

## МЕТОДИКА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Рассмотрена методика организации интеллектуального анализа в прикладных системах в медицине. Основной идеей статьи является применение нового подхода при нечетком логическом анализе «сырых» данных в информационных системах. Исследования проводятся на базе реальной разработки информационной системы «Электронная история болезни» и модуля системы поддержки принятия решений «В помощь молодому специалисту».

### Введение

Информационное обеспечение здравоохранения является на сегодня актуальной и приоритетной задачей. Медицинские компании и министерства здравоохранения многих стран мира вкладывают большие средства для повышения уровня эффективности и технологичности этой отрасли, снижая при этом остальные накладные расходы.

С другой стороны, разработчикам приходится учитывать специфику предметной области, поскольку требуется принятие мер по аудиту и контролю работы, что требует расширения штатов и увеличения заработной платы сотрудникам, обеспечивающим работу подобных систем. Более того, необходимо проводить работу по управлению документацией и вести ее архивирование.

Но и, собственно, лечение и диагностика требуют повышенного внимания и точности, поскольку информационные потоки очень большие и они плохо структурированы. В ходе лечебно-диагностического процесса специалисты передают друг другу большое количество сведений об объекте этого процесса – пациенте. Информация о состоянии пациента и ходе его лечения, которой обмениваются между собой медики, составляет лечебно-диагностический процесс и обеспечивает верную последовательность действий. Ведение истории болезни, или, выражаясь более точно, документирование лечебно-диагностического процесса, является базисным информационным процессом в любом медицинском учреждении.

Однако документирование лечебно-диагностического процесса не решает задачу диагностики и лечения, поскольку информация о пациенте должна быть обработана и проанализирована. Поэтому специалисту может потребоваться профессиональная помощь: произвольная выборка для статистических наблюдений, справочная информация, подсказка в той или иной ситуации. И если медицинские учреждения действительно намерены в результате информатизации получать актуальные и достоверные данные, необходимые для принятия решений, то создание подобных информационных систем должно быть главным направлением вложений средств на сегодня.

В данной статье описывается разработка комплексной методики сбора, обработки и анализа данных, реализуемой в информационной системе (ИС) «Электронная история болезни» и подключаемом модуле системы поддержки принятия решений (СППР) «В помощь молодому специалисту». Методика разрабатывается с учетом нечетких начальных условий вывода. ИС «Электронная история болезни» разрабатывалась с учетом требований специалистов областной больницы им. Н. Н. Бурденко. Программная реализация ИС

[1] внедрена на кафедре хирургии Медицинского института Пензенского государственного университета, а также спроектирован модуль СППР «В помощь молодому специалисту». В роли молодого специалиста выступает начинающий доктор, которому требуется квалифицированная консультация.

### **1 Постановка задачи**

Для разработки экспертных систем не существует четких методик и алгоритмов. Каждая предметная область имеет специфику, учет которой всегда требует нестандартных решений и вычислительных приемов.

Поскольку исходная предметная область – медицина – требует особой точности и оперативности принимаемых решений, целью данной работы является проектирование и реализация ИС, которая обеспечивает:

- 1) надежное хранение и оперативный доступ к информации;
- 2) разграничение прав и защиту от несанкционированного доступа к данным;
- 3) наглядное и интуитивно понятное представление информации в базе данных (БД);
- 4) минимальные трудозатраты по обеспечению работоспособности ИС.

Кроме того, подключаемый модуль СППР должен оказывать поддержку молодым специалистам при принятии решений.

В повседневной практике наиболее вероятен сценарий, когда молодой специалист нуждается в помощи опытного доктора в принятии решения, однако возможности такой нет, а решение принять нужно как можно быстрее. В этом случае помочь могут экспертные системы поддержки принятия решений как хранилище обоснованных и проверенных знаний.

Классическая экспертная система имеет схему, представленную на рис. 1. Как видно, в схеме определены потребитель и источник знаний. Скелет схемы строится на базе знаний (БЗ) и подсистеме логического вывода. Однако при реализации современных систем данная схема претерпевает значительные изменения. Модернизируются как состав экспертной системы, так и алгоритмы, лежащие в основе каждого из звеньев системы. Примером могут служить более эффективные алгоритмы поиска информации в БЗ, распределенные вычислительные системы и системы хранения, на основе которых реализуется подсистема логического вывода и т.д. [2].

