

выше вес признаков ω_k , $k=1..q$, тем ниже качество классификации состояний.

Таким образом, для успешного решения поставленной задачи наиболее оптимальным является алгоритм нечеткой классификации Данна-Беждека позволяющий системе работать в условиях неопределенности, что особенно важно, так как с этой проблемой приходится сталкиваться практически в любых ситуациях.

ЛИТЕРАТУРА

- [1]. Гриняев С.Н. Интеллектуальное противодействие информационному оружию. – М.: СИНТЕГ, 1999.
- [2]. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 415с.
- [3]. Сокал Р.Р. Кластер-анализ и классификация: предпосылки и основные направления// Классификация и кластер /Под ред. Дж. Вэн Райзина. – М.: Мир, 1980. С.7-19.
- [4]. Заде Л.А. Размытые множества и их применение в распознавании образов и кластер-анализе//Классификация и кластер /Под ред. Дж. Вэн Райзина. – М.: Мир, 1980.

ОБ ОДНОМ ПОДХОДЕ К СОГЛАСОВАНИЮ СРЕДСТВ ИЗВЛЕЧЕНИЯ ЗНАНИЙ

В.В. Смирнов

Московский государственный инженерно-физический институт (технический университет),
Каширское ш., 31, Москва, 115409, РОССИЯ, тел. (095) 324-28-85, Vitaly_Smirnov@mail.ru

АННОТАЦИЯ

В работе рассматривается подход к согласованию средств извлечения знаний из экспертов, проблемно-ориентированных текстов и баз данных, использованный в инструментальном комплексе АТ-ТЕХНОЛОГИЯ. Подход предусматривает как согласование обмена данными между компонентами, так и передачи им управления, что обеспечивается организацией взаимодействия компонент по типу “классной доски”, а также применением следующих средств инструментального комплекса АТ-ТЕХНОЛОГИЯ: лингвистического процессора, интерпретатора сценариев диалога с экспертом, средств верификации поля знаний и интеллектуального планировщика.

1. ВВЕДЕНИЕ

В работе [1] были описаны вопросы автоматизированного построения баз знаний (БЗ) для интегрированных экспертных систем (ИЭС) на основе задачно-ориентированной методологии (ЗОМ) и поддерживающего эту методологию инструментального комплекса АТ-ТЕХНОЛОГИЯ [2]. В работе [3] рассмотрено расширение ЗОМ за счет применения баз данных (БД) как дополнительного источника знаний при формировании БЗ. Автоматизированное построение БЗ средствами комплекса АТ-ТЕХНОЛОГИЯ осуществляется в несколько этапов. В данной работе рассматривается под-

ход к согласованию выделенной группы средств, которая обеспечивает этап извлечения знаний (ИЗ) из экспертов, проблемно-ориентированных текстов и баз данных. Фрагмент информационно-логической (ИЛ) модели данной группы средств показан на Рис. 1.

Реализованные компоненты средств ИЗ относятся к одной из следующих групп:

- компонентам специализированного лингвистического процессора, обеспечивающим определение типа решаемой задачи [4] для активизации сценария диалога с экспертом, сбор лексики системного аналитика [5,6] и эксперта из текста с описанием решаемой задачи [7] и протокола интервьюирования эксперта (ПИЭ) [5] с формированием вспомогательных словарей новой лексики [6], получение информации о наличии НЕ-факторов [6,8];
- компонентам интерпретатора сценариев диалога с экспертом, обеспечивающим активизацию сценариев диалога с экспертом, каждый из которых отражает структуру решаемой задачи и обеспечивает способ ввода информации от эксперта по методу “имитация консультации” [9], для выявления дифференцирующих условий поддерживает метод репертуарных решеток [9] и организует взаимодействие с компонентами, извлечения знаний из данных, описанных в работе [3];

- компонентам извлечения знаний из данных по технологии Knowledge Discovery in Databases (KDD) [3];
- компонентам поддержки протоколов интервьюирования экспертов ПИЭ [9];
- компонентам доступа к полю знаний.

