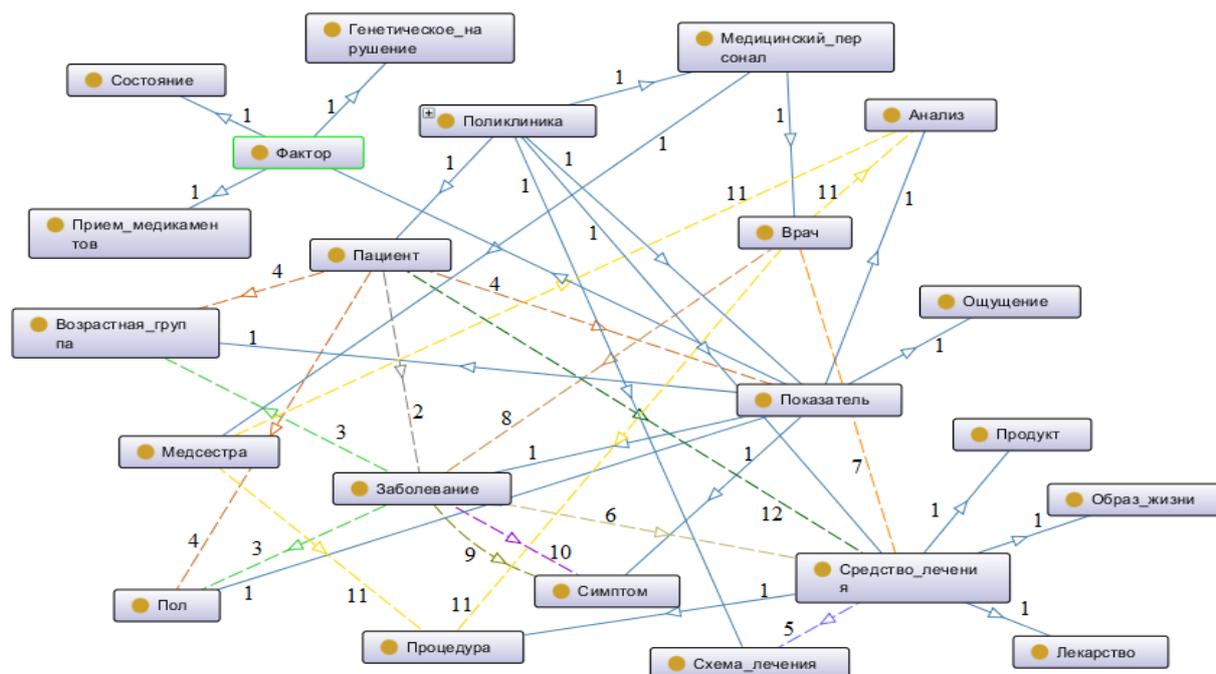


Рис. 2 – Онтология лечебной деятельности поликлиники

Под межведомственным классификатором понимается информационный ресурс, консолидирующий межведомственную информацию для обеспечения межведомственного взаимодействия.

Ведомственный классификатор утверждается министерством (ведомством) для применения в рамках данного министерства (ведомства). Ведомственный классификатор разрабатывается на специфические виды информации, циркулирующей внутри ведомства и необходимой при решении его функциональных задач.

Под системным классификатором понимается классификатор, утвержденный Главным конструктором автоматизированной системы и принятый для применения в рамках данной системы. Системные классификаторы содержат информацию, необходимую для решения задач в конкретной автоматизированной системе и отсутствующую в национальных и ведомственных классификаторах.

Таким образом, можно сказать, что в классификаторах представлены систематизированные описания объектов предметной области, в которых наименования объектов имеют определенные коды.

Расширение онтологии предметной области данными классификаторами может повысить эффективность поисковой системы, улучшить поддержку принятия

решений, а также упростить аналитическую деятельность.

На представленной на рисунке 2 онтологии можно увидеть, что некоторые ее классы можно связать с целым множеством объектов, представленных в различных классификаторах от международных до системных.

Так, например, класс «Заболевание» можно связать с международной классификацией болезней. Данная классификация используется по всему миру для организации и кодирования медико-санитарной информации о заболеваемости и использования этих данных для различного рода статистических исследований и эпидемиологии, для управления здравоохранением и исследованиями, для профилактики и лечения заболеваний.

Класс «Лекарство» может быть связан со справочником лекарственных препаратов, который содержит сведения о различных лекарственных средствах, что может быть крайне необходимо при лечении, а также ведении учета медицинских препаратов и сбора статистических данных.

Класс «Процедура» непосредственно связан с внутренним системным перечнем медицинских услуг, предоставляемых исследуемым медицинским учреждением.

