

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ БАЗЫ ЗНАНИЙ ИНТЕЛЛЕКТУАЛЬНОЙ МЕДИЦИНСКОЙ СИСТЕМЫ

Ю.А.Прокопчук

Институт технической механики НАН и НКА Украины,
Украинский государственный химико-технологический университет,
Днепропетровский областной диагностический центр

49600, Днепропетровск ГСП-5, Лешко-Попеля 15, ИТМ НАНУ (0562) 47-25-10, Yury@rdc.dp.ua

У статті розглядаються питання побудови концептуальної моделі бази знань інтелектуальної медичної системи. Кінцева мета проекту полягає у створенні інтелектуальної системи "Віртуальний лікар" яка може використовуватися в медичній практиці та учбовому процесі.

In the paper the conceptual model of knowledge base of intellectual medical system is considered. The ultimate goal of the project consists in creation of system "the Virtual doctor".

Основные принципы построения модели

Прежде чем перейти к построению концептуальной модели определенной части Базы Знаний интеллектуальной медицинской системы (ИМС) [1-5] дадим общую характеристику "знания о предметной области".

Под субъектом предметной области (ПрО) будем понимать любую систему S , способную реализовать некоторый проект из ПрО. Примеры систем: ООН, государство, научное сообщество, технопарк, организация, интеллектуальная компьютерная система и т.д. Жизненный цикл существования системы может ограничиваться одним единственным проектом. Очевидно также, чем сложнее проект, тем сложнее должна быть система, которая его реализует. Любая сложная система состоит из элементов и подсистем. Следовательно, сложный проект (глобальный проект) должен быть преобразован в иерархию проектов более низкого уровня, реализуемых элементами системы или подсистемами. Под "проектом" будем понимать как глобальный проект, так и всю совокупность подчиненных проектов.

Понятие "проект" может включать в себя простейшую школьную задачу, написание/чтение статьи или книги, поход в магазин за покупками, создание компьютерной программы, организацию лечения больного, разработку сложной математической модели, изучение материалов других проектов, проект полета человека на Марс и т.д. Новый проект может открываться в момент заключения хозяйственного, страхового, научно-технического или иного договора. Обязательным для любого проекта является наличие осознанных целей, этапа планирования, а, значит, решения задач выбора, необходимых ресурсов и обозримого срока выполнения. Цели проекта задают, в свою очередь, критерии "успешности" реализации проекта. Критерии включают в себя также и ограничения, накладываемые на процесс реализации проекта. Для конкретности будем рассматривать только те проекты, которые реализованы какой-либо системой S и ее элементами (за весь срок их существования) или включены в план работ и/или исследований системы S (как известно, полет человека на Марс запланирован NASA).

Судить о наличии знания и о глубине знания системой ПрО будем, в общем случае, по способности системы реализовать тот или иной проект данной ПрО. Более того, будем считать, что две системы обладают эквивалентным знанием о ПрО, если они способны потенциально реализовать одно и то же множество проектов. Связывание знания с проектом позволяет обеспечить необходимую целостность и системность знания. Однако нас интересует, главным образом, возможность создания эквивалентного знанию системы S "машинного знания" (в приведенном выше смысле), как основы работы интеллектуальной системы, поэтому необходимо выделить те компоненты знания, которые могут служить прообразом "машинного знания".

Под базовой информационной задачей σ_{PS} в рамках проекта P , реализуемого системой S , будем понимать задачу разработки максимально полного информационно – документального обеспечения проекта. Без потери общности будем считать, что базовая информационная задача решается также системой S .

Необходимо отметить, что задача разработки информационно – документального обеспечения в свою очередь также является проектом. В ряде случаев может существовать стандарт (предприятия, отраслевой и т.д.) информационно – документального обеспечения проекта, что значительно облегчает решение базовой задачи.

