

**В.Л. МАЛЫХ,**

к.т.н., заведующий лабораторией Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия

Я.И. ГУЛИЕВ,

к.т.н., руководитель Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия

А.Н. КАЛИНИН,

м.н.с. Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия

А.В. КОЛУПАЕВ,

инженер-исследователь Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия

С.Г. ЮРЧЕНКО,

м.н.с. Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ РЕЧЕВОГО ИНТЕРФЕЙСА И СИСТЕМ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ В МИС

УДК 61:007, 519.711.3

Малых В.Л., Гулиев Я.И., Калинин А.Н., Колупаев А.В., Юрченко С.Г. *Возможности применения речевого интерфейса и систем автоматической обработки текстов в МИС* (Институт программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия)

Аннотация. Статья посвящена рассмотрению концептуальных возможностей применения речевого интерфейса и систем автоматической обработки текстов в современных МИС. Рассматриваются программные прототипы подобных систем. Статья будет полезна архитекторам и разработчикам современных МИС.

Ключевые слова: медицинские информационные системы, системы автоматической обработки текстов, работа на ограниченном языке, квазиестественный командный язык, речевой интерфейс.

UDC 61:007, 519.711.3

Malykh V.L., Guliev Y.I., Kalinin A.N., Kolupaev A.V., Jurchenko S.G. *The possibility of using a speech interface and automatic text processing systems in MIS* (Ailamazyan Program Systems Institute of RAS, Pereslavl-Zalessky, Russia)

Abstract. The article is devoted to possibility of using a speech interface and automatic text processing systems in modern MIS. Describes software prototypes of such systems. The article will be useful for architects and developers of modern MIS.

Keywords: medical information systems, automatic text processing systems, controlled language, quasi-natural command language, speech interface.

Введение

Системы автоматической обработки текста начали свою историю с 60-х годов XX века и непрерывно развиваются, используя возможности современных компьютеров, технологий и языков программирования [1]. Несмотря на все приложенные усилия, естественно-языковой интерфейс для компьютеров все еще находится в стадии экспериментальных лабораторных



разработок, не сулящих быстрых практически значимых результатов. Сложность проблемы компьютерного анализа естественного языка хорошо известна. Возможные на данном этапе научного развития практические решения лежат в переходе к работе на ограниченном языке (Controlled Language), к переходу к квазиестественным языкам, отказу от «глубокой» семантической разборки (анализа) текста на естественном языке. Проблемно-ориентированные системы автоматической обработки текстов могут легче поддаваться практической реализации, чем универсальные системы. Мы считаем, что на данном этапе технологического развития эксперименты в МИС с речевым интерфейсом и системами автоматической обработки текстов вполне оправданы и актуальны. Заметим, что на передний край науки выходят еще более смелые исследования, связанные с построением интерфейсов мозг-компьютер [2]. В работе рассматриваются эксперименты, выполненные в 2008–2013 гг. в Исследовательском центре медицинской информатики ИПС им. А.К. Айламазяна РАН, связанные с изучением возможностей практического применения речевых интерфейсов и систем автоматической обработки текстов в МИС.

Ввод лечебно-диагностических назначений в МИС на естественном языке

Ввод лечебно-диагностических назначений в МИС требует немалых временных затрат от врачей — пользователей МИС. Связано это с тем, что врач делает назначение по форме, но не по содержанию, «не свободно». Для аналитических целей в системе требуется разбирать назначения на отдельные семантические элементы и согласовывать значения этих элементов со словарями системы. Например, назначаемые лекарственные средства могут выбираться исходя из их наличия в аптечке отделения или в аптеке медицинской организации (МО), количество и едини-

цы дозировки, а также периодичность применения должны быть выделены и разобраны на элементы, чтобы на основании исполненных назначений автоматически можно было сформировать документ о материальном расходе. При оказании платных услуг необходимо выбирать назначаемые услуги из соответствующих прейскурантов. Контроль качества лечебно-диагностического процесса на соответствие стандартам оказания медицинской помощи требует ассоциирования назначений с элементами стандартов. Чтобы удовлетворить всем этим требованиям, врач вынужден вводить отдельные семантические элементы назначения в отдельных полях, пользоваться различными списками выбора. Для облегчения и ускорения работы врача можно применять различные шаблоны назначений, но все равно при этом сохраняется компьютерная парадигма ввода, резко отличающаяся от рукописной или речевой формы задания назначений. При наблюдении за работой врачей в системе было неоднократно замечено, что многие врачи пытаются «сохранить» привычную для них рукописную форму ввода назначений. Например, в поле, предназначенное для ввода наименования лекарственного средства, пытаются записать текст медикаментозного назначения целиком, игнорируя другие поля ввода, оставляя их пустыми или заполненными значениями по умолчанию. С когнитивной точки зрения для врача оказывается удобней и привычней ввести формулировку назначения неразделенным целостным текстом. Нами была поставлена и исследована следующая проблема: Возможно ли вводить лечебно-диагностические назначения в систему в виде ограниченно свободного текста, с практически достаточной надежностью автоматически разбирать текст формулировки назначения, удовлетворив при этом аналитические требования системы? Было решено в исследовательских целях разработать архитектуру для решения поставленной задачи и провести численные эксперименты