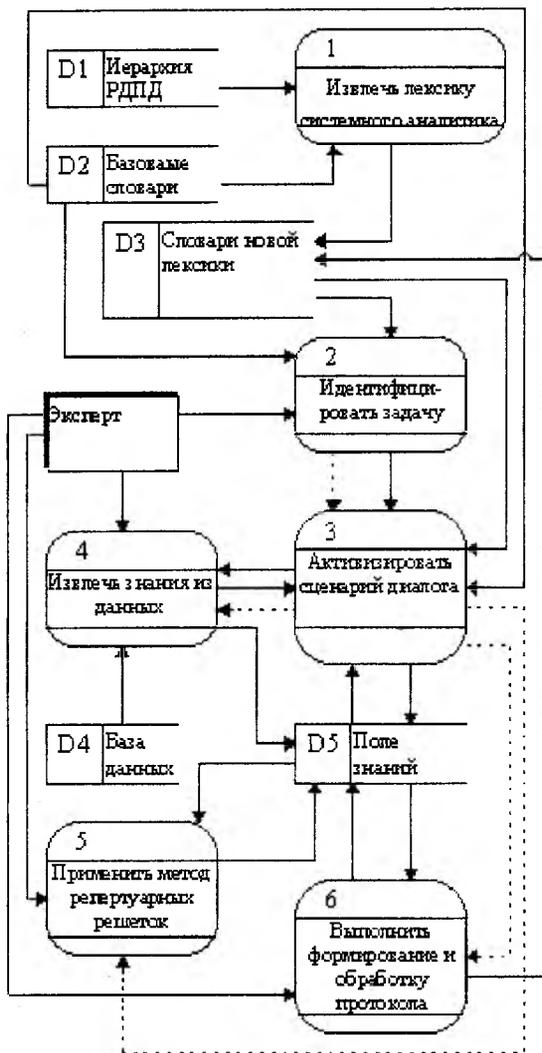


Рисунок 1. Фрагмент ИЛ модели средств ИЗ

Согласование средств извлечения знаний будем рассматривать с точки зрения обмена данными между компонентами средств ИЗ (поток данных на Рис. 1) и с точки зрения последовательности решения задач извлечения знаний (поток управления на Рис. 1).

При согласовании средств ИЗ по обмену данным решаются следующие задачи:

- формирование фрагментов поля знаний (ПЗ) в едином формате ПЗ комплекса АТ-ТЕХНОЛОГИЯ, элементами которого являются объекты и правила [10];
- объединение всех фрагментов поля знаний в единой структуре для независимого доступа

- к ним любого компонента средств построения БЗ;
- отнесение фрагментов поля знаний к конкретному типу решаемой задачи;
- учет источника знаний в каждом фрагменте поля знаний;
- совмещение совпадающих атрибутов во фрагментах поля знаний и приведение их к одному типу для обеспечения возможности их использования в последующих сеансах извлечения знаний, а также для формирования БЗ на ЯПЗ комплекса АТ-ТЕХНОЛОГИЯ или другого инструментального средства [4,10];
- выявление аномалий (противоречивых, циклических, избыточных, пересекающихся, пропущенных и др. правил), вызванных объединением нескольких фрагментов поля знаний.

Среди компонент, обеспечивающих согласование средств ИЗ с точки зрения обмена данными можно выделить компоненты обеспечения доступа к «классной доске», лингвистического процессора и верификации поля знаний [9].

При согласовании передачи управления решаются следующие задачи:

- поддержание порядка активизации средств ИЗ в соответствии с расширенной ЗОМ;
- определение набора используемых функций и последовательности их применения для каждого средства ИЗ на основе информации тип текущей решаемой задачи.

Для согласования средств ИЗ с точки зрения управления применяются компоненты интерпретатора сценариев диалога и компоненты интеллектуального планировщика комплекса АТ-ТЕХНОЛОГИЯ [11,12].

Ниже рассматриваются некоторые особенности реализации средств комплекса АТ-ТЕХНОЛОГИЯ, обеспечивающих согласованность средств ИЗ.