Под документированием некоторого образа будем понимать фиксацию проекций образа на одну или несколько шкал в цифровом виде, которые содержат значимые для проекта характеристики образа. В частности, это могут быть электронные публикации и описания проектов, компьютерные программы, цифровые аудио и видео материалы и т.п. Документирование предполагает, во-первых, осмысленность образа, процесса (т.е. выделение смысловых и логических элементов и этапов), во-вторых, наличие естественного или формального языка описания образа, процесса и, в третьих, наличие средств для создания цифровой проекции

образа. Например, одна система (больница) может иметь томограф, а другая нет, следовательно, первая система обладает большими возможностями документирования хода лечебно-диагностического процесса. Как правило, расширение возможностей документирования означает более глубокое понимание процессов, протекающих в рамках проекта, и, следовательно, ведет к более качественному выполнению проекта.

Документальное обеспечение проекта должно отражать все этапы жизненного цикла проекта, например: техническое задание, эскизный проект, технический проект, рабочий проект, любые инструкции, чертежи, схемы, описания математических моделей, вычислительных алгоритмов и т.д. Отображение (документальное) информационных и технологических процессов не обязательно должно быть тождественным первичным процессам (чаще всего это принципиально невозможно). Главное, чтобы были отображены существенные для проекта механизмы.

Информационное обеспечение включает в себя все незадокументированные (в определенном выше смысле) информационные образы процессов, протекающих в рамках проекта, которые, однако, могут быть потенциально задокументированы. Необходимо также отметить, что подготовка и реализация некоторых проектов ПрО может вообще не предполагать какого-либо документального сопровождения, однако всегда имеет место информационное сопровождение (целеполагание, планирование и т.д.).

Вербальным знанием системы S о ПрО будем называть совокупность сведений, позволяющих максимально полно задокументировать базовые информационные задачи всех реализованных и реализуемых системой проектов ПрО. Вербальные знания должны быть доступны всем элементам системы. Благодаря вербальному знанию различные системы или элементы систем могут обмениваться знаниями между собой.

Невербальным знанием системы S будем называть информационное отображение процессов, способствующих реализации некоторых проектов ПрО, но не поддающихся документированию системой S. Примером таких процессов могут служить процессы, протекающие на подсознательном уровне человека. Невербальные знания могут образоваться и в случае, если какой-либо элемент системы выполняет свою часть проекта с грифом "секретно". Следует отметить, что в результате приобретения новых знаний невербальные знания системы S могут перейти в разряд вербальных. Наличие невербального знания приводит к образованию "белых пятен" при документировании базовых информационных задач проектов.

Таким образом, знание системы S о ПрО является результатом объединения вербального и невербального знания. Реализация нового проекта приводит к появлению нового знания у системы S. Знание сложной системы зависит от совокупности знаний ее элементов.

Будем считать, что вербальное знание любой системы S может быть представлено в виде "машинного знания", соответствующей интеллектуальной системы. Однако, в общем случае, машинного знания недостаточно для реализации проектов, аналогичных тем, которые реализует система S.

Интегральным знанием о ПрО будем называть гипотетическое знание, которое образовалось бы в результате объединения всех субъектов ПрО в единую систему. Подчеркнем, что интегральное знание могло бы проявить себя только при реализации каких-либо проектов ПрО.

В общем случае, никакая система не обладает интегральным знанием о ПрО, в частности, ввиду того, что отдельно взятая система обладает ограниченными ресурсами для реализации проектов (проекты, требующие больших ресурсов, не включаются в план работ системы, а значит, не образуются информационно – документальное обеспечение, составляющее основу вербальных знаний системы о ПрО). Отметим также, что существует, как правило, множество механизмов (способов, алгоритмов) реализации проекта и каждый механизм предполагает свои ресурсы. При одних ресурсах система может реализовать проект, т.е. обладает соответствующим знанием технологии решения (информационно-документальным обеспечением), а при других нет. Более того, один и тот же механизм реализации проекта у разных систем требует в общем случае разных ресурсов.

