

УДК 62.506.2

Б. В. ДЗЮНДЗЮК, канд. техн. наук, *В. Я. ТЕРЗИАН*, канд. техн. наук

**ПРИНЦИПЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНОГО ОБУЧАЕМОГО
ЕСТЕСТВЕННО-ЯЗЫКОВОГО ИНТЕРФЕЙСА ДЛЯ РЕШЕНИЯ ЗАДАЧ
БОРЬБЫ С ПРОФЕССИОНАЛЬНЫМИ ЗАБОЛЕВАНИЯМИ**

Эффективность использования вычислительной техники для решения важнейших народнохозяйственных задач во многом определяется созданием естественно-языковых интерфейсов, обеспечивающих общение непрограммирующих профессионалов с ЭВМ на естественном языке. Методы и процедуры обработки информации, используемые в подобных интерфейсах, могут быть задействованы и для решения самих задач, которые обслуживаются данными интерфейсами.

В данной работе исследуется возможность применения естественно-языкового (ЕЯ) интерфейса для решения задач диагностики, профилактики и лечения профессиональных заболеваний. К настоящему времени накоплен определенный опыт по созданию на базе ЭВМ

автоматизированных систем медицинской диагностики. Однако традиционные системы имеют ряд существенных недостатков, из которых выделим следующие.

1. Тестовые вопросы, посредством которых система опрашивает пациента, либо заранее оговариваются, либо программно встраиваются в систему. Информация в ответ на эти вопросы вводится в систему по строго определенному формату и обрабатывается по жесткой программе. Чтобы даже незначительно изменить или пополнить набор тестовых вопросов или изменить форму ответа, потребуются существенные изменения в программном обеспечении системы.

2. Специальные медицинские знания системы формируются заранее либо в форме процедурных знаний, либо на специализированных языках формирования знаний, недоступных непрограммирующим профессионалам. Отсутствует возможность декларативного обучения системы в процессе ее функционирования. То есть при изменении научных представлений о характере тех или иных заболеваний, о действии тех или иных лекарственных препаратов; при необходимости ввести знания о новых заболеваниях или препаратах; при переходе от одной области медицины к другой потребуются коренная перестройка программного обеспечения системы, выполняемая системными программистами. Этот недостаток является очень существенным, поскольку научные представления в области медицины меняются очень часто и заранее все учесть даже в узкой подобласти не представляется возможным.

3. Отсутствует возможность расширения функциональных возможностей системы по мере накопления ею опыта работы. Система не может синтезировать новые знания (формировать умозаключения и осуществлять вывод), не может подключать к работе те или иные свои процедуры по декларативному запросу.

4. Система не способна выработать рекомендации по устранению объективных причин, вызвавших то или иное заболевание.

Для создания эффективных систем медицинской диагностики необходимо устранить указанные недостатки. Это можно сделать на основе методов искусственного интеллекта, использования ЕЯ интерфейсов, принципов обучения, самообучения, независимости программного и информационного обеспечения системы.

Конечной целью данного исследования является создание диалоговой обучаемой ЕЯ системы диагностики, лечения и предупреждения профессиональных заболеваний (ДЕСТА — «Диагноз»). К основным функциональным возможностям системы предъявляются следующие требования.

1. Возможность обучения системы на естественном языке знаниям о заболеваниях, вредных воздействиях, средствах защиты, лекарственных препаратах, процедурах, алгоритмах расчета параметров защитных средств практически «с нуля» непрограммируемыми профессионалами.

2. Установление диагноза путем целенаправленного оптимального тестирования пациента вопросами, синтезируемыми системой автоматически на основе ее знаний о заболеваниях.

3. Назначение оптимального набора процедур и лекарственных препаратов для лечения заболеваний на основе знаний о показаниях и противопоказаниях препаратов.

4. Выдача рекомендаций по оптимальному изменению условий работы на производстве с целью максимально уменьшить их вредное воздействие на человека на основе знаний системы о средствах защиты, вредных воздействиях, условиях работы на конкретном участке и т. п.

5. Предварительное планирование оптимального набора условий работы на производстве, предполагающих надежную защиту рабочих от вредных воздействий.

6. Возможность использования системы в любой области медицины на основе принципа независимости программного и информационного обеспечения системы, позволяющего переходить от одной предметной области к другой без изменения программного обеспечения с помощью процесса обучения.