на реальных данных. Основная надежда на успешное решение задачи заключалась в том, что мы имеем дело с профессиональным ограниченным языком. Для возможных значений различных семантических элементов текста на ограниченном языке имеются или относительно легко могут быть созданы словари. В контексте уже определенных по словарям семантических элементов другие семантические элементы могут быть найдены, исходя из ряда правил, исходя из представлений о взаимном позиционировании семантических элементов, или, другими словами, о порядке слов в данном тексте.

Поясним идею алгоритмов автоматической обработки текстов назначений на примере медикаментозных назначений. Семантически в формулировке медикаментозного назначения в общем случае можно выделить следующие элементы: наименование лекарственного средства, форму выпуска, концентрацию (для растворов), предписанную дозировку (единицы измерения дозировки и количество), кратность применения (величину периода и кратность исполнения назначения за период), способ введения, дополнительную инструкцию по применению, признак срочности исполнения *sito*. Возможные значения отдельных семантических элементов могут быть сведены в соответствующие словари. В качестве наименования лекарственного средства может выступать торговое наименование зарегистрированного и разрешенного к применению в России лекарственного средства или же его международное непатентованное наименование. Существуют соответствующие электронные фармацевтические справочники, позволяющие сформировать словарь для данного семантического элемента. Фармацевтические справочники — это не лингвистические словари. В них наименование лекарственного средства не выступает как словоформа, мы не найдем в фармацевтических справочниках наименования лекарственного средства в различных падежах, в различном роде

и числе и т.п. Но мы и не предполагаем давать чисто лингвистическое описание значений наших семантических элементов. Разборка и «понимание» текста назначения выполняются не на основе лингвистической модели, но на основе семантической модели. В нашем случае возможное ошибочное написание наименования лекарственного средства также может быть (по нашему мнению, должно быть) включено в словарь в качестве синонима для правильного написания наименования. Либо ошибочное написание должно быть правильно интерпретировано с учетом возможных ошибок в написании, например, с учетом возможности замены, удаления или вставки одной буквы в правильном написании наименования. В качестве второго важнейшего элемента медикаментозного назначения рассмотрим единицу дозировки лекарственного средства. К этому семантическому элементу применима та же логика, что и к наименованию лекарственного средства. Составляется словарь возможных значений единицы дозировки. Даже с учетом синонимии он будет небольшим по размеру. В нем мы найдем следующие синонимы: «*таблетка*», «*таблетки*», «*таблетку*», «*таблеток*», «*табл*», «*таб*», «*т*» с возможным добавлением в конце аббревиатур точек. При разборке мы пытаемся выделить единицу дозировки по значению, по совпадению элемента со значением, указанным в словаре. Если это удалось, то далее для определения количества лекарственного средства применяем следующий позиционный принцип (правило). Слева от выделенной единицы дозировки пытаемся выделить число, которое в случае успеха и будем считать предписанным количеством лекарственного средства. При этом можно учесть неоднозначность в написании количества и, например, понимать «*1/2*», «*1\2*», «*0,5*», «*0.5*», «*половина*», «*пол*» как синонимы. В этом случае следующие дозировки будут эквивалентны: «*пол таблетки*», «*1/2 таб.*», «*0,5 т.*». Все сказанное про дозировку верно





и для периодичности применения — формируется словарь для единиц периода и применяется позиционное выделение числа, например, «3 раза в сутки». Аналогичный подход, связанный как с формированием словарей значений семантического элемента, так и с позиционированием семантических элементов относительно уже выделенных элементов, используется для остальных элементов. Например, для способа введения создается соответствующий словарь, в котором значения «перорально», «пер ос», «п/о», «п\о» будут являться синонимами. При успешном нахождении таких основных семантических элементов медикаментозного назначения, как наименования лекарственного средства, дозировки и кратности, прочие элементы могут находиться в соответствии с позиционным принципом. Например, дополнительная инструкция выделяется как связанная часть текста, не отнесенная к другим семантическим элементам, позиционно находящаяся в конце текстовой формулировки медикаментозного назначения. Заметим, что в предложенном подходе к формированию словарей значений семантических элементов будет не всегда возможно однозначно интерпретировать значение. Например, единица дозировки «кап.» может означать и каплю, и капсулу в зависимости от контекста — формы выпуска лекарственного средства. К счастью, содержательная интерпретация назначения делается не машиной, а исполняющим его человеком (медсестрой), что исключает неоднозначность в интерпретации назначения. Семантический разбор медикаментозного назначения должен также основываться на ряде правил, выделенных в ходе разбора реальных назначений. Мы приводим ниже реальные примеры медикаментозных назначений, написанные непосредственно врачами и хранящиеся в БД МИС. Отметим следующие особенности этих назначений: **1)** часть семантически значимых элементов может быть опущена; **2)** дозировка и кратность могут задаваться с помощью числовых