В архитектуре модуля «В помощь молодому специалисту» можно выделить следующие особенности:

- возможность ввода информации с нескольких рабочих мест несколькими экспертами;
- возможность онлайн-обсуждения группой экспертов того или иного факта, вносимого в БЗ, что позволяет корректировать БЗ на этапе заполнения;
- решение задачи является открытым процессом, т.е. суждения системы выводятся пользователю – молодому специалисту;
- подсистема логического вывода должна иметь комплексную структуру. Это означает совместное использование сразу нескольких приемов и методов логического вывода – от продукционного вывода до алгоритмов нечеткой логики.

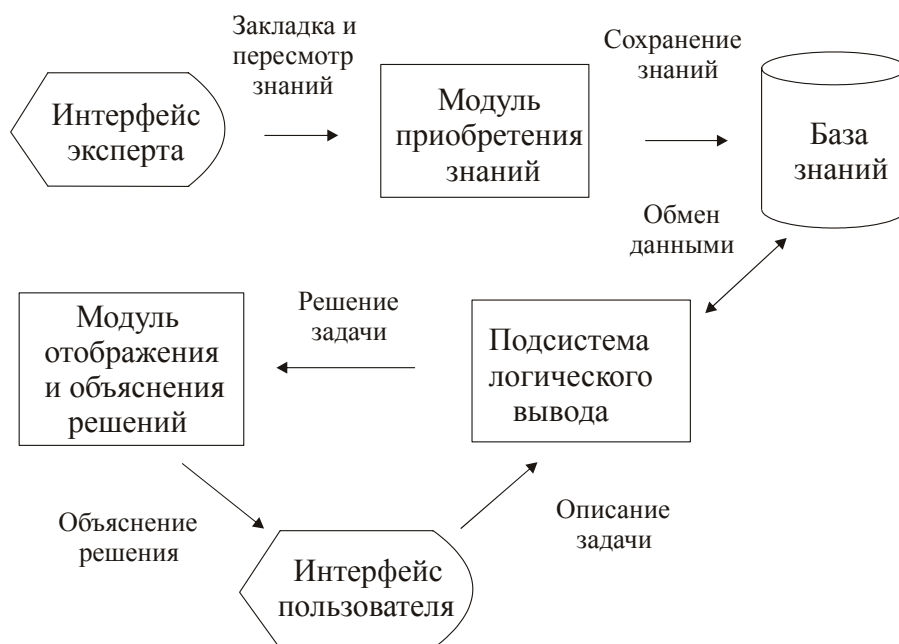


Рис. 1 Модель классической экспертной системы

## 2 Математическое описание модели анализа данных

В данной работе предлагается новый подход, используемый для организации БЗ в виде системы хранения лингвистических описаний для хранения в ИС данных и реализации механизма логического вывода модуля СППР в составе ИС «Электронная история болезни».

Поскольку число комбинаций фактов (симптоматика, результаты исследований и т.п.) во входном наборе ограничено, то для каждого извлеченного параметра возможен ввод соответствующего предиката. Кроме того, для каждого из параметров возможны уточняющие данные: «повышенный» – «сниженный» и т.д., для чего в предикат необходимо ввести свойство, повысив порядок предиката. Например, для пигментации можно выделить не только цвет, но и текстуру.

Модель последовательности работы алгоритма рассматриваемой системы представлена на рис. 2.

На начальном этапе входные данные поступают из БД в виде результирующего набора нечеткого запроса, т.е. запроса, где критерии сопоставляются в общем случае не с конечным значением, а с множеством значений функции принадлежности. Далее результирующий набор разбивается на предикаты высших порядков [3] для того, чтобы облегчить формальное представление исходной задачи перед подсистемой логического вывода. На этапе сопоставления входная цепочка предикатов сравнивается с фактами из БЗ, где представлены накопленные знания СППР «В помощь молодому специалисту» [4]. В слотах фрейма находятся знания, условия применимости которых указаны в виде предикатных выражений. В результате наложения получают существенно больший набор, который будет считаться системой исходных правил на этапе логического вывода.