В качестве ресурсов могут быть задействованы возможности других систем. В пределе, при наличии, например, достаточных финансовых средств некоторый субъект ПрО может использовать в качестве ресурсов возможности всех остальных субъектов ПрО и, таким образом, система - "субъект ПрО + доступный ресурс" будет обладать интегральным знанием. Примерно такая ситуация имеет место, когда специалист использует мощный интегральный пакет прикладных программ, например, MATLAB. Действительно, в этом пакете и аналогичных пакетах сконцентрированы знания сотен и тысяч субъектов ПрО, что позволяет говорить о некотором приближении к интегральному знанию человеко-машинной системы "специалист + ЭВМ + пакет MATLAB". Будем считать, что именно интеллектуальные информационные системы позволят отдельному субъекту ПрО в максимальной степени приблизиться к интегральному знанию о ПрО.

Проведенный выше анализ показывает, что в общем виде базовая информационная задача σ_{PS} (на любом уровне детализации или иерархии) может быть представлена в виде кортежа: $\sigma_{PS} = \langle A_1, A_2, \alpha, K, R_\alpha \rangle_{PS}$, где A_1 – описание исходного состояния объектов задачи; A_2 – описание результирующего состояния объектов задачи; α – описание оператора (механизма, способа), осуществляющего отображение: $\alpha: A_1 \rightarrow A_2$, или отношения вида $A_1 \alpha A_2$ (α может выражать, в частности, субъективные и объективные причинно – следственные, родо-видовые, ситуативные, функциональные, ассоциативные и другие связи между объектами задачи); K – описание целей и критериев качества реализации проекта, включая любые ограничения; R_α – описание ресурсов, потребных или выделенных для реализации проекта и документирования. Решение базовой

информационной задачи состоит в том, чтобы определить все элементы кортежа.

Пример описания упрощенного проекта из медицины: <"Пациент болен", "Пациент здоров", "Лечение", "Только терапевтическим путем", "В районной больнице">.

Обычно в задаче неизвестен один или несколько элементов, что позволяет разделить задачи, например, на диагностические ($A_1?$), прогностические ($A_2?$), выбора стратегии и тактики управления ($A_2, \alpha, K, R_\alpha?$), поиска закономерностей (в частности, $\alpha?$) и т.д. Предельно сложный случай задачи имеет место, когда неизвестен ни один элемент кортежа: $\sigma_{PS} = \langle ?, ?, ?, ?, ? \rangle$, например: "на улице нашли лежащего человека в коматозном состоянии" и проект заключается в том, чтобы вернуть человеку максимально возможное здоровье.

Концептуальная модель диагностических знаний

Определим основные понятия:

S – произвольная фиксированная система – субъект ПрО.

ω - произвольный объект ПрО (реальный или абстрактный), который может быть задокументирован системой S .

Ω_ω - класс объектов, однотипных ω (с точки зрения документирования), или некоторое множество объектов, например, агрегат. Любой ω может входить во множество различных классов - множеств: $\Omega_\omega^1, \Omega_\omega^2, \dots, \Omega_\omega^n$. Любой класс Ω_ω также является объектом ПрО. Запись $\Omega.\omega$ будет означать, что рассматривается объект ω из класса Ω . Поскольку могут быть образованы классы классов, то цепочка вложений может быть сколь угодно длинной, например: $\Omega_1.\Omega_2.\dots.\Omega_n.\omega$

Возможность вхождения объекта в разные классы предполагает наличие у него различных имен, алиасов, кодов и т.д., которые должны быть связаны между собой. Совокупность всех кодов объекта может быть объединена в отдельный класс.

Σ - вербальные знания системы S о предметной области (База Знаний - БЗ). Примем, что любое $\omega \in \Sigma$. Сама Σ также является объектом. Согласно определению вербальные знания позволяют максимально полно задокументировать решение базовой информационной задачи любого проекта ПрО.

$\sigma = \langle A_1, A_2, \alpha, K, R_\alpha \rangle$ - документальное описание произвольной информационной задачи (проекта), решенной или решаемой системой S (обозначения те же). Очевидно, σ является объектом БЗ. С помощью σ могут быть описаны любые отображения, проблемные ситуации и связи известные системе. Будем использовать нотацию $\sigma \Rightarrow \omega$, если на каком-либо уровне описания задачи σ используется объект ω .

