

ранговой степени $R(t_k, c)$ для каждого из выбранных термов является предметом отдельного исследования и будет освещен автором в отдельной работе.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Блюмин С.Л., Шуйкова И.А., Сараев П.В., Черпаков И.В. Нечеткая логика: алгебраические основы и приложения: Монография. – Липецк: ЛЭГИ, 2002. – 113 с.
2. Ха Т.Ч., Юрчук С.Н. Создание текстовой выборки на основе электронного архива данных лаборатории ELDIC для исследования задач автоматической обработки текстов на естественном языке // Труды Всероссийской научной школы-семинар молодых ученых, аспирантов и студентов Таганрог: "Интеллектуализация информационного поиска, скантехнологии и электронные библиотеки". – Таганрог: Изд-во ТТИ ЮФУ, 2008. – С. 82-86.
3. Igor Kuralenok, Vladimir Dobrynin, Igor Nekrestyanov, Mikhail Bessonov, and Ahmed Patel. Distributed search in topic-oriented document collections. In Proc. of World Multiconference on Systematics, Cybernetics and Informatics (SCI'99), volume 4, pp. 377-383, Orlando, Florida, USA, August 1999.
4. Dumais S.T., Platt J., Heckerman D., Sahami M. Inductive learning algorithms and representations for text categorization. In Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management (Bethesda, MD, 1998), 148-155.
5. Haris Z. Mathematical Structures of Language. Interscience Publishers John Wiley & Sons, New York. – 1968, 80. – 230 p.
6. McCallum A., Nigam K. A Comparison of Event Models for Naive Bayes Text Classification. In AAAI/ICML-98 Workshop on Learning for Text Categorization, 1998. p. 41-48.
7. Sebastiani F. Machine Learning in Automated Text Categorization. ACM Computing Surveys, 34(1): 1-47, 2002.

УДК 517.714.3

Г.В. Уралев

КОНЦЕПТУАЛЬНАЯ СПЕЦИФИКАЦИЯ НЕКОТОРЫХ ЗАДАЧ ИНФОРМАТИКИ*

Введение. Здесь под *задачей информатики* мы понимаем задачу, решение которой достигается с помощью программ, использующих экспертное знание. Важным этапом в разработке таких программ является *концептуальная спецификация* структуры этого знания, в частности, знания предметной области и методов решения задачи. Концептуальная спецификация состоит в идентификации соответствующих понятий предметной области, связей между ними, схем, правил, процедур и т.п.

Часто концептуальная спецификация выполняется на неформальном уровне. Но тенденция такова, что все чаще используются формальные языки для концептуальной спецификации, благодаря которым удается строить концептуальные модели приложений, допускающие формальную интерпретацию и, следовательно, машинную обработку. Процесс построения таких моделей называют *концептуальным моделированием* [1]. Формализмы концептуальной спецификации называются *концептуальными языками* или *языками концептуального моделирования* [2]).

Одним из важных требований, предъявляемых к концептуальным языкам, является их когнитивная адекватность, т.е. близость их конструкций к тем, какие использует эксперт, выполняющий неформальную концептуализацию. Так как такая

* Работа выполнена при финансовой поддержке РФФИ (грант № 08-01-00465).

концептуализация выполняется с использованием понятий, то концептуальные языки должны быть понятийно-ориентированными. В частности, «нормальный» концептуальный язык для спецификации структуры понятий должен содержать основные конструкторы понятий (агрегацию, таксономию, конкретизацию, обобщение и т.п.)

Формальное описание предметной области, выполненное в концептуальном языке, называется *концептуальной схемой*. Концептуальные схемы называют также *онтологиями*.

В работах Г.С. Плесневича была предложена система концептуальных языков, названная «Бинарная Модель Знаний» (БМЗ) [4, 5]. С помощью предложений этих языков можно определять формальные понятия моделируемой предметной области, которые могут быть двух типов – *классы* и *бинарные связи*. В БМЗ имеются следующие языки:

- ◆ Язык структурной спецификации (ЯСС). С помощью предложений этого языка определяется атрибутивная структура объектов – элементов универсума формального понятия, а также определяется таксономия понятий.
- ◆ Язык для определения таксономической структуры понятий (ЯТС). С помощью предложений этого языка определяется иерархическая сеть понятий и указываются дефолты для атрибутов.
- ◆ Язык запросов к базам фактов (ЯЗ), структурированных по схеме, составленной из предложений языков ЯСС и ЯТС.
- ◆ Язык манипуляции фактами (ЯМФ). С помощью предложений этого языка можно вводить в базу фактов новые факты, исключать факты и модифицировать факты.
- ◆ Язык спецификации типы данных (ЯСТД). С помощью предложений этого языка определяются абстрактные типы данных, которые могут служить значениями атрибутов понятий.
- ◆ Язык логической спецификации (ЯЛС). С помощью предложений этого языка определяются логические связи между понятиями.
- ◆ Язык модальной спецификации (ЯМС). С помощью предложений этого языка определяется модальные связи между понятиями.
- ◆ Язык темпоральной спецификации (ЯТС). С помощью предложений этого языка специфицируются понятия, содержащие неявно атрибуты, значениями которых служат моменты или интервалы времени, и определяющие темпоральные отношения между такими атрибутами.

Приведем примеры предложений из языков системы БМЗ.

Примеры.

1) Предложения ЯСС:

Преподаватель[ФИО: String, Должн: String, Зарплата: Integer,
Работает_на: Кафедра]

% Преподаватель имеет фамилию имя и отчество (представляется строчкой символов), занимает некоторую должность (представляется строчкой символов), работает на некоторой кафедре. Вообще, предложение ЯСС, специфицирующее универсум класса имеет вид <имя понятия> [список атрибутивных пар]%

Студент[ФИО: String, Группа: String, Руководитель: Преподаватель],
(Студент СдалЭкзамен Предмет)[ВремяСдачи: Date, Оценка: Integer,
Принял: Препод].

% Студент сдал экзамен по некоторому предмету в определенную дату и получил некоторую оценку, причем экзамен принял некоторый преподаватель. Вообще, предложение ЯСС, специфицирующее универсум бинарной связи, имеет вид (<имя понятия> <имя бинарной связи> <имя понятия> [<список атрибутивных пар>] %

Кафедра[Название: String, ЗавКафедрой: Преподаватель],
Предмет[Название: String].

2) Предложения ЯЗ:

?X.ФИО – X СдалЭкзамен(Оценка = 5; Принял.(ФИО = ‘И.И. Иванов’;
Должность = профессор) Предмет(Название = матлогика).

% Найти всех студентов (точнее, их ФИО) группы А13-02, которые сдали на
отлично профессору И.И. Иванову экзамен по математической логике %.

?(X.ФИО, LLIST Y.ФИО) – X.Руководитель = Y; Y. Работает_на. На
звание = ПМ.

% Для каждого преподавателя кафедры прикладной математики составить
список руководимых им студентов %

3) Предложения ЯЛС:

EACH Препод.(Должность = профессор OR доцент; Работает_на.Название =
ПМ) Руководит SOME Студент; (X Руководит Y) IFF Y.Руководитель = X.

% Каждый профессор или доцент кафедры прикладной математики руково-
дит хотя бы одним студентом %

IF X IN Преподаватель.Должность = профессор) THEN X.Зарплата > 10000.

% Зарплата профессора больше 10000 рублей %

В ЯТС имеются предложения, относящиеся к интервальной логике Аллена
[6]. Декларация события C выполняется предложением $EVENT C$. Базовые отно-
шения между событиями C и D выражаются следующими предложениями: C
 $BEFORE D$ (раньше), $C MEETS D$ (встречает), $C OVERLAPS D$ (перекрывает), C
 $FINISHES D$ (заканчивает), $C STARTS D$ (начинает), $C DURING D$ (в течение), C
 $EQUAL D$ (равно). Обратные отношения обозначаются с помощью приставки i .
Например, предложение $C i-BEFORE D$ эквивалентно предложению $D BEFORE C$.

Дизъюнкция предложений с одинаковыми парами событий обозначается так,
как обозначается конечное множество. Например, дизъюнкция трех предложений
 $C BEFORE D$, $C i-MEETS D$ и $C EQUAL D$ обозначается так: $C \{BEFORE, i-MEETS,$
 $EQUAL\} D$.

В настоящей работе мы показываем возможность применения БМЗ к двум за-
дачам:

- ◆ проектирование потоков работ с учетом динамики требований и целей
[7]. Точнее, БМЗ может применяться в качестве инструмента для созда-
ния онтологий, которые могут быть использованы для моделирования по-
ведения системы потока работ. Такие онтологии могут быть использованы
при разработке семантических веб-сервисов для бизнес-процессов;
- ◆ создание медицинской онтологии, используемой для диагноза вирусных
гепатитов.