2. СОГЛАСОВАНИЕ СРЕДСТВ ИЗ С ТОЧКИ ЗРЕНИЯ ОБМЕНА ДАННЫМИ

Для обеспечения согласованного обмена данными между средствами ИЗ при сохранении их независимости их взаимодействие организовано по типу «классной доски». В качестве классной доски выступает специальный компонент ТКВControl, который обеспечивают 3 уровня согласования. Верхний уровень обеспечивает согласование по типу решаемой задачи, обеспечивая настройку всех средств ИЗ на ре-

шение задачи одного типа. Второй уровень обеспечивает согласование на уровне сеансов ИЗ, обеспечивая настройку средств ИЗ на один и тот же сеанс и отнесение каждого фрагмента поля знаний к конкретному сеансу ИЗ. Следует отметить, что при создании двух фрагментов поля знаний разными средствами ИЗ они будут отнесены к различным сеансам извлечения знаний. Последний уровень обеспечивает согласование средств ИЗ по конкретным объектам и правилам поля знаний. В частности, обеспечена возможность сохранения и восстановления данных сеанса на любом шаге сценария диалога с экспертом [5], даже если процесс создания очередного правила еще не закончен.

Заметим, что каждое средство ИЗ участвует в обеспечении согласованного обмена данными, так как свое внутренне представление поля знаний преобразует в единый формат после завершения создания очередного фрагмента поля знаний.

Другим средством, участвующим в обеспечении согласования средств ИЗ является специализированный лингвистический процессор, который решает следующие задачи согласования:

- идентификацию решаемой задачи и размещение данных о ее типе на классной доске;
- согласование терминов из разных фрагментов поля знаний за счет приведения слов к канонической форме;
- выявление числовых ограничений из ответов эксперта с последующей корректировкой правил с целью подготовки их к формированию БЗ на ЯПЗ комплекса или других инструментальных средств.

Дадим краткую характеристику базовым функциям специализированного лингвистического процессора комплекса АТ-ТЕХНОЛОГИЯ.

Основной функцией специализированного лингвистического процессора (ЛП) инструментального комплекса АТ-ТЕХНОЛОГИЯ является преобразование фраз ограниченного естественного языка (ЕЯ) в предикатно-аргументные структуры на языке CAREL[13]. Например, фраза «Диагностировать у пациента интерстициальный фиброз легких» может быть представлена в виде следующего CAREL-выражения:

(диагностировать) (МОД (INFN))
 (D (пациент) (Оду)))
 (O (фиброз) (Заб)
 (H (интерстициальный) (Хар))
 (H (легкое) (Нео)))

где

D - глубинный падеж «быть адресатом действия»;

O - глубинный падеж «быть объектом действия»;

H - обозначение связи типа «быть характеристикой»;

МОД,

INFN - зарезервированные слова для обозначения модальности;

Оду, Заб, Хар, Нео - обозначения типов семантических категорий в принятой кодировке использованного в данном случае словаря.

Обработка ЕЯ-фраз, поступающих на вход ЛП, выполняется при помощи морфологического, синтаксического и семантического анализа с использованием словарей предикатов, понятий, характеристик, квазифлексий и классификатора семантических классов и категорий.

При идентификации решаемой задачи используется классификатор семантических классов и категорий, а также вместе с базовыми словарями применяются словари новой лексики. Извлечение новой лексики поддерживается следующими особенностями организации словарей.

Для морфологического анализа текущая версия ЛП использует процедурный метод, который предполагает использование словарей основ со ссылкой на строки в таблице возможных аффиксов. В первых версиях ЛП словари для хранения предикатов и характеристик также были организованы по принципу разделения основ и аффиксов, однако это существенно усложняло процесс наполнения словарей, так как вынуждало заносить слова, имеющие изменяющуюся основу, в словарь исключений. Использование словарей квазиоснов и квазифлексий [14], позволило отказаться от применения словаря исключений. Применение квазифлексий позволило также во многих случаях решить задачу автоматического определения морфологических характеристик новых слов при извлечении лексики СА и эксперта. Для предварительного определения морфологических характеристик новых слов использован дополнительный словарь, содержащий индексированные в порядке убывания длины квазифлексии, причем каждая такая квазифлексия имеет ссылку на запись в основном словаре, соответствующую определенному типу словоизменения. Окончательная корректировка морфологической зоны словарной статьи выполняется с помощью сравнения со словами, являющимися образцами склонения и спряжения.

Рассмотрим, каким образом выполняется согласование фрагментов поля знаний, полученных с применением компонент KDD и интервьюирования эксперта по методу “имитация консультации”. При использовании средств KDD с точки зрения реализации сценария диалога с экспертом возможны три основных способа активизации функций KDD:

- перед интервьюированием эксперта;
- после интервьюирования эксперта;
- включение функций KDD в процесс интервьюирования эксперта.