$\Sigma_\sigma = \{ \omega \mid \omega \in \Sigma \wedge \sigma \Rightarrow \omega \}$ – множество объектов БЗ, используемых для описания задачи σ .

t – время (значения времени t могут принадлежать разным порядковым шкалам: числовым, лингвистическим и т.д.). Объекты ПрО могут появляться и исчезать, менять свои свойства с течением времени, поэтому имеют смысл записи: ω, Σ_t

Если любые две системы S_1 и S_2 обладают собственными БЗ - Σ_1 и Σ_2 соответственно, то возникают задачи сравнения ($\Sigma_1 \cap \Sigma_2, \Sigma_1 \setminus \Sigma_2$ и т.д.) и/или интеграции (θ - объединения) знаний различных систем: $\Sigma_1 \theta \Sigma_2$, где θ - различные операции (способы) объединения при реализации тех или иных проектов (например, операции слияния знаний, взаимообогащения, вариант консилиума и т.д.). Так, если $S_{1,2}$ – чисто компьютерные системы, то теоретически возможно слияние знаний. Если $S_{1,2}$ - человеко – машинные системы, то возможны либо консилиум, либо взаимообогащение знаниями. Нас будут интересовать, в первую очередь, системы "Врач - ИМС", "Пациент - ИМС" и "ИМС".

Любая информация об объектах ПрО, например, их свойствах и характеристиках, параметрах управления и т.д., может быть выявлена только с помощью специальных тестов. Выполнение любого теста означает реализацию отдельного проекта. Пусть τ_ω - тест, предназначенный для исследования ω .

$T_\omega = \{ \tau_\omega \}$ - множество всех тестов, предназначенных для исследования ω .

T_Ω - множество тестов, предназначенных для исследования класса объектов Ω .

В общем случае справедливо:

$$T_\omega = \bigcup_{i=1}^n T_{\Omega_\omega^i}$$

Примем, что $\forall \omega \in \Sigma, \tau_\omega, T_\omega, T_\Omega$ также являются объектами ПрО и, следовательно, принадлежат Σ .

Любой тест τ может иметь один или несколько механизмов реализации, информационные модели (описание) которых обозначим через $\mu_{\tau v}$, где $v \in N_\tau$. Все множество механизмов реализации теста τ обозначим через M_τ .

При выполнении любого теста $\tau \in T_\omega$ с помощью механизма реализации $\mu \in M_\tau$ должны быть соблюдены определенные метрологические требования (условия), которые обозначим через s_{μ} . Очевидно, в конкретной ситуации могут быть задействованы только такие механизмы реализации теста, которые реально осуществимы и для которых выполнимы требования метрологии.

Если необходимо подчеркнуть, что тест τ реализуется через механизм μ с соблюдением метрологических требований s , то могут быть использованы следующие нотации: $\tau/\mu, s; \tau/\mu; \tau/s$.

Выбор системой S того или иного механизма реализации теста зависит от ряда критериев K_τ , которые принимают различные значения в зависимости от конкретных обстоятельств β в модели действительности. Значения данных критериев для конкретного механизма μ и обстоятельств β обозначим через $K_{\tau\beta}(\beta)$. Очевидно, у системы S должен существовать механизм μ_{K_τ} выбора наилучшего (допустимого, оптимального) способа реализации теста в конкретных условиях β .

Примем, что $\forall \omega \in \Sigma, \forall \tau \in T_\omega, \mu_\tau (v \in N_\tau), M_\tau, c_\tau, K_\tau, \mu_{K_\tau}$ также являются объектами ПрО и, следовательно, принадлежат Σ .

С каждым тестом $\tau \in T_\omega$ связано множество вопросов $q \in Q_\tau$, например: $q_1 =$ "Может ли быть выполнен тест τ (с помощью механизма μ при выполнении метрологических условий c)?"; $q_2 =$ "Проводился ли тест τ ?"; $q_3 =$ "Есть ли результаты теста τ ?"; $q_4 =$ "Каковы результаты теста τ ?"; $q_5 =$ "Каковы механизмы реализации теста τ ?"; $q_6 =$ "Каковы критерии выбора того или иного механизма реализации теста τ ?"; $q_7 =$ "Каковы чувствительность (Se) и специфичность (Sp) теста τ ?" и т.д.