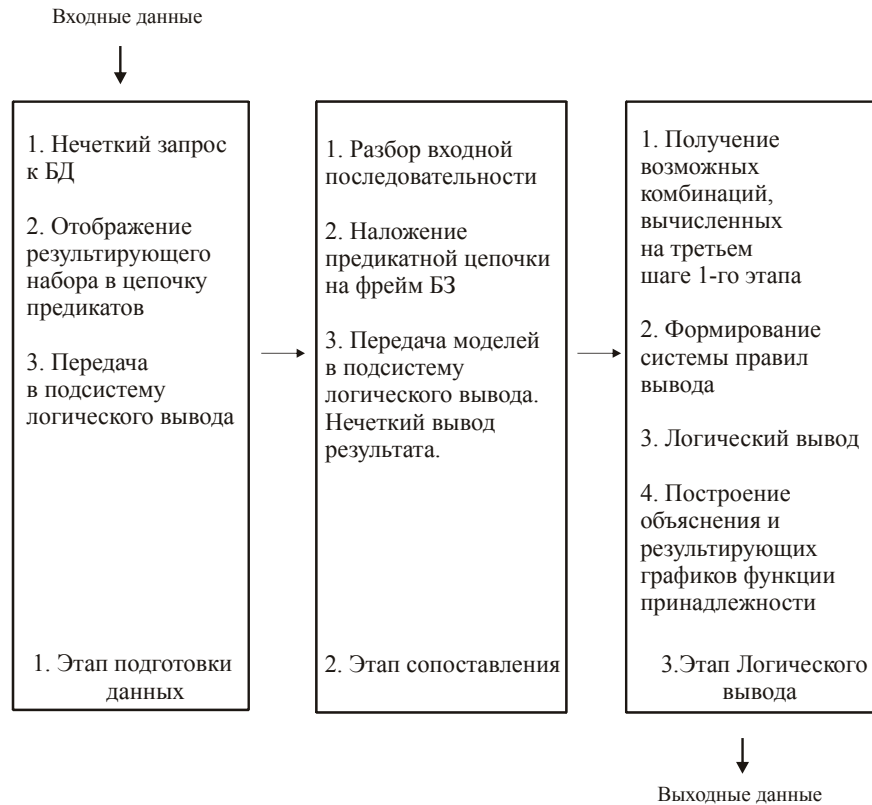


Рис. 2 Последовательность работы алгоритма модуля СППР «В помощь молодому специалисту»

Поиск решения проводится на основе алгоритмов нечеткого вывода. Результат логического вывода является конечным этапом работы модуля поддержки принятия решения «В помощь молодому специалисту». Если в ходе работы результат подтверждается, он записывается в БЗ. Иначе специалист проводит дополнительные исследования и консультации.

Пополнение БЗ может производиться и опытными экспертами средствами ИС.

Для начала определим набор данных, получаемых из БД при нечетком запросе. В случае четкого критерия результирующий набор вполне определенный, но если критерий нечеткий, спрогнозировать данные в результирующем наборе сложно и модель должна учитывать это обстоятельство. Сделав запрос по критериям  $t_1, t_2, t_3, \dots, t_n$ , соответствующих некоторым условиям  $e_1, e_2, e_3, \dots, e_m$ , получают результирующую выборку, представимую в виде набора векторов:

$$t_i = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_l \end{pmatrix}, \quad (1)$$

где  $i = \overline{1, n}$ .

Векторы отражают в общем случае столбцы в полученной выборке и помогают систематизировать данные для представления в виде цепочки предикатов. Каждая цепочка строится с использованием полученных значений  $d_i$ . В результате построения получается  $l$  цепочек, длиной  $n$ :

$$q_j = (p_{j,1}, p_{j,2}, \dots, p_{j,n}), \quad (2)$$

где  $j = \overline{1, l}$ .

Отображение данных, отобранных из БД ИС в предикатную форму, происходит согласно словарю, поскольку необходимо согласование данных в БД и БЗ, где условия применимости того или иного слота заданы в предикатной форме. Построение и ведение такого словаря происходит при внесении новых знаний в БЗ, т.к. СППР должна различать те или иные условия применимости, представленные в предикатной форме. Таким образом, если в БЗ появляется новый слот, то ему в соответствие ставится условие его применимости, а в словарь записываются данные об условиях отображения «сырых» данных и соответствующие им предикаты.

Рассмотрим пример пополнения БЗ. На рис. 3 показан пример фрейма для лекарственных препаратов. Изначально во фрейме знаний «Препарат» присутствует категория в виде слота «Антисептик» и два вложенных слота: «Антисептик 1», «Антисептик 2». Над слотами указаны предикатные условия применимости значений слотов, т.е. показания, при которых эти препараты следует применить.