В работе рассматривается согласование правил только для первого способа реализации сценария диалога с экспертом. Сценарий диалога включает как использование “открытых вопросов”, называющих тему или предмет, и оставляющих эксперту полную свободу по выбору формы и содержания ответа, так и безличных вопросов, направленных на выявление распространенных и общепринятых закономерностей ПО [10]. Кроме того, введенные на предыдущих шагах термины заносятся во вспомогательный словарь, облегчающий ввод информации на новом шаге диалога [4,9]. Как отмечено в работе [3] представляется маловероятным существование специфичной БД, набор атрибутов которой может отображаться на структуру поля знаний, сформированного в результате интервьюирования эксперта, поэтому в интерпретатор сценариев диалога включены функции, позволяющие выполнить обратное отображение. Для такого отображения во вспомогательный словарь включены термины не только текущего сеанса, но и термины фрагмента поля знаний, полученного средствами KDD. Для того, чтобы учесть возможные расхождения форм слов, термины приводятся к канонической форме с помощью ЛП комплекса АТ-ТЕХНОЛОГИЯ.

В работах [6,8] рассматриваются вопросы извлечения знаний с НЕ-факторами из ответов эксперта, в данной работе описывается процесс использования выявляемой в ответах эксперта информации о неопределенности, а также точных числовых значений с целью согласования правил поля знаний по типам значений атрибутов. В случае применения “открытых вопросов” в процессе диалога границы числовых интервалов или отдельные числовые значения в ответах эксперта могут встречаться в тексте в сочетании со словами, указывающими на тип ограничения. Например слова “больше”, “от”, “с” и др. указывают на левую границу интервала, “меньше”, “до”, “по” и др. - на правую, а

“равно”, “составляет”, и др. - на число, являющееся одновременно левой и правой границей. В результате удается представить ответы эксперта, содержащие числовые ограничения, в виде логических выражений, содержащих атрибуты числового типа. Например, в процессе интервьюирования эксперта имел место такой шаг диалога с экспертом:

Система: Сформулируйте вопрос по поводу состояния пациента так, как будто Вы спрашиваете пациента.

Вопрос: В течение какого периода наблюдалось недомогание?

Система: Сформулируйте возможный вариант ответа пациента на данный вопрос.

Ответ: Симптомы недомогания наблюдались более 6 недель.

На основе обработки данного ответа эксперта в правило было включено числовое ограничение:

ТИП Тх
 ЧИСЛО
 КОММЕНТАРИЙ период

ОБЪЕКТ ОБЪЕКТу
 АТТРИБУТЫ

 АТТРИБУТ АТТРИБУТz
 ТИП Тх
 КОММЕНТАРИЙ недомогание

 КОММЕНТАРИЙ пациент

ПРАВИЛО
 ЕСЛИ & ОБЪЕКТу.АТТРИБУТz>6 |
 ТО

Для объединения и согласования фрагментов поля знаний, полученных последовательным применением компонент KDD и интервьюирования эксперта, используются компоненты верификации поля знаний, возможности которых описаны в работах [9,15]. Информация об обнаруженных аномальных правилах (противоречивых, избыточных, циклических, пропущенных, конфликтных, пересекающихся) и аномальных атрибутах (без ссылок, с неучтенным значением, неверного типа) заносится в протокол верификации для того, чтобы на его основе инженер по знаниям мог принять решение о том, какие средства использовать для корректировки поля знаний.