диапазонов; **3)** дозировка и кратность могут задаваться «курсовым» выражением.

- «Диклофенак на ночь», «Церукал раствор внутримышечно ситуационно» — в назначениях опущена дозировка лекарственного средства.

- «Хумулин Р раствор Подкожно 8–5–8» — дозировка и кратность задаются «курсовым» выражением. Назначение выполняется 3 раза в день в дозировке по 8, 5 и еще раз 8 единиц, соответственно. Единицы дозировки опущены, сравни с «Хумулин Р раствор Подкожно 8–5–8 ед».

- «Ново-пассит Перорально по 10–15млх 3–4р/д» — Дозировка и кратность задаются числовыми диапазонами, символ «х» играет роль разделителя (знака умножения). В полной формулировке назначение выглядело бы так: «Ново-пассит перорально по 10–15 миллилитров 3–4 раза в день».

- «Вит.С 4,0 в\в стр.» — Ошибка в сокращенном наименовании лекарственного средства (имеется в виду витамин С), сравни «Вит.С 5,0 в\в стр.», отсутствуют единицы дозировки, способ введения — «внутривенно струйно».

Анализ примеров (прецедентов), которые порождают трудности в разборе, позволяет выделять и формулировать новые правила разборки и повышает качество разборки. Приведем логику разборки одного из вышеуказанных назначений: «Ново-пассит Перорально по 10–15млх3–4р/д». Числа или числовые диапазоны предварительно отделяются от примыкающих к ним символов. Текст назначения приобретает после такого разделения следующий вид: «Ново-пассит Перорально по 10–15 млх 3–4 р/д». «Ново-пассит» — выделяется как наименование лекарственного средства, исходя из наличия такого значения в словаре лекарственных средств. Способ введения лекарственного средства «Перорально» выделяется по словарю способов введения. Единица дозировки «млх» выделяется как синоним основной единицы «мл.» с



использованием словаря единиц дозировки и правила, позволяющего «приписывать» символ «х» в качестве последнего символа к словарным значениям единиц дозировки. Единица периода «р/д» («раза в день») выделяется с использованием словаря единиц периода. Относительно выделенных единиц дозировки и периода выделяются числовые характеристики дозировки и кратности. Предлог «по» считается семантически неразбираемым элементом. Окончательный итог: текст назначения семантически полностью разобран и правильно «понят».

Рассмотрим чуть более сложный случай — назначение медикаментозного комплекса. В медикаментозном комплексе присутствуют одновременно несколько компонентов — лекарственных средств, каждый компонент в своей дозировке. Все вышеуказанные принципы разборки медикаментозного назначения переносятся и на медикаментозный комплекс. Основное отличие комплекса в том, что в нем, используя словари, надо находить несколько лекарственных средств с отдельной дозировкой для каждого из них. Для упрощения разборки можно потребовать, чтобы компоненты комплекса отделялись друг от друга с помощью некоторого разделителя, например, символа «+». Ограничительной мерой для повышения качества разборки являются формальные требования к используемому ограниченному языку.

Описанный подход переносится и на другие типы назначений, например, диагностические, инструментальные и лабораторные исследования. Важно, чтобы тип назначения был явно специфицирован перед разборкой, так как каждому типу может соответствовать своя семантическая модель и свои соответствующие словари значений семантических элементов и правила разборки. Основным отношением на элементах словарей будет отношение синонимии, но не только оно. Элементы словарей могут быть связаны с другими справочниками системы, например, с

прейскурантами платных услуг или с элементами стандартов оказания медицинской помощи, с нозологиями, с исполняющими лечебно-диагностические назначения подразделениями и т.п. Фактически речь идет о формировании тезауруса предметной области.