Представленная онтология, дополненная описанными выше классификаторами, может быть использована для автоматизации медицинской деятельности поликлиники, в том числе: для разработки системы поддержки принятия решений врачами при постановке диагнозов и проведении лечения; системы регистрации пациентов и записи на прием; системы сбора и анализа статистических данных деятельности медицинского учреждения и др.

IV. МЕТОДИКА ОНТОЛОГИЧЕСКОГО СВЯЗЫВАНИЯ

Для привязки классификаторов к онтологии предметной области предлагается методика, состоящая из последовательности обязательных действий (процедур), выполняемых автоматизировано. Далее представлены действия в соответствующей последовательности.

1) Представление исследуемой онтологии в машинно-интерпретируемом виде. Перевод имеющейся онтологии в формат owl.

2) Загрузка системы классификаторов в базу данных классификаторов.

3) Выбор классификатора для привязки к онтологии. Из базы данных классификаторов выбирается очередной классификатор.

4) Выделение из метаописания классификатора его наименования. Анализ метаописания классификатора с целью выбора его полного наименования.

5) Семантический анализ по наименованию. Поиск с помощью блока семантического анализа по наименованию классификатора, семантически идентичного объекта в онтологии.

6) Привязка классификатора. В случае однозначного соответствия объекта онтологии и классификатора выполняется:

а) создание объекта в онтологии, соответствующего данному классификатору;

б) выделение метаданных из классификатора;

в) запись метаданных классификатора в свойства объекта, созданного в онтологии.

7) Выбор наиболее подходящего объекта в онтологии. В случае нахождения в онтологии нескольких семантически близких к классификатору объектов, оператор выбирает наиболее подходящий объект из предложенных вариантов.

8) Привязка классификатора ручным способом. В случае если не удалось найти семантически схожий объект, оператор определяет место классификатора в

онтологии по следующему принципу:

а) выделение на первом уровне онтологии ветки, наиболее подходящей по смыслу к конкретному классификатору;

б) поиск в выбранной ветке соответствующего классификатору объекта;

в) в случае, если подходящий в онтологии объект найден – выбор его для привязки классификатора. В противном случае переход в онтологии на уровень ниже и повторение действий, описанных в пунктах «б» и «в».

V. ВЫВОДЫ

В статье предложена методика расширения онтологического описания предметной области автоматизированной системы с помощью классификаторов. Использование методики не только облегчает работу аналитиков при разработке формального описания предметной области АС, но также помогает строить модели предметной области на более глубоком концептуальном уровне.

Классификаторы являются важной составляющей АС и содержат разнообразную информацию для описания предметной области, которая может использоваться при моделировании.

Предлагаемая методика может быть полезна при разработке автоматизированных (информационных) систем, систем поддержки принятия решений, экспертных систем, основанных на базах знаний и требующих широких концептуальных описаний предметных областей.

БИБЛИОГРАФИЯ

- [1] Кулинич А.А. Концептуальные «каркасы» онтологий в поддержке принятия решений в условиях неопределенности. [Электронный ресурс]. URL:<http://raai.org/resurs/papers/kii-2012/doclad/Kulinich.doc> (дата обращения: 27.03.2018).
- [2] Постановление Правительства Российской Федерации от 1 июня 2016 г. №487 «О первоочередных мерах, направленных на создание государственной информационной системы «Единая информационная среда в сфере систематизации и кодирования информации».
- [3] ПР 50.1.024–2005. Основные положения и порядок проведения работ по разработке, ведению и применению общероссийских классификаторов. – М.: Стандартинформ, 2006. – 34 с.

Development of methods for ontological binding of objects in automated systems using classifiers

E. S. Dzhumaylo, V. V. Baranyuk

Abstract – This paper analyzes knowledge representation approaches and is devoted to the problem of developing formal conceptual knowledge representation model for automated systems. A development of any automated system begins from building its knowledge representation model. A process of building such a model is rather complicated and requires special skills and experience. Methods for simplifying knowledge representation modeling and expanding developing model are proposed. The paper considers the possibility of using presented in automated systems information resources, such as classifiers while knowledge representation model developing. Information resources are the important components of automated systems. These resources also contain a lot of useful information about the domain knowledge of an automated system, so they can be useful for building domain knowledge model. Classifiers are one of the important informational resources. They provide the uniform data representation and the qualitative information interaction between the components of the system. Classifiers contain a lot of useful information about the objects of an automated system, thus they can be used for building a knowledge representation model.

Keywords – automated system, information resource, classifier, ontology, domain knowledge, knowledge representation model.