Онтологии для потока работ. На самом верхнем уровне поток работ пред-
ставляется ориентированным графом, вершинами которого являются задачи (или
действия, активности), а дугами служат бинарные связи между задачами. Этот
уровень специфицируется следующими двумя структурными предложениями:

ПотокРабот[Задачи: Задача(*), НачальнаяЗадача: Задача,

КонечнаяЗадача:Задача], (1)

(Задача Связь_с Задача(*)). (2)

В предложении (1) значением атрибута НачальнаяЗадача служит задача в
данном потоке работ, которая должна выполняться первой. Класс Задача может
иметь ряд атрибутов, определяемых условиями моделируемой предметной обла-
сти. Например, может быть включен атрибут, определяющий длительность выпол-
нения задачи, атрибут, указывающий ее директивные сроки, дедлайн и т.п. Дан-

ная задача может в дальнейшем уточняться в терминах ее подзадач, совокупность которых составляет свой поток данных. Это обстоятельство можно формально представить структурным предложением:

Задача[Уточнение: ПотокЗадач].

Бинарная связь Связь_с имеет атрибут ТипСвязи, значениями которого, например, могут быть управляющая, синхронизирующая, ресурсная и т.п. Значения атрибута ТипСвязи определяют соответствующие понятия. Например, понятие УправляющаяСвязь специфицируется с помощью предложения:

УправляющаяСвязь = Связь_с(ТипСвязи = управляющая).

Это понятие имеет атрибут Паттерн, значениями которого служат следование, параллельное_расщепление, исключаящий_выбор, простое_слияние и др.

Рассмотрим следующий типичный сценарий, возникающий при концептуальном моделировании веб-сервиса для организации, обслуживающей покупателей. Запросы покупателей принимаются веб-сервисом, который обрабатывает их с помощью подходящих софт-агентов. Софт-агенты имитируют работу экспертов в данной области, в результате чего софт-агенты формируют план обслуживания. Этот план включает тесты и другие действия по обработке запросов. Затем план исполняется, и его результаты компилируются в соответствующий отчет.

Предположим, что при обработке конкретного запроса формируются план (обозначаемый именем класса План), содержащий 5 диагностических тестов: Тест1, Тест2, Тест3, Тест4 и Тест5. Имена Тестj мы рассматриваем как имена классов, каждый из которых может иметь свои атрибуты. Более того, естественно считать их именами событий. Поэтому в онтологию мы записываем предложения:

Event(Тест), Тестj ISA Тест (j = 1,2,3,4,5). Кроме того, класс План включает атрибуты Входной_запрос и Компиляция_плана и компоненту Ситуация_доступа. Структуру класса План специфицируем при помощи предложения

План[Входной_запрос: Запрос, Ситуация_доступа,
Компиляция_плана: Отчет].

Атрибут Компиляция_плана является вычислимым, т.е. представляется в языке ЯСТД для спецификации абстрактных типов данных. С другой стороны, этот атрибут мы рассматриваем как событие, т.е. как класс, с которым ассоциирован временной интервал. Точно так же, атрибут Входной_запрос и компонента Ситуация_доступа рассматриваются как события.

Между событиями в сценарии существуют бинарные связи, выражаемые отношениями алгебры Аллена. Предположим, что эти связи (в обозначениях Аллена) указаны в следующем графе отношений между событиями (рис. 1).

В языке БМЗ темпоральных отношений мы записываем отношения между этими событиями с помощью следующих предложений:

(Входной_запрос {BEFORE, MEET, OVERL} Ситуация_доступа),
(Ситуация_доступа {BEFORE, MEET, OVERL} Тест1),
(Ситуация_доступа {BEFORE, MEET, OVERL} Тест4),
(Тест1 {BEFORE, MEET, OVERL, i-DURING, i-FINISH*} Тест2),
(Тест2 {BEFORE, MEET} Тест3), (Тест2 {BEFORE} Компиляция_плана),
(Тест3 {BEFORE, MEET} Компиляция_плана),
(Компиляция_плана {BEFORE, MEET} Тест5),(Тест4 {START, MEET} Тест5).