Ответы на вопрос $q \in Q_\tau$ могут выбираться или формироваться из разных множеств ответов D_{qv} (доменов). Для фиксации того, что в качестве множества ответов на вопрос q используется домен D , будем использовать следующую нотацию: q/D . В общем случае D может не быть множеством готовых ответов, например, D может быть лексическим деревом [1,2,4] (ответ формируется из стандартных лексем) или множеством всех фраз естественного (профессионального) языка. Следовательно, в общем случае можно говорить лишь о выводимости ответа d на базе конструкций множества D , что будем отражать с помощью нотации: $D \vdash d$. Домены могут быть неполны, т.е. не все элементы доменов могут быть известны. Используя разные домены, можно управлять общностью (масштабом) ответа. Примеры доменов:

Темп.тела $\in D_1 = [35-41]^\circ\text{C}$

Темп.тела $\in D_2 = \{\text{Нормальная, повышенная, субфебрильная, ...}\}$

Боль $\in \{\text{Сильная, слабая, ...}\}$

Продолжительность $\in \{\text{часы}\}, \{\text{дни}\}, \{\text{месяцы}\}, \{\text{года}\}, \{\text{несколько часов, дней, мес, лет; ...; недавно; точно не известно; ...}\},$ и т.д.

Пусть $\forall \omega \in \Sigma, \tau \in T_\omega, q \in Q_\tau$. Универсальную схему исследования объекта определим следующим образом:

$\tau/\mu, c \quad q/D \quad ?d$, где $D \vdash d$

Если используются механизм реализации теста и множество ответов, принятые по умолчанию, то запись может быть упрощена: $\tau \quad q?d$. По умолчанию может применяться и вопрос: "Каков результат теста?". В этом случае можно использовать нотацию: $\tau?d$ или $q/D?d$

Рассмотрим примеры.

Темп.тела / D_1 ? 37,5°C

Темп.тела / D_2 ? Повышенная

Введем следующие сокращения: МКБ-10 = "Международный классификатор болезней 10-го пересмотра", КлинDs = "Клинический диагноз". Справ_Sp = {Низкая, Неопределенная, Высокая}. Используем данные обозначения для записи следующих (упрощенных) примеров:

$\Omega =$ Пациенты, $\Omega.\omega =$ Пациенты.Петров, $\tau =$ Пациенты.Петров.Диагноз, $q = \tau$

Пациенты.Список? Петров, Сидоров, Тарасов,...

Пациенты.Петров.Номер_ИБ? 1234

τ Дата установления? 12.01.2004

Пациенты.Петров.Диагноз /МКБ-10 ? G23

Пациенты.Петров.Диагноз /КлинDs? Острая, левосторонняя пневмония

Пациенты.Петров.Диагноз/ Врач Иванова Se ? 76%

Пациенты.Петров.Диагноз/ Консилиум, Больница Se ? 83%

Пациенты.Петров.Диагноз/ ИМС Se ? 96%

Пациенты.Петров.Диагноз/ Врач Иванова Sp ? 82%

Пациенты.Петров.Диагноз/ Врач Иванова Sp /Справ_Sp? Высокая

В последних пяти примерах "Врач Иванова", "Консилиум" и "ИМС" играют роль различных механизмов реализации теста, а "Больница" играет роль необходимых условий для реализации механизма "Консилиум".

Рассмотрим другие примеры.