7. Совершенствование функциональных возможностей системы по мере накопления ею опыта работы.

8. Извлечение при выдаче рекомендаций из базы знаний новых фактов, в явном виде в ней не содержащихся (формирование умозаключений и вывод).

9. Синтезирование по ЕЯ запросу последовательности программных модулей для решения задач расчета параметров средств защиты.

ЕЯ интерфейс, обеспечивающий функционирование данной системы, базируется на принципах, лежащих в основе создания диалоговой ЕЯ системы ДЕСТА [1]. Процессы интерпретации запросов, вывода, метавывода, формирования умозаключений, обеспечивающие решение задач, на которые ориентирована система, основываются на принципах и механизмах семантического анализа в системе ДЕСТА [2—5].

Рассмотрим принципы формирования, представления и хранения знаний системы ДЕСТА — «Диагноз».

Собственно-лингвистические знания. К декларативным собственно-лингвистическим знаниям системы относится морфолого-синтаксическая информация (МСИ) словоформ, составляющих словарный запас системы. Форма декларативного задания этих знаний следующая [6],

(словоформа) 1 (местоимение) 2 (вопрос) 3 (предлог)

Например: магнитный2какой; магнитным2каким; магнитного2какого; магнитному2какому; машинный2какой; защитный2какой.

Собственно-лингвистические знания системы хранятся в ТВ-структуре [6]. ТВ-структура для рассмотренного примера приведена на рис. 1. Как видно из рисунка, организация структуры позволяет определить МСИ для словоформ «машинным», «машинного», «машинному», «защитным», «защитного», «защитному», которые не были занесены в ТВ-структуру.

Знания системы о внешнем мире. Этот традиционный для ЕЯ систем тип знаний в данной системе включает в себя знания о конкретных заболеваниях, о функциональном назначении и противопоказа-

ниях тех или иных лекарственных препаратов, о вредных воздействиях, о методах и средствах защиты от вредных воздействий. Эти знания задаются системе с помощью текстов на естественном языке, что могут сделать в процессе функционирования системы специалист в области медицины и охраны труда без привлечения программистов.

Классифицироваться знания будут по заглавиям текстов, с помощью которых они были заданы системе. Заглавия текстов определяют имена семантико-прагматических отношений, посредством которых эти знания будут увязаны в единую семантическую сеть знаний системы о внешнем мире. Рассмотрим форму задания знаний о внешнем мире.

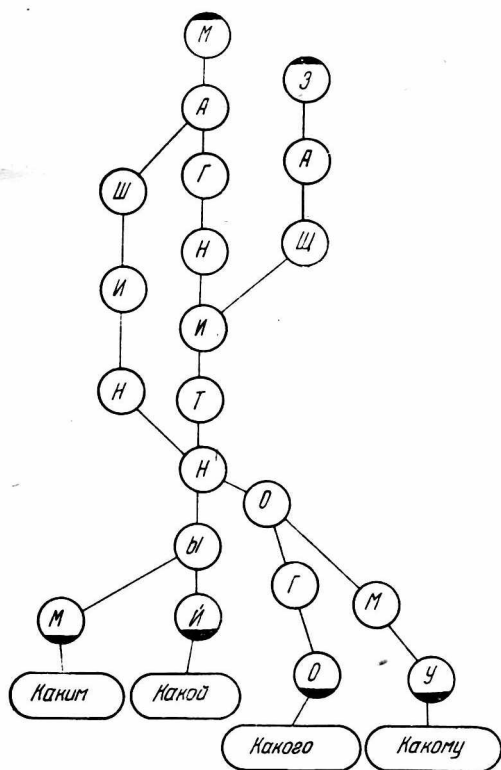


Рис. 1

или симптома 1, на фоне которого противопоказано использование данного препарата). (наименование заболевания или симптома 2). ... (наименование заболевания или симптом N). ✖END.

3. Знания о вредных воздействиях. ТЕКСТ: **результат** (наименование вредного воздействия). (наименование заболевания или симптома 1, порождаемого данным воздействием). (наименование заболевания или симптома 2). ... (наименование заболевания или симптома N). ✖END. ТЕКСТ: **условие** (наименование вредного воздействия). (наименование условия 1, приводящего к возникновению данного воздействия). (наименование условия 2). ... (наименование условия N). ✖END.