В общем случае семантическая модель назначения описывает все отдельные семантические элементы, которые могут присутствовать в тексте назначения. Для семантических элементов, определяемых «по значению», указываются соответствующие словари, а также определяются правила для выделения семантических элементов. Одни правила могут основываться на знании порядка семантических элементов в тексте. Примеры таких правил: **1)** «Числовое значение или числовой диапазон размерной физической величины стоит слева от единиц измерения»; **2)** «В медикаментозных назначениях одиночный символ 'х' или 'Х', за которым следует число или числовой диапазон, стоит перед семантическим элементом кратность»; **3)** «При ненахождении в медикаментозном назначении наименования лекарственного средства по словарю лекарственных средств и нахождению семантического элемента дозировка весь текст от начала до элемента дозировка считать наименованием лекарственного средства». Другие правила могут быть связаны с семантической интерпретацией элементов текста в данном контексте. Примеры таких правил: **1)** «В медикаментозном назначении группа чисел, разделенная знаком '-' при отсутствии выделенных единиц дозировки и кратности, считается курсовым выражением, задающим дозировку лекарственного средства для одного дня»; **2)** «В назначении медикаментозного комплекса знак '+' отделяет один компонент комплекса от другого». Третьи правила учитывают особенности поиска семантического элемента «по значению». Пример такого правила: «Если сразу несколько элементов из словаря значений семантического элемента найдено в тексте, то предпоч-





тение отдается значению большей длины». Четвертые правила описывают предварительную обработку текста. Примеры таких правил: **1)** «Выделить числа и числовые диапазоны в тексте пробелами»; **2)** «Подавить в тексте все незначащие пробелы». Чем большее число правил удастся выделить и учесть при разборке — тем «умнее» будет наш разборщик, будет уменьшаться вероятность того, что текст не удалось разобрать, но при этом одновременно может расти вероятность неверной разборки текста. Самым лучшим подходом, обеспечивающим качество разборки, является выдвижение жестких формальных требований к используемому ограниченному языку и необходимость следовать этим требованиям. Пример таких формальных требований для назначений медикаментов: **1)** никогда не опускать количество и единицы дозировки; **2)** названия лекарственных средств всегда употреблять в именительном падеже; **3)** всегда использовать разделитель «+» при задании медикаментозного комплекса; **4)** не допускать «склеивания» в тексте отдельных семантических элементов. Следует заметить, что при использовании речевого интерфейса для ввода назначений вместо прямого текстового ввода с клавиатуры часть проблем, связанных с семантической разборкой, уйдет. Это объясняется тем, что в речи значения семантических элементов будут проговариваться, как правило, полностью без сокращений, не будут возникать возможные на письме случаи сцепления элементов между собой. При условии надежной разборки речи речевой ввод может быть предпочтительнее письменного.

В Исследовательском центре медицинской информатики ИПС им. А.К. Айламазяна РАН был разработан программный прототип для разборки текстовых формулировок назначений медикаментов и медикаментозных комплексов. На основании имеющегося в распоряжении исследователей объемного фактографического материала были построены словари для следующих семантических

элементов: **1)** наименования лекарственного средства; **2)** единиц дозировки; **3)** единиц периода; **4)** способа введения. Кроме того, был построен словарь для нестандартного задания числовых значений («пол», «1\2», и т.п.). Был выделен, сформулирован и программно реализован ряд правил разборки. Численные эксперименты показали, что при условии полноты формулировки и успешного выделения вышеуказанных семантических элементов по словарям можно достичь практически 100% семантически верной разборки текста назначения. Проблемы начинаются тогда, когда в разбираемом тексте появляются значения семантических элементов, не присутствующие в словарях, или когда возникает необходимость во введении в систему нового правила разборки. Поэтому для получения практически значимого результата совершенно необходимо не только построить некоторый проблемно-ориентированный ограниченный язык и реализовать разборщик текстов на этом языке, но и включить в архитектуру системы механизм актуализации и концептуализации новых знаний.

Несколько слов о программной реализации разборщика. Разборщик строится на основе теории категорий. Разбираемые тексты соответствуют объектам теории категорий. Соответствующие правила разборки оформляются в виде функций, соответствующих морфизмам теории категорий. Применение функции к объекту либо оставляет объект неизменным (автоморфизм), либо модифицирует объект. Например, успешное выделение по некоторому правилу какого-либо семантического элемента приводит к модификации — вставке в разбираемый текст разметки (тегов), выделяющих данный семантический элемент. Множества объектов (текстов) объединяются в классы, отражающие уже выделенную и известную для объектов данного класса семантику. Например, класс объектов, у которых выделены наименование лекарственного средства и дозировка лекарственного сред-