3. СОГЛАСОВАНИЕ СРЕДСТВ ИЗ С ТОЧКИ ЗРЕНИЯ ПЕРЕДАЧИ УПРАВЛЕНИЯ

Порядок применения методов “имитация консультации” и репертуарных решеток определяется сценарием интервьюирования эксперта. Для текущей версии комплекса АТ-ТЕХНОЛОГИЯ разработан сценарий диалога с экспертом для задачи диагностики, позволяющий активизировать компоненты поддержки метода репертуарных решеток в двух случаях. В первом случае, после ввода каждого нового заключения о состоянии пациента или диагностируемого объекта, эксперту предлагается либо уточнить заключение, указав наиболее определяющие (дифференцирующие), признаки, либо продолжить процесс интервьюирования без уточнения заключения. Если эксперт выбирает первый вариант, то активизируется диалог, для выявления дифференцирующих признаков методом репертуарных решеток. Во втором случае, после ввода очередного ответа эксперта на вопрос выполняется анализ введенной фразы с помощью ЛП на предмет наличия понятий, указывающих на то, что речь идет не о единичном признаке, а о некотором синдроме. Если в очередном ответе эксперта на вопрос содержится указание на синдром, то эксперту предлагается уточнить данный ответ и в случае согласия активизируются компоненты поддержки метода репертуарных решеток. Для того, чтобы ЛП был способен выявлять синдромы, в словарь занесены названия часто встречающихся синдромов со ссылкой на соответствующий семантический класс.

С точки зрения единого управления проектом по созданию ИЭС, согласование процессов ИЗ осуществляется на основе сценариев реализации типовых проектных процедур, использованных в интеллектуальном планировщике комплекса АТ-ТЕХНОЛОГИЯ и описанных в работе [12].

Исследования выполнялись при поддержке РФФИ РАН (проект № 00-01-00679).

ЛИТЕРАТУРА

[1]. Рыбина Г.В. Автоматизированное построение баз знаний для интегрированных экспертных систем // Изв. РАН. Теория и системы управления. № 5, 1998. С.152-166.
[2]. Рыбина Г.В. Задачно-ориентированная методология автоматизированного построения интегрированных экспертных систем для статических проблемных областей // Изв. РАН. Теория и системы управления. № 5, 1997. С.129-137.

[3]. Калинина Е.А., Рыбина Г.В. Применение технологии Data Mining для автоматизированного построения баз знаний интегрированных экспертных систем // В кн. КИИ-2000. Седьмая нац. конференция с межд. участием. Тр. конф. М.: Физматлит, 2000, Т.2. С. 119-127.
[4]. Пышагин С.В., Рыбина Г.В., Смирнов В.В. Инструментальный комплекс АТ-ТЕХНОЛОГИЯ для поддержки проектирования интегрированных экспертных систем // В кн.: КИИ-96 Пятая нац. конференция с межд. участием “Искусственный интеллект-96”. Сборник научных трудов в трех томах. Том 3. Казань, 1996. С. 522-527.
[5]. Рыбина Г.В., Пышагин С.В., Смирнов В.В., Чабаяев А.В., Пашина И.А. Автоматизированное построение интегрированных экспертных систем на основе средств инструментального комплекса АТ-ТЕХНОЛОГИЯ (версия MS-Windows) // В кн.: Международная летняя школа-семинар по искусственному интеллекту для студентов, аспирантов и молодых ученых. Сборник трудов. Минск, 1997, С. 119-137.
[6]. Рыбина Г.В., Душкин Р.В., Душкина Е.Н. О новых возможностях лингвистического процессора инструментального комплекса АТ-ТЕХНОЛОГИЯ // В кн.: Научная сессия МИФИ-2001. Сборник научных трудов. Том 3. М.:МИФИ, 2001, С. 134-135.
[7]. Рыбина Г.В., Смирнов В.В., Кустикова И.А., Ледовская Т.В., Солонович Е.А., Файбисович М.А., Гриценко Ф.А. Автоматизированное построение базы знаний для интегрированных экспертных систем на основе средств комплекса АТ-ТЕХНОЛОГИЯ // В кн.: Научная сессия МИФИ-98. Сборник научных трудов. Ч.5. М.:МИФИ, 1998, С. 34-37.
[8]. Рыбина Г.В. Душкин Р.В. Душкина Е.Н. Лингвистические аспекты извлечения знаний, содержащих неопределенность, неточность и нечеткость // В кн.: Международный научно-практический семинар “Интегрированные модели и мягкие вычисления в искусственном интеллекте”. Сборник трудов. М.: Физматлит: 2001. С. 168-172.
[9]. Кустикова И.А., Рыбина Г.В., Смирнов В.В. Об одном подходе к автоматизированному построению базы знаний для интегрированных экспертных систем: аспекты тестирования // В кн.: КИИ’98 Шестая нац. конференция по искусственному интеллекту с межд. участием. Сборник научных трудов в трех томах. Том 2. Пушино: РАИИ, 1998. С. 138-145.
[10]. Колобашкина М.В. Рыбина Г.В. Сергиевская О.Г. Смирнов В.В. Задачно-ориентированная методология приобретения знаний для компьютерного построения интегрированных экспертных систем // В кн.: КИИ-96 Пятая нац. конференция с межд. участием “Искусственный интеллект-