4. Знания о средствах защиты от вредных воздействий. ТЕКСТ: **назначение** (наименование средства защиты). (наименование вредного воздействия 1, для борьбы с которым используется данное средство защиты). (наименование вредного воздействия 2). ... (наименование вредного воздействия N). ✖END.

Рассмотрим форму представления знаний системы о внешнем мире. Знания о внешнем мире преобразуются в форму R-представлений

[1], основу которых составляют синтактико-семантические отношения (ССО) и семантико-прагматические отношения (СПО).

Формат ССО: $M: B, A, C (HE, T_l, D_l^1, D_l^2, E_l^1, E_l^2, X_l, K_l) = (HE, T_r, D_r^1, D_r^2, E_r^1, E_r^2, X_r, K_r)$, где B, A и C — это МСИ терминальной словоформы T_r ; B — имя ССО, представленное вопросительным словом; A — местоимение; C — предлог; HE — частица; T — нормализованный вид словоформы; D^1, D^2 — соответственно первая и вторая приставки словоформы; E^1, E^2 — соответственно первый и второй суффиксы словоформы; X — предметная переменная; K — квантор; M — метка ССО. Индексы l и r обозначают соответственно элементы левой и правой частей ССО.

Формат СПО: $MC: L (H^i) = (andH^i)$, где MC — метка СПО; L — имя СПО; H — естественно-языковая конструкция (ЕЯК) (ССО, словосочетание, факт, ситуация, текст).

Рассмотрим пример R -представления для простейшего текста, который вводится в систему в следующей последовательности:

ТЕКСТ: симптомы ангины. Боль в горле. Повышенная температура. Нагноение миндалин. Затруднена речь. $\neq END$.

Используя правила, изложенные в [1, 2], получаем следующее R -представление для данного текста: $MH1$: что, она = (ангина); $MH2$: что, она = (боль); $M1$: где, в, оно (боль) = (горло); $MH3$: что, она = (температура); $M2$: какая (температура) = (высокий, по); $MH4$: что, оно = (нагноение); $M3$: чего, они (нагноение) = (миндалины); $MH5$: что, она = (речь); $M4$: какая (речь) = (грудный, за); $MФН1$: ($MH2, M1$); $MФН2$: ($MH3, M2$); $MФН3$: ($MH4, M3$); $MФН4$: ($MH5, M4$); $MC1$: симптом ($MH1$) = ($MФН1$); $MC2$: симптом ($MH1$) = ($MФН2$); $MC3$: симптом ($MH1$) = ($MФН3$); $MC4$: симптом ($MH1$) = ($MФН4$). Приведенное R -представление система формирует автоматически. Здесь MH — метка неполного ССО; $MФН$ — метка словосочетания (неполного факта).

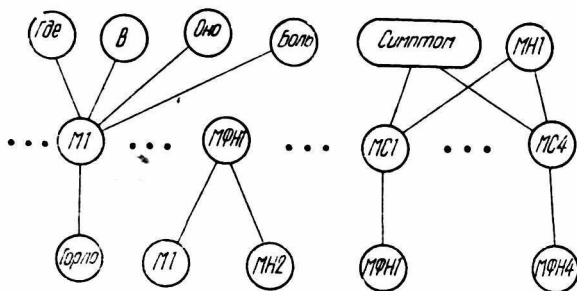


Рис. 2

Знания в форме R -представлений хранятся в S -структуре [1]. S -структура для рассмотренного примера приведена на рис. 2.

Семантические знания системы. К ним относятся знания о взаимосвязи предметов и явлений внешнего мира, представимые в форме СПО: «синоним», «элемент класса», «часть целого», «противоположно» и др. С помощью данных СПО осуществляется вывод в процессе интерпретации входных запросов в терминах базы знаний системы. Задаются эти знания с помощью ЕЯК. Например: «Вибрационная болезнь — элемент класса профессиональные заболевания». Семантические знания система переводит в форму R -представлений и хранит в S -структуре.

Знания системы о прикладных процедурах. Система ДЕСТА [1] позволяет декларативно описывать знания о конкретных процедурах,

входящих в программное обеспечение системы, с целью последующего подключения их по запросу на естественном языке. Рассмотрим пример декларативного описания процедуры с помощью следующего текста.