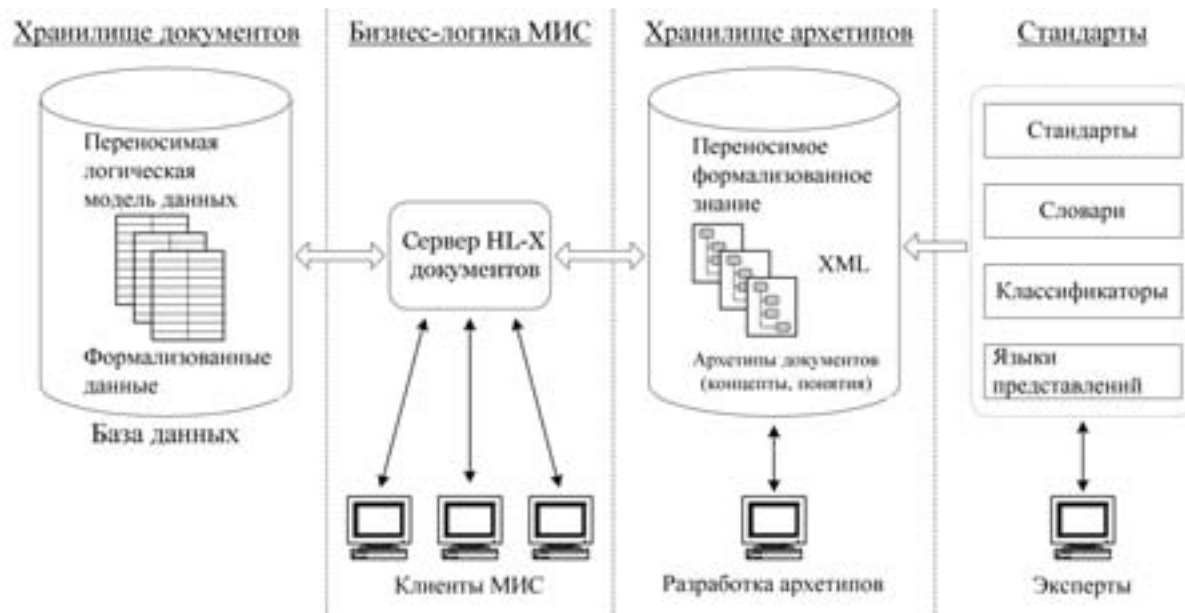


Рис. 1. Эволюционный процесс концептуализации предметной области в архитектуре HL-X

ства. Для каждого класса определяется упорядоченное множество морфизмов (функций), применимых ко всем объектам данного класса. Используемые при разборке правила помечаются. Объект становится терминальным, когда к нему применены все возможные морфизмы, на этом разборка заканчивается. Допускается возможность разветвлений, когда семантический элемент выделяется неоднозначно, и далее рассматриваются все варианты возможной дальнейшей разборки. В этом случае разборщик выдаст множество терминальных объектов, из которого по некоторым критериям можно будет сделать автоматически окончательный выбор либо передать решение о выборе наиболее предпочтительной семантически правильной разметки в руки пользователя.

Актуализация и концептуализация знаний, формирование тезауруса предметной области

В более ранних наших работах мы уже отмечали целесообразность и необходимость

встраивания механизма концептуализации предметной области в архитектуру МИС [3]. Цитируем: «Основная концептуальная идея архитектуры ... введение процесса концептуализации предметной области непосредственно в саму информационную систему». Ниже приведен рис. 1 из работы [3], иллюстрирующий процесс концептуализации предметной области.

Процесс концептуализации в указанной работе рассматривался только с точки зрения построения моделей клинических документов МИС. Справедливо отмечалось, что клинические документы являются важнейшими источниками знаний о предметной области. Теперь эту точку зрения можно расширить. К источникам знаний — моделям клинических документов следует добавить знания, заключенные в унифицированной модели лечебно-диагностического процесса в целом [4, 5]. Отметим, что описанные выше семантические элементы рассмотренных текстов — это те же понятия и концепты предметной области. Рассматривая частную проблему семантической разборки текста, мы неизбежно сталкиваемся