- 96". Сборник научных трудов в трех томах. Том 2. Казань, 1996. С. 270-274.
- [11] Рыбина Г.В., Пышагин С.В. Интеллектуальная поддержка разработки интегрированных экспертных систем // В кн.: КИИ'98 Шестая нац. конференция по искусственному интеллекту с межд. участием. Сборник научных трудов в трех томах. Том 2. Пушкино: РАИИ, 1998. С. 419-426
- [12] Левин Д.Е., Пышагин С.В., Рыбина Г.В., Смирнов В.В. Новые возможности инструментального комплекса АТ-ТЕХНОЛОГИЯ, предназначенного для поддержки построения интегрированных экспертных систем // В кн.: КИИ'2000. Седьмая национальная конференция по искусственному интеллекту с международным участием. Труды конференции. М.: Изд-во физико-математической литературы, 2000. Т. 2. С. 751-757.
- [13] Рыбина Г.В. Модель диалога в естественно-языковой системе ДИСАР // В кн.: Автоматизированная информационная технология. М.: Энергоатомиздат, 1990, С.29-36.
- [14] Автоматизация анализа научного текста. Киев, "Наукова думка", 1984.
- [15] Рыбина Г.В., Пышагин С.В., Смирнов В.В., Левин Д.Е., Душкин Р.В. Инструментальный комплекс АТ-ТЕХНОЛОГИЯ для поддержки разработки интегрированных экспертных систем: Учебное пособие. М.: МИФИ, 2001.

ОРГАНИЗАЦИЯ БАЗ ДАННЫХ ДЛЯ ГИБРИДНЫХ СИСТЕМ МОДЕЛИРОВАНИЯ И ОПТИМИЗАЦИИ

А.В. Жиренкин

Сибирский Государственный Университет Телекоммуникаций и Информатики, РОССИЯ,
Новосибирск, тел. (3832) 39-65-34, zhirenkin@katren.ru

АННОТАЦИЯ

В настоящее время в Лаборатории Математического Моделирования и Информационных Сетей Института Вычислительной Математики и Математической Геофизики Сибирского Отделения Российской Академии Наук ведутся работы по созданию гибридных систем моделирования и оптимизации транспортных и коммуникационных сетей. В рамках этой темы возникает вопрос об исследовании и разработке баз данных для подобных систем, включающий в себя создание структуры хранения информации, средств доступа к данным и средств обработки и анализа данных. В данной работе описана модель базы данных, наиболее полно отвечающая предложенным критериям хранения и обработки информации для гибридных систем моделирования.

1. ВВЕДЕНИЕ

Классической методикой проектирования баз данных является создание отдельной таблицы для каждой описываемой моделью данных сущности, затем, в процессе нормализации – выделение отдельных таблиц для хранения атрибутов сущности (таблицы-справочники). Такой подход хорошо работает для баз данных с относительно небольшим количеством описываемых объектов (десятки) и при несложных и статичных связях между ними. Однако любое

изменение структуры хранимых данных приводит к внесению изменений в структуру таблиц, эти данные отображающих. Несложная на этапе разработки, эта операция становится крайне проблематичной при больших объемах данных и отсутствии у разработчика непосредственного доступа к БД (например, если она стоит у заказчика). Многим, наверное, знакомы громоздкие, медленные и требующие огромного дискового пространства операции по конвертированию БД при переходе на новую версию продукта. Не менее неприятна работа с базами данных, исторически разросшимися до сотен таблиц, структуру, которой сложно даже изобразить в читабельном виде. В связи с этим, встает вопрос – а нельзя ли создать структуру данных, не требующую переделок при появлении новых сущностей, позволяющую хранить произвольную информацию и при этом достаточно простую и эффективную? Чудес, увы, не бывает. Любая универсальная система менее эффективна, чем специализированная. Однако возможно создание решения, сочетающего приемлемую производительность и простоту с достаточной степенью универсальности.

2. МОДЕЛЬ БАЗЫ ДАННЫХ

Сформулируем основные принципы, на которых строиться наша модель данных.