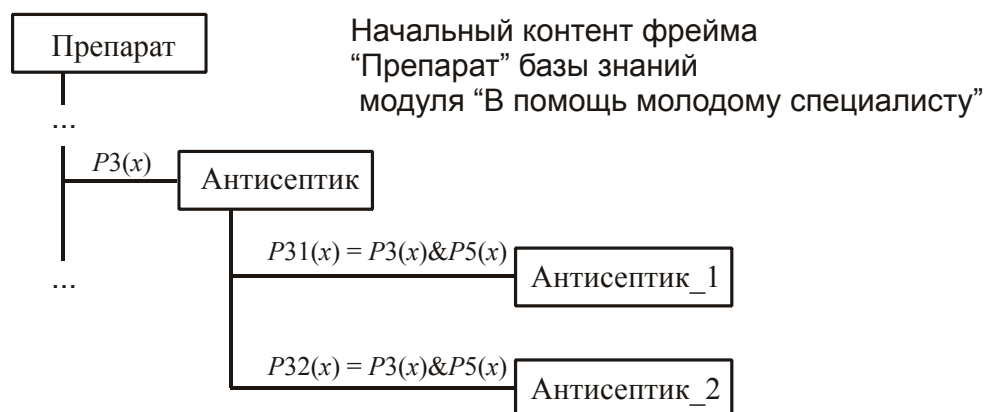


Рис. 3 Графическое описание фрейма «Препарат»

На рис. 4 показано добавление нового слота со значением «Антисептик 3». Из него видно, что при добавлении нового описания препарата класса антисептиков над слотом вносится предикатное выражение-описание (условие применимости).

Вновь добавленные термы  $P33(x)$ ,  $P7(x)$ ,  $P8(x)$  вносятся в словарь, где им в соответствие ставятся лингвистические описания. Поскольку число подобных описаний ограничено для каждой области медицины, их объем не будет слишком большим, и замедления работы всего модуля заметно не будет. Словарь также представляется фреймом (или иерархической таблицей), причем лингвистические описания должны учитывать морфологию слов в

записанных в БД «сырых» данных с использованием механизма регулярных выражений. Фреймовая структура выбрана не случайно, поскольку предикаты высших порядков предполагают некоторый уровень вложенности свойств тех или иных объектов проблемной области. Для цепочек предикатов  $q_i$  можно привести одну из последовательностей, построенную на основе значений параметров из БД. Представим для примера извлеченные из БД выборки о гипертензии одного из больных:

$$\begin{aligned} p_{i,1} &= \exists d(High(d)), \\ p_{i,2} &= \exists f(Pigm(f)) \wedge \exists P(P(f) \wedge property(P) \supset P = 'Red'), \\ p_{i,3} &= \exists w(Large(w)), \end{aligned} \quad (3)$$

где терм  $d$  означает давление, а  $p_{i,1}$  описывает нечеткое выражение «артериальное давление – высокое». В данном случае речь идет не о числовом, а о лингвистическом значении параметра. Это во многом упрощает логический вывод, и получаемый результат становится более понятным [5].

$p_{i,2}$  тождественен предикатному выражению второго порядка, который описывает дополнительное свойство (цвет кожи – красный). В общем случае описание пигментации может быть различным – нормальная, бледная, красная, желтая и т.п. В данном случае вводятся уточняющие данные, а именно свойство, описываемое формулой, входящей в  $p_{i,2}$   $property(P) \supset P = 'Red'$  и означает, что присутствует ненормальная пигментация кожи со свойством «красный».  $p_{i,3}$  означает наличие избыточного веса.