Жалобы? Есть

Жалобы.Одышка? Есть

Жалобы.Одышка.Особенности проявления? Постоянная

Жалобы.Головные боли? Есть

С помощью отступов от левого края можно указывать уровень вложенности теста в иерархии тестов. Это позволяет сократить число повторов тестов в протоколе обследования. Примеры:

Консультация терапевта? Есть
Жалобы? Есть
Одышка? Есть
Причины появления? В покое, после физ.нагрузки
Особенности проявления? Периодическая
Длительность? Несколько лет
Головные боли? Есть
Локализация? Височная область, лобная область
Характер проявления? Распирающие
Отрыжки? Не беспокоят
Анамнез? Есть
Желтуха? Нет
Болезни родителей.Рак? Нет
Консультация окулиста? Есть
Жалобы? Запорошенность, светобоязнь, отек век
Анамнез.давность? 2 месяца
Об-но? Есть
Веки? Гиперемия, невусы
Склера? Желтая

Полезно ввести специальный класс тестов – Тест –агрегат. Тест –агрегат представляет собой совокупность элементарных тестов (тестов, не имеющих подуровней) и записывается следующим образом:

ТЕСТ(Парам1, Парам2,...ПарамК)
АН_КРОВИ (Лейкоциты, Эритроциты,...)

Результат теста – агрегата записывается следующим образом:

ТЕСТ? Парам1=знач1, Парам2 = знач2,..., ПарамК = значК
АНАЛИЗ МОЧИ? белок = отс., глюкоза = отс.
ЭКГ? Гипертрофия лев.жел.= отс., Очаговые изменения = отс.

Или упрощенно (если определялись все параметры): ТЕСТ? знач1, знач2,..., значК

Каждый Парам1,..., ПарамК представляет собой элементарный тест из класса ТЕСТ, поэтому справедливы записи: ТЕСТ.Парам1? знач1 и т.д.

Для целей диагностики чрезвычайно полезным является класс тестов – Тест-АДД (АДД-Алгоритм Дифференциальной Диагностики). Данный класс тестов может быть реализован на основе лексического дерева [1,2] и специально здесь не рассматривается (это предмет отдельной статьи).

Конкретизация теста в виде Теста – агрегата, Теста-АДД и других структур является по сути частичным или полным описанием одного из возможных механизмов реализации сложного теста.

Заключение

Несмотря на то, что рассматривалась модель диагностических знаний, предлагаемая модель является более общей. Действительно, в качестве теста можно рассматривать произвольное лечебное воздействие (все зависит от механизма реализации). Более того, в реальной клинической практике часто бывает так, что только вид реакции на проводимое лечение позволяет установить диагноз. Дополнением к рассматриваемой модели является работа [3], в которой изучались вопросы формализации клинического диагноза.

Предлагаемая концептуальная модель позволяет построить реляционную модель базы знаний ИМС, т.е. такую модель, которая эффективно реализуется современными программными средствами [1-3]. Конечной целью проекта является создание ИМС "Виртуальный доктор" и "Виртуальный пациент".

Литература

1. Прокопчук Ю.А., Костра В.В. Интеллектуальная обработка данных в открытых информационных системах // Проблемы программирования. N 1-2. 2002. С.390-395
2. Прокопчук Ю.А. Введение в проблематику баз данных и знаний. Конспект лекций. – Днепропетровск: ИПК ИнКомЦентра УГХТУ, 2003.- 185с.
3. Прокопчук Ю.А. К вопросу формализации понятия "Клинический диагноз". // Сб. докладов Междунар. конф. "Информационные технологии и кибернетика на службе здравоохранения" (Днепропетровск, июнь 2003 г.). – Днепропетровск: ИПК ИнКомЦентра УГХТУ, 2003.- С.85-91
4. Yuriy Prokopchuk, Vladimir Kostra: Representation of Algorithmic Knowledge in Medical Information Systems // Knowledge-Based Intelligent Information and Engineering Systems, 7th International Conference, KES 2003, Oxford, UK, September 3-5, 2003, Proceedings, Part II, Series: Lecture Notes in Artificial Intelligence, Vol. 2773: Palade, Vasile; Howlett, Robert J.; Jain, Lakhmi C. (Eds.) Springer-Verlag Berlin Heidelberg 2003, LI, ISBN: 3-540-40804-5, pp. 1069-1074.
5. Prokopchuk, Yuriy A., Kostra, Vladimir V. Intelligent Modules of Medical Information Systems / Proceedings of World Multiconference on Systemics, Cybernetics and Informatics (July 22-25, 2001 Orlando, Florida, USA). –V. XVIII.-Pp.201-204