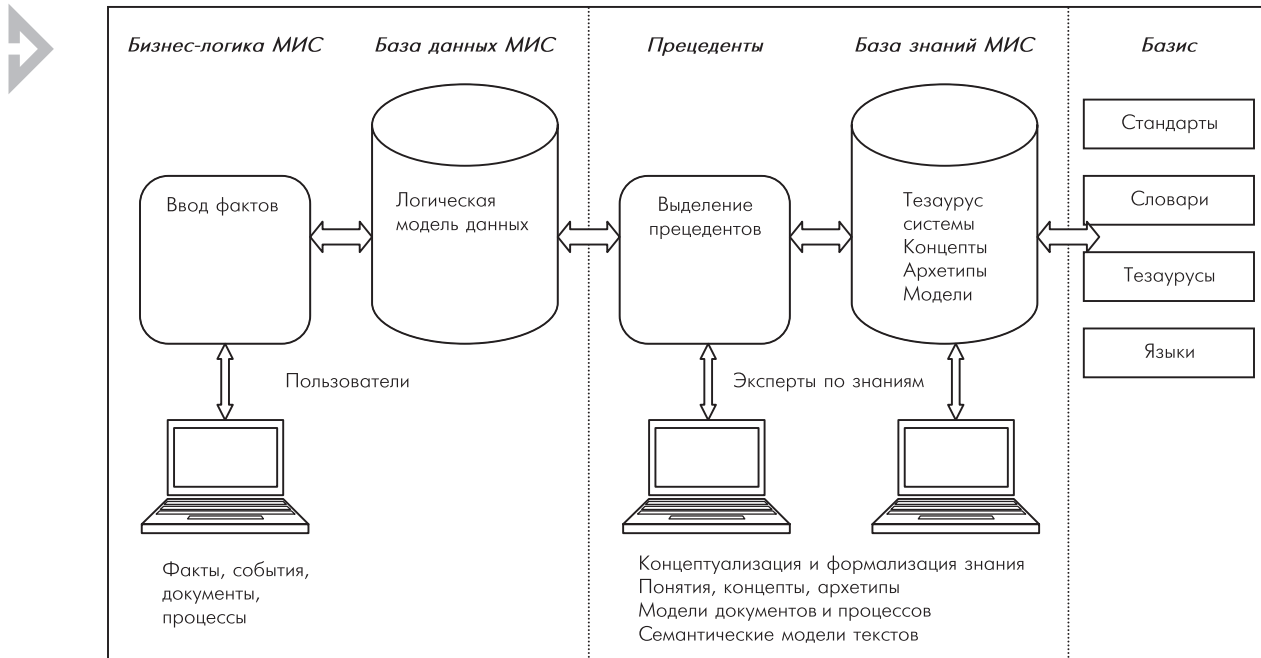


Рис. 2. Эволюционный процесс концептуализации и формализации предметной области в МИС

с общесистемными проблемами. Речь идет о необходимости усвоения и добавления в систему новых знаний, о построении в системе непрерывного процесса концептуализации и формализации нового знания. Оказывается, что можно предложить достаточно общий подход для решения этой общесистемной проблемы. Этот подход, получивший название прецедентного подхода, изложен в работе [6]. Появление нового прецедента в МИС равносильно появлению новых знаний в системе, за которым следует возникновение соответствующих событий, например, уведомления о необходимости контроля нового прецедента: необходимости оценки корректности прецедента, необходимости концептуализации и ввода в систему новых знаний для корректных прецедентов. Появление новых знаний в системе тут же открывает возможности их немедленного функционального использования. Для клинических документов прецеденты могут заключаться в

появлении новых документов, в появлении в структуре документа ранее не формализованных понятий и концептов. В лечебно-диагностическом процессе новые прецеденты — это ранее не наблюдавшиеся события (факты). Для рассматриваемой нами проблемы семантической разборки текста новый прецедент — это новое, ранее не наблюдавшееся значение семантического элемента текста или ранее не наблюдавшаяся совокупность значений некоторого множества семантических элементов. Обработка нового прецедента семантической разборки заключается в пополнении соответствующих словарей и тезауруса системы, а также в формулировке новых или уточнении старых правил разборки. Для семантической разборки медикаментозных назначений новые прецеденты приведут к пополнению словарей новыми значениями для наименования лекарственного средства, дозировки, кратности применения, способа введения и т.д. На рис. 2 пред-



ставлена расширенная схема встроенного в МИС процесса концептуализации и формализации предметной области.

Представленная на *рис. 2* архитектура была практически апробирована в МИС Интерин PROMIS и полностью оправдала все возлагаемые на нее надежды.

Использование командного квазиестественного языка и речевого интерфейса в МИС

Речевой интерфейс является одним из интерфейсных каналов, по которому в систему могут вводиться тексты. Тексты могут быть достаточно свободными, не подлежащими семантической разборке и интерпретации. Тексты могут использовать ограниченный проблемно-ориентированный язык, высказывания на таком языке потребуются семантически разбирать и интерпретировать. Примером такого ограниченного языка были рассмотренные выше медикаментозные назначения.

Хорошим примером использования свободного естественного языка в МИС является заполнение многочисленных клинических документов: осмотров, дневников, диагностических протоколов, эпикризов и т.п. В этих документах, наряду с формализованными и выбираемыми из списков значениями, присутствуют поля для ввода свободного текста: анамнеза, жалоб и др. Использование речевого интерфейса для заполнения клинических документов при условии приемлемого качества распознавания слов, в первую очередь профессиональных терминов, может оказаться вполне востребованным. Здесь мы неизбежно столкнемся с проблемой задания пунктуации при речевом вводе и укажем давно известное ее решение. Каждый, кто диктовал текст машинистке в докомпьютерную эпоху, помнит, что знаки пунктуации просто проговаривались: «точка», «запятая», «тире» и т.п. Естественно, что при использовании речевого интерфейса для заполнения в МИС клинического документа нам захочется ввести

в речевой интерфейс речевые команды навигации по полям и управляющим интерфейсным элементам электронного документа. Поэтому следует сразу рассматривать общий случай работы с командным квазиестественным языком (КЕЯ) или языками.