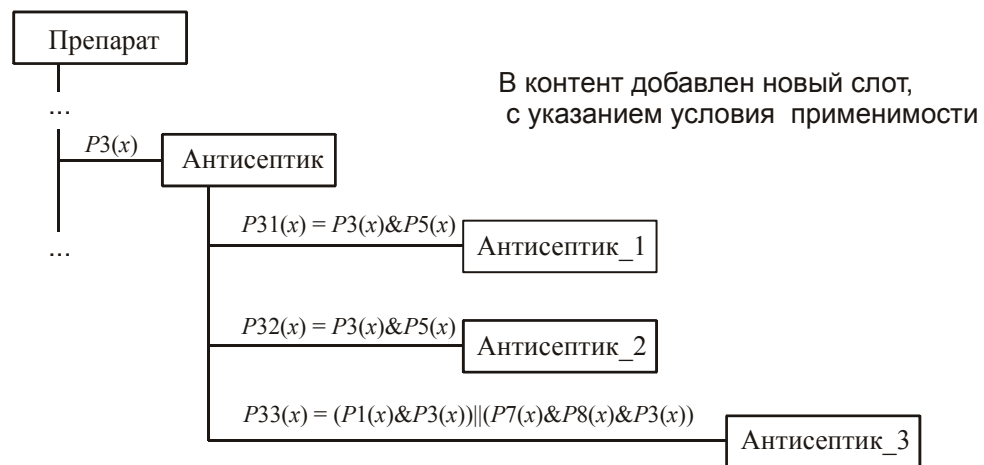


Рис. 4 Добавление нового слота во фрейм «Препарат»

### 3 Описание функциональной модели анализа данных

Любая объектная модель подлежит описанию. Наиболее подходящей спецификацией является UML [6], которая позволяет описать семантику, архитектуру и идеологию модели любой объектной или модульной распределенной [7] системы.

В ходе проектирования изначально выделяются формальные сущности – участники взаимодействия. Поэтому на начальном этапе определим прецеденты разрабатываемой системы – Use Case.

В контексте работы ИС состав внешних субъектов – актеров – ограничивается только «Пользователем», который непосредственно взаимодействует с клиентским приложением для БД.

Для проектируемой ИС предполагается реализация нескольких функций по сбору и обработке данных. На начальном этапе пользователь проходит аутентификацию для доступа к ресурсам. Часто работа начинается с поиска уже введенных данных, поэтому центральной сущностью является «Произвести поиск». Доступ для добавления, редактирования или удаления определяется правами доступа данного пользователя. Назначение прав доступа проводится в административной утилите «Администратор».

Каждый из пользователей может производить поиск и просматривать разрешенные администратором системы записи. Для упрощения данной модели для актера «Пользователь» не указан описатель, содержащий уровень назначенных привилегий.

Сущность «Добавить запись» является начальной функциональной единицей, когда пользователь начинает работу на вновь установленной системе (или системе с заархивированной и очищенной БД). Добавление записи – интерактивный процесс, когда проводится наполнение БД пользователем. Приведенная на рис. 5 диаграмма отражает процесс работы пользователя с ИС «Электронная история болезни».

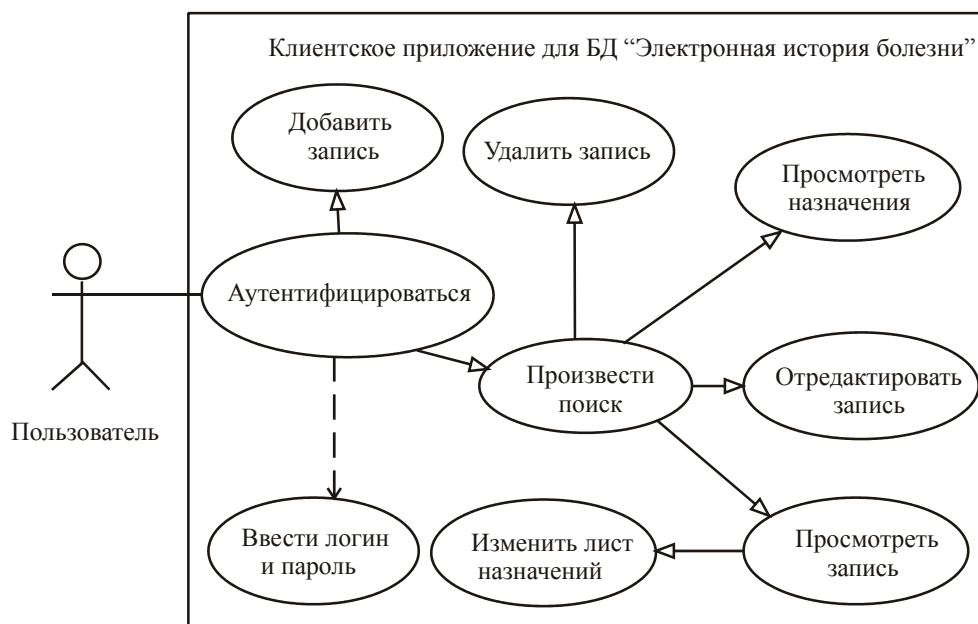


Рис. 5 Диаграмма Use Case для ИС «Электронная история болезни»

На рис. 6 приведена диаграмма прецедентов модуля СППР. Данный модуль является функциональной надстройкой для ИС и предназначен для интеллектуальной обработки данных БД ИС.