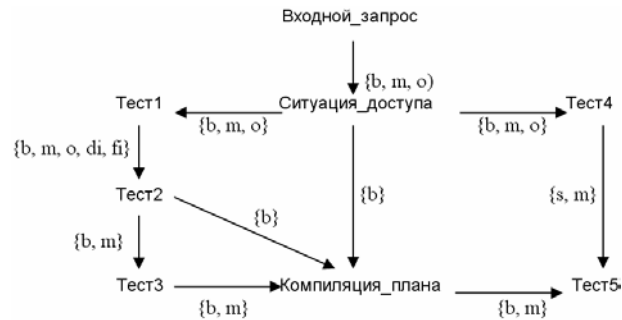


Рис. 1. Граф отношений между событиями

Онтологии медицинского знания. Мы покажем, как БМЗ можно использовать для построения медицинских онтологий, взяв пример, связанный с задачей диагноза гепатита В, на основе выявления антигенов и антител [8].

Для диагноза гепатитов выполняются серологические исследования, при которых могут быть обнаружены антитела к вирусам гепатита. Первая фаза после заражения вирусом гепатита В – инкубационный период. К концу инкубационного периода обнаруживаются определенные антитела, которые указывают на начало острой фазы заболевания. Затем наступает период выздоровления, при котором происходит сероконверсия антигенов вируса гепатита в антитела. Последний период – фаза выработки иммунитета, характеризуемого положительным тестом на антитела. Возможны различные течения гепатита. Эти течения описываются как курсы, которые представляются в виде системы событий, определяемых появлением и исчезновением определенных антигенов и антител. Имеется 4 основных курса, анализ которых показывает, что (1) количественная темпоральная информация (длины временных интервалов событий) представляется средними значениями, тогда как, индивидуальные значения могут варьировать от случая к случаю. Но качественная темпоральная информация (например, что данное антитело появляется в курсе раньше другого) является устойчивой; (2) все курсы выглядят одинаково с точки зрения серологических данных; они различаются порядком их появления. Это подсказывает, что в формальную модель (онтологию) для представления медицинского знания о течении болезней, следует включать темпоральные отношения.

Введем события Инкубация, Острый_период, Выздоровление, Иммунизация, представляющие фазы течения болезни. Тогда эти события представляются следующими предложениями: event(Инкубация), event(Острый_период), event(Выздоровление), event(Иммунизация). Гепатит В возникает при инфицировании вирусом HBV. Следующие предложения представляют некоторое общее знание, касающееся гепатита В:

```

Гепатит = Внутренняя_болезнь(Место_поражение = Печень)
Вирусный_гепатит = Гепатит(Этиология.Фактор = Вирус),
Вирус[Тип:String, Антигены:Антиген(*)],
(Антиген Сероконверсия Антитело),
Гепатит_В = Вирусный_гепатит(Вирус.Тип = 'HBV').
  
```

При рутинном серологическом исследовании на диагноз гепатита В проводятся тесты на:

- ◆ поверхностный антиген HBsAg и антитело к нему anti-HBs;
- ◆ оболочный антиген HBeAg и антитело к нему anti-HBe;
- ◆ антитела anti-HBc IgN и anti-HBc к ядерному антигену HBc.

Обозначим через $C\langle HBsAg \rangle$ событие, во время которого существует антиген при тесте на антиген $HBsAg$. Подобные обозначения сделаем для событий, определяемых другими антигенами и антителами. Декларируя все эти события в онтологии, записываем следующие предложения: $event(C\langle HBsAg \rangle)$, $event(C\langle anti-HBs \rangle)$, $event(C\langle HBeAg \rangle)$, $event(C\langle anti-HBe \rangle)$, $event(C\langle Igm_anti-HBc \rangle)$, $event(C\langle anti-HBc \rangle)$.

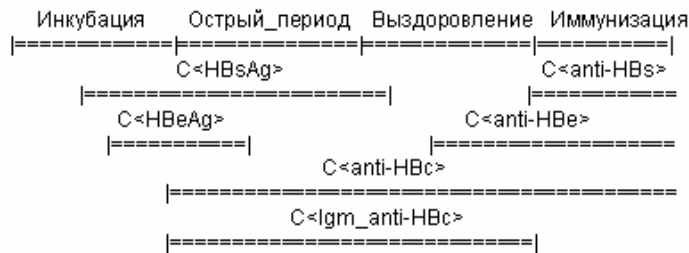
Введем классы, означающие обнаружение антигена и антитела:

ОбнаружАнтигена = $C\langle anti-HBs \rangle \mid C\langle anti-HBe \mid C\langle anti-HBc \rangle \mid C\langle Igm_anti-HBc \rangle$,

ОбнаружАнтитела = $C\langle HBsAg \rangle \mid C\langle HBs \rangle \mid C\langle HBeAg \rangle$

Гепатит В имеет четыре основных типа течения.