Очевидно, что речевой интерфейс — это один из возможных, но не единственный из интерфейсных каналов ввода информации в систему на командном квазиестественном языке. Возможности практического применения речевого интерфейса в МИС будут определяться исключительно качеством и удобством данного интерфейса. Организовать такой интерфейс технологически не составляет большого труда. В настоящее время промышленно выпускается целый спектр мобильных устройств (мобильные телефоны, планшеты и ноутбуки), обладающих возможностями распознавания речи. Нами в исследовательских целях был реализован программный прототип такого речевого интерфейса с МИС Интерин PROMIS. Для планшета с OS Android с уже имеющейся на нем программой распознавания речи была написана программа, осуществляющая распознавание «речевых пакетов» и их передачу в распознанном текстовом виде совместно с контекстной информацией (для какого пользователя и приложения предназначен пакет) через беспроводную сеть Wi-Fi в базу данных МИС. В МИС выполнялась семантическая разборка поступающих речевых пакетов с учетом указанного контекста. Разборщик выделял в блоках команды и элементы разметки, превращал понятия команды на квазиестественном языке в команды для программы — интерпретатора команд. Далее программа — интерпретатор на клиентской машине в реальном времени интерпретировала (исполняла) построенную разборщиком программу. Интерпретатор команд мог эмулировать работу клиента в клиентском интерфейсе, формируя последовательности ввода символов с клавиатуры, или команды манипуляции с мышью. Интер-



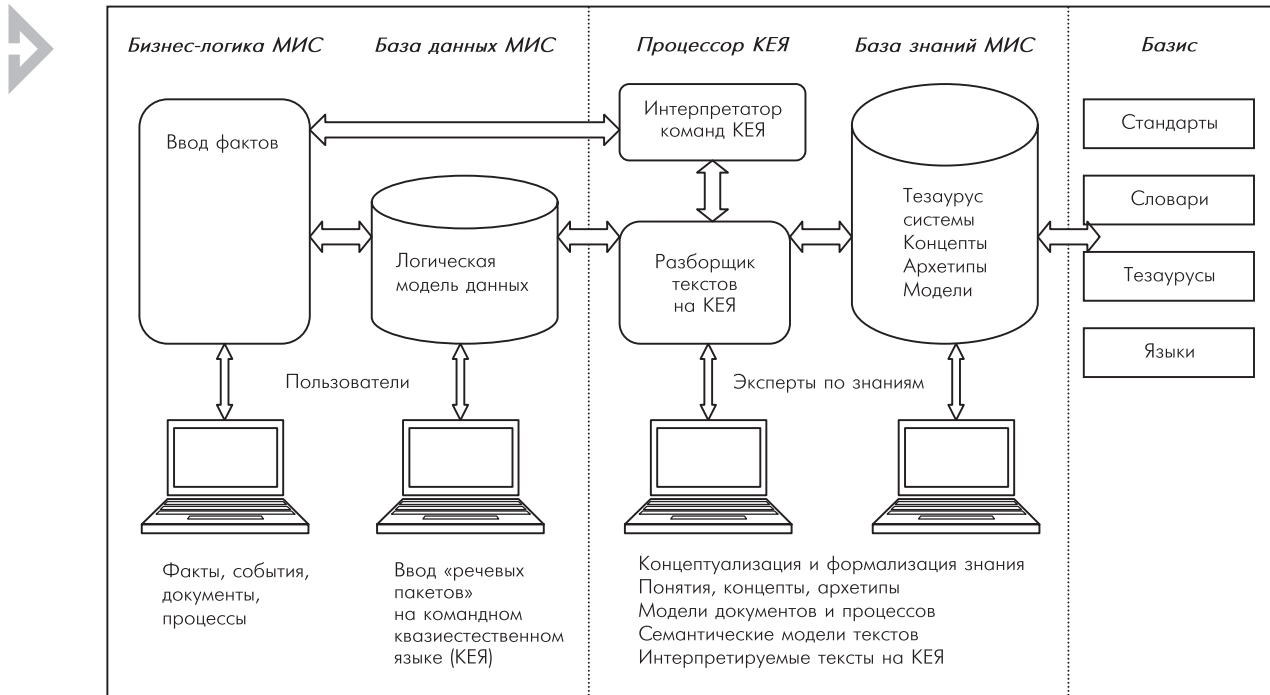


Рис. 3. Архитектура обработки текстов на командном квазиестественном языке в МИС

прематор мог обращаться непосредственно к БД МИС, формируя транзакции для сервера БД. Для пользователя все это выглядело как быстрая реакция системы на поток его речевых команд, выбор нужного поля ввода, ввод в поле произнесенного текста с указанной пунктуацией, нажатие на форме различных кнопок, управление другими элементами интерфейса. Описанная архитектура представлена на *рис. 3*.