Рис. 6 Диаграмма Use Case для модуля СППР «В помощь молодому специалисту»

Для начала определим внешних актеров – это собственно пользователи системы «Эксперт-администратор» и «Молодой специалист» и сама система поддержки принятия решения – «СППР». В контексте работы с СППР роли этих актеров различаются. «Эксперт-администратор» в этом случае получает функцию поддержки работы действующего лица «Молодой специалист» с системой. «Молодой специалист» является потребителем результатов работы актера «СППР» через сущность «Вывести результат». Сущности модели отражают работу алгоритма, реализующего математическую модель логического вывода по исходным данным в виде цепочек предикатов вида (3). Основные этапы функционирования актера СППР представлены сущностями: «Получить запрос от клиента», «Преобразовать входные данные», «Выполнить логический вывод», «Вывести результат». Наличие промежуточных сущностей «Извлечь данные из БД», «Провести наложение на фрейм», «Получить систему правил», «Сформировать входную лингвистическую переменную» и «Выполнить приведение к четкости» в модели необходимы по ряду причин. Во-первых, они отражают специфику работы актеров в этой UML-модели. Во-вторых, детализация диаграммы Use Case помогает более точно описывать объекты всей системы.

Применение добавочных сущностей «Сформировать входную лингвистическую переменную» и «Выполнить приведение к четкости» не всегда целесообразно, поскольку их назначение состоит в получении четких числовых выводов, что может повлиять на решение эксперта. Эти функциональные единицы могут быть реализованы в виде подключаемых внешних модулей. Они применяются по усмотрению пользователя, отраженного в модели актером «Эксперт-администратор».

### **Заключение**

На данном этапе развития средств интеллектуального анализа имеется возможность создания интеллектуальных ИС в любой отрасли.

В результате проведенной работы получена методика построения СППР в области медицины в условиях нечетких исходных данных. Кроме того, разработаны методы перехода от кортежей БД к предикатным выражениям исходной системы логического вывода. Была применена методика представления знаний во фреймах с условиями применимости в виде предикатов высших порядков.

Разработка может применяться в лечебно-профилактических учреждениях, а также представляет методический интерес для специалистов медицинских вузов для использования ИС «Электронная история болезни» при обучении студентов работе с медицинской документацией.

### **Список литературы**

1. **Макаров, М. М.** Использование интеллектуального анализа данных для повышения эффективности использования информации в реляционных базах данных / М. М. Макаров // Сборник научных трудов XXXIII МНТК ПАИИ. – Пенза, 2005. – С. 355–357.
2. **Кузин, Е. С.** Перспективы развития вычислительной техники / Е. С. Кузин, А. И. Ройтман, И. Б. Фоминых, Г. К. Хахалин // Справ. пособие / под ред. Ю. М. Смирнова. – М. : Высш. шк., 1989. – 2 кн. – 160 с.
3. **Такеути, Г.** Теория доказательств / Такеути Г. ; пер. с англ. С. К. Соболева ; под ред. С. И. Адяна. – М. : Мир, 1978. – 412 с.
4. **Кузин, Л. Т.** Основы кибернетики : в 2-х т. : учеб. пособие для вузов. Т. 2 Основы кибернетических моделей / Л. Т. Кузин. – М. : Энергия, 1979. – 584 с.
5. **Вашкевич, Н. П.** Применение методологии и принципов нечеткой логики в информационной системе «Электронная история болезни» / Н. П. Вашкевич, С. А. Зинкин, М. М. Макаров // Вопросы радиоэлектроники. – Вып. 2. – 2007. – С. 5–14. – (Серия ЭВТ).
6. **Буч, Г.** UML : руководство пользователя / Г. Буч, Д. Рамбо, А. Джейкобсон. – М. : ДМК, 2000. – 432 с.
7. **Макаров, М. М.** Виртуальный массив хранения данных / М. М. Макаров // Актуальные проблемы современной науки : труды 5-й Международной конференции молодых ученых и студентов. – Самара : Изд-во СамГТУ, 2004. – 19 ч. – С. 6–7.