Инкубация Острый_период Выздоровление Иммунизация
 Инкубация Острый_период Выздоровление Иммунизация



Информацию о типе 1 течения гепатита В можно представить в онтологии следующими предложениями:

(Инкубация meets Острый_период), (Инкубация overlaps C<HBsAg>),
 (Инкубация meets C<anti-HBc>), (Инкубация meets C<Igm_anti-HBc>),
 (Острый_период meets Выздоровление),
 (Острый_период starts C<anti-HBc>),
 (Острый_период starts C<Igm_anti-HBc>),
 (Острый_период i-overlaps C<HBeAg>), (Иммунизация overlaps C<anti-HBs>),
 (Иммунизация during C<anti-HBe>),
 (Иммунизация during C<anti-HBc>),
 (Иммунизация i-overlaps C<Igm_anti-HBc>),
 (Выздоровление meets Иммунизация),
 (Выздоровление i-overlaps C<HBsAg>), (C<HBeAg> meets C<anti-HBe>),

Замечание 1. Так как в диаграмму интервалов для типа 1 входят 10 событий, то имеется $10^2 = 100$ предложений, выражающих базовые темпоральные отношения между этими событиями. Однако вышеуказанные 14 предложений составляют минимальное множество предложений, из которых остальные 86 предложений являются логическими следствиями этих 14 предложений.

Аналогичные минимальные системы предложений можно указать для типов 2, 3 и 4. Из минимальных систем предложений для указанных четырех типов течений гепатита В можно получить обобщенный тип течения болезни. Этот тип представляется следующими предложениями:

(Инкубация meets Острый_период), (Инкубация overlaps C<HBsAg>),
 (Инкубация meets C<anti-HBc>), (Инкубация meets C<Igm_anti-HBc>),
 (Острый_период meets Выздоровление),
 (Острый_период starts C<anti-HBc>),
 (Острый_период starts C<Igm_anti-HBc>),
 (Острый_период i-overlaps, C<HBeAg>),
 (Иммунизация {overlaps, starts, i-overlaps} C<anti-HBs>),
 (Иммунизация {during, i-overlaps} C<anti-HBe>),
 (Иммунизация {during, i-before} C<anti-HBc>),
 (Иммунизация {i-overlaps, i-meets, i-before} C<Igm_anti-HBc>),
 (Выздоровление meets Иммунизация),
 (Выздоровление {i-overlaps, i-before} C<HBsAg>).

Замечание 2. Интерпретатор БМЗ содержит компоненту, выполняющую обобщение темпоральных предложений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Advances in Conceptual Modeling. Challenges and Opportunities (ER 2008 workshops, Barcelona, Spain, October 2008). – LNCS, vol. 5232, 2008. – 458 p.
2. Allen J.F. Maintaining knowledge about temporal intervals. Communications of the ACM. – v. 26, no. 11. 1983.
3. <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/menzies/>.
4. G.S. Plesniewicz. Binary Data and Knowledge Model. Knowledge-based Software Engineering (Proc. of the 6th JCBSE). – IOS, 2004. – pp. 237-244.
5. Плесневич Г.С. Бинарная модель знаний // Сб. трудов 3-й Международной конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (Колмна, май 2005 г.). – М.: Физматлит, 2005. – С. 88-93.

6. *Г.С. Плещевич*. Бинарные модели знаний // Труды Международных научно-технических конференций «Интеллектуальные системы», «Интеллектуальные САПР» (AIS'08, CAD-2008). – М.: Физматлит, 2008. – С. 135-14.
7. *P. Lawrence* (ed.) *Workflow Handbook: 1997, Workflow Management*. – John Wiley and Sons, N.Y., 1997. – 561 p.
8. *K.-P. Adlassnig, Horak W.* Development and retrospective evaluation of HEPAXPERT-1: a routinely-used expert system for interpretive analysis of hepatitis A and B serological findings. *Artificial Intelligence in Medicine*, 1995, v. 7, pp. 1-24.