ТЕКСТ: программа *ELMDPR (D, DVL, E, TG)*. *Z* Назначение. Найти глубину проникновения электромагнитного излучения в биологическую ткань. *Z* Входные параметры. *DVL* — длина волны электромагнитного излучения в вакууме. *E* — диэлектрическая проницаемость биологической ткани. *TG* — тангенс угла потерь в биологической ткани. *Z* Выходные параметры. *D* — глубина проникновения электромагнитного излучения в биологическую ткань. *END.

На данном примере рассмотрим фрагмент диалога с системой. В процессе диалога система по ЕЯ запросу пользователя (при необходимости преобразуя запрос с помощью процедур вывода) находит в базе знаний требуемый программный модуль, а затем, используя ЕЯ описание данного модуля, запрашивает необходимые для его подключения параметры. Когда все параметры определены, система подключает к работе требуемый программный модуль и выдает результаты, указанные в списке выходных параметров в ЕЯ описании. Рассмотрим диалог.

ПЛЗ: Вычислите глубину проникновения в ткань организма излучения от электромагнитного источника.

ДЕСТА: Вам необходимо найти глубину проникновения электромагнитного излучения в биологическую ткань?

ПЛЗ: Да.

ДЕСТА: При задании исходных величин пользуйтесь единой системой единиц, не указывая размерность.

ДЕСТА: Вы знаете длину волны электромагнитного излучения в вакууме?

ПЛЗ: Да.

ДЕСТА: Задайте значение этой величины.

5.671E—26 (это и последующие числа взяты произвольно).

ДЕСТА: Вы знаете диэлектрическую проницаемость биологической ткани?

ПЛЗ: Да.

ДЕСТА: Задайте значение этой величины.

2.1401

ДЕСТА: Вы знаете тангенс угла потерь в биологической ткани?

ПЛЗ: Да.

ДЕСТА: Задайте значение этой величины.

0.097.

ДЕСТА: Вам выдать на печать значение глубины проникновения электромагнитного излучения в биологическую ткань?

ПЛЗ: Да.

ДЕСТА: Печатать на дисплей (консоль)?

ПЛЗ: Да.

9.6421635E—3

ДЕСТА: Вам необходимо найти глубину проникновения электромагнитного излучения в биологическую ткань?

ПЛЗ: Нет.

(конец диалога).

Если пользователь не может задать значение какого-либо из запрашиваемых параметров, например, тангенс угла потерь, то система формирует запрос сама для себя: «Вычислите тангенс угла потерь в биологической ткани». Теперь система решает ту же задачу отыска-

ния необходимого программного модуля, но уже для нового параметра, и действует так до тех пор, пока пользователь не сможет определить все входные данные, после чего система в обратном порядке возвращается к вычислению всех промежуточных переменных до нахождения ответа на исходный запрос. Таким образом, по ЕЯ запросу может быть синтезирована программа, состоящая из последовательного подключения программных модулей, решающая поставленную задачу в процессе диалога с непрограммирующим пользователем. Знания системы, хранимые в С-структуре, нагляднее представлять в форме семантико-прагматического графа (СПГ), вершины кото-

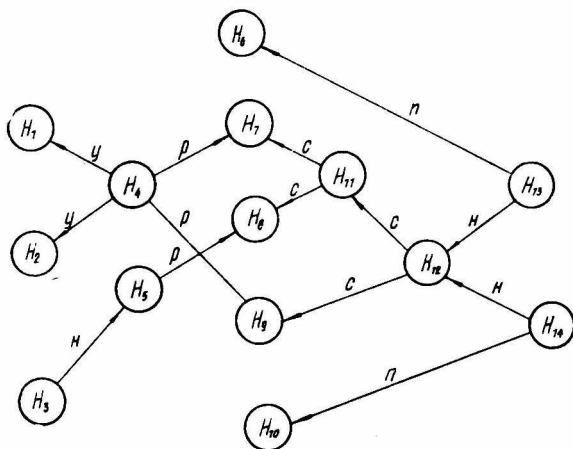


Рис. 3

рого — это метки ЕЯК (ССО, словосочетаний, фактов, ситуаций, текстов и т. д.), а ребра — это СПО, связывающие соответствующие ЕЯК. По метке ЕЯК, используя ее R -представление, всегда можно синтезировать поверхностную структуру ЕЯК и наоборот. Пример СПГ базы знаний системы ДЕСТА — «Диагноз» приведен на рис. 3. Здесь y означает СПО условие; p — результат; n — назначение; c — симптом; n — противопоказание; H — ЕЯК.