Очевидно, что речевой интерфейс не является единственным каналом ввода в систему текстов на КЕЯ. Готовые к интерпретации программы на КЕЯ могут запускаться непосредственно пользователями из любого ресурса: файла, БД. Возможности КЕЯ в МИС не исчерпываются только лишь вводом информации в клинические документы или вводом лечебно-диагностических назначений. Очень перспективным выглядит формирование на КЕЯ программ тестирования. В ходе тестирования программных модулей зачастую

тестировщиком выполняются (имитируются) последовательные действия пользователя в рамках определенного бизнес-процесса. Если эти выполняемые действия можно выразить на КЕЯ в виде интерпретируемой программы, то ведение процесса и ввод данных при тестировании можно будет поручить автомату, а пользователю останется только запускать программы тестирования и визуально контролировать процесс тестирования. Интерпретируемые программы на КЕЯ, дополненные возможностями остановки хода выполнения программы, возможностями диалога с интерпретатором, возможностями отката уже выполненных команд, могут также сыграть важную роль при обучении пользователей. Программа на КЕЯ будет выступать в роли учителя, обучая пользователей последовательности действий, позволяя прерывать исполнение программы, позволяя возвращаться и повторно исполнять шаги процесса. Использование квазиестественного языка



существенно упростит подготовку программ для тестирования и обучения.

В данный момент мы можем лишь в самых общих чертах указать на возможности использования командных квазиестественных языков в МИС и предложить лишь самые общие архитектурные принципы построения таких систем. При обсуждении проблемы речевого интерфейса и текстового интерфейса с врачами выяснилось, что врачи видят определенное удобство и целесообразность в возможности независимого (вне МИС) ввода текстовой информации и речи на различные мобильные устройства с последующей передачей этой информации в МИС. Командный квазиестественный язык также заинтересовал врачей, в первую очередь им хотелось бы автоматизировать с помощью КЕЯ часто выполняемые в системе действия: поиск медкарты, печать документа, запуск отчета.

Заключение

В работе исследованы возможности применения систем автоматической обработки

текстов в МИС. Основным выводом работы в том, что проблемно-ориентированные системы автоматической обработки текстов на ограниченном языке могут найти практическое применение в МИС. В первую очередь такие системы могли бы обеспечить прямой текстовый ввод лечебно-диагностических назначений. Возможно обеспечить прямой речевой ввод клинических документов на не ограниченном свободном профессиональном языке, включающем в себя знаки пунктуации. Системы автоматической обработки текстов в МИС обязательно должны быть поддержаны системой концептуализации и актуализации знаний, построенной на основе выделения прецедентов нового знания. Перспективна разработка командного квазиестественного языка для целей реализации речевого интерфейса с МИС, а также для разработки интерпретируемых программ на КЕЯ, решающих задачи тестирования и обучения. Разработанные в ходе выполнения работы программные прототипы и проведенные численные эксперименты подтверждают эти выводы.

ЛИТЕРАТУРА



1. Мальковский М.Г., Грацианова Т.Ю., Полякова И.Н. Прикладное программное обеспечение: системы автоматической обработки текстов. — М.: Издательство МАКС «Пресс»; Издательский отдел факультета вычислительной математики и кибернетики МГУ им. М.В. Ломоносова, 2000.
2. Фролов А.А., Рощин В.Ю. Интерфейс мозг-компьютер. Реальность и перспективы. Лекции по нейроинформатике 2008. <http://neurolectures.narod.ru/>.
3. Гулиев Я.И., Малых В.Л. Архитектура HL-X поддержки документов в медицинских информационных системах//Информационно-управляющие системы. — 2009. — № 2. — С. 63–69.
4. Малых В.Л., Гулиев Я.И. Моделирование лечебно-диагностического процесса в классе управляемых стохастических процессов с памятью//Врач и информационные технологии. — 2013. — № 2. — С. 6–15.
5. Малых В.Л., Гулиев Я.И. Управляемый стохастический прецедентный процесс с памятью как математическая модель лечебно-диагностического процесса//Информационные технологии и вычислительные системы. — 2014. — № 2. — С. 62–72.
6. Малых В.Л., Гулиев Я.И. Прецеденты в медицинских информационных системах// Программные продукты и системы. — 2009. — № 2(86). — С. 19–27.