На основе приведенного примера покажем некоторые режимы работы системы. Тестирование пациента с целью установления диагноза происходит следующим образом. Система сначала синтезирует в вопросительной форме первую попавшуюся ЕЯК, соответствующую одному из конечных элементов в цепочке симптомов СПГ (например, H6). Пусть на этот вопрос пациент ответил отрицательно, на следующий вопрос (H7) утвердительно. Теперь система следующие вопросы задает целенаправленно. В СПГ базы знаний возбуждение передается от элемента H7 к элементу H11, и значит, далее система задает вопросы, соответствующие остальным симптомам, входящим в H11 (а именно — H8). Пусть и на этот вопрос пациент ответил утвердительно. Теперь элемент H11 возбуждается полностью и от него возбуждение передается элементу H12. Это означает, что следую-

щим будет задан вопрос Н9. Пусть и на этот вопрос пациент ответил утвердительно. Теперь элемент Н12 возбуждается полностью, т. е. диагноз поставлен (наименование заболевания можно синтезировать из ЕЯК Н12). Теперь по СПГ системы можно определить, какие лекарственные препараты необходимо назначить для лечения данного заболевания (им соответствуют ЕЯК Н13 и Н14). При этом возбуждаются элементы Н6 и Н10, соответствующие противопоказаниям к данным лекарственным препаратам. На вопрос Н6 пациент уже отвечал отрицательно, следовательно, лекарственный препарат Н13 подходит. Для определения пригодности лекарственного препарата Н14 система задает пациенту вопрос Н10. Пусть пользователь (пациент) ответит утвердительно. Это говорит о непригодности препарата Н14, следовательно, в данном конкретном случае может быть использован только препарат Н13. Параллельно с процессами диагностики и лечения процедуры работы с СПГ устанавливают наличие вредных воздействий на производстве, которые могли привести к возникновению данного заболевания, а также рекомендуют средства защиты от них. Для данного примера можно установить, что симптомы Н7 и Н9 рассматриваемого заболевания возникают в результате вредного воздействия Н4, а симптом Н8 — в результате воздействия Н5. Из СПГ можно установить, что воздействие Н5 может быть ослаблено при использовании защитного средства Н3. Также система устанавливает причины возникновения воздействия Н4, а именно условия Н1 и Н2. В результате система выдает рекомендации по использованию средства защиты Н3 и изменению условий Н1 и Н2 работы на производстве, которые дадут возможность сократить случаи заболевания болезнью Н12.

Таким образом, рассмотренные задачи сводятся к задачам распространения возбуждений в цепочках СПГ. Данный подход не настроен на работу с каким-либо конкретным набором отношений (его можно корректировать или полностью изменять), а значит, может быть использован в различных областях, где имеет место диалог человека и искусственной системы.

Список литературы: 1. *Ловицкий В. А.* Обучаемая диалоговая естественно-языковая система. Техн. кибернетика // Изв. АН СССР.— 1983.— Вып. 5.— С. 114—127. 2. *Терзиян В. Я.* Принципы организации анализа естественно-языковых высказываний в системах общения пользователей с ЭВМ. Сообщ. 1 // Пробл. бионики.— 1985.— Вып. 34.— С. 47—51. 3. *Терзиян В. Я.* Принципы организации анализа естественно-языковых высказываний в системах общения пользователей с ЭВМ. Сообщ. 2 // Пробл. бионики.— 1985.— Вып. 35.— С. 17—24. 4. *Головина Е. А., Терзиян В. Я.* Экспресс-анализ естественно-языковых высказываний // Интерактивные системы: Материалы пятой шк.-семинара.— Тбилиси, 1983.— С. 385—388. 5. *Головина Е. А.* Принципы проверки семантической правильности естественно-языковых высказываний // Пробл. бионики.— 1984.— Вып. 32.— С. 64—72. 6. *Ловицкий В. А.* Структурный подход к решению морфологических задач // Пробл. бионики.— 1980.— Вып. 25.— С. 39—43.

Поступила в редколлегию 17.01.87