

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.89

Каешко
Алексей Игоревич

Модели и средства построения медицинских информационных систем на
основе семантических технологий

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1-31 80 10 «Теоретические основы информатики»

Научный руководитель
Голенков Владимир Васильевич
д.т.н., профессор

Минск 2016

ВВЕДЕНИЕ

Объектом исследования выбрана достаточно обширная предметная область интеллектуальных медицинских информационных систем. Изначально было уточнено понятие интеллектуальности: определено, что интеллектуальность означает способность системы генерировать новую информацию на основе уже известной, то есть увеличивать количество информации введенной в информационную систему. Такой подход к интеллектуальности информационной системы требует особых технологий, применяемых для их создания. Часто интеллектуальность системы заключается в отдельный компонент, обрабатывающий данные, который и называется экспертной системой. Однако результаты работы таких компонентов обратно в систему не поступают, или, что еще хуже такая система работает по принципу «черного ящика». В медицине же не допустимы такие системы, так как окончательное решение принимает медицинский специалист и ему не нужно просто решение с определенной степенью «правильности». Медицинский специалист должен четко видеть логику работы системы.

Таким образом, оптимальная экспертная система для медицины это система, основанная на знаниях. А особенностью таких систем является необходимость накопления этих знаний.

Однако при изучении объекта исследования выяснилось, что до настоящего времени не разработана технология для создания интеллектуальных медицинских систем, основанных на знаниях, позволяющих накапливать знания без значительных трудозатрат.

Несмотря на все возможности и в настоящее время развитие ИМИС находится в самом начале. Существуют отличные примеры решения частных диагностических и клинических задач посредством применения технологий машинного обучения, нечеткой логики, линейных алгоритмов. Однако системы, основанные на знаниях, обладают более широкими диапазонами применения, а появившаяся в начале XXI века технология Semantic Web предвещала революцию в разработке прикладных интеллектуальных медицинских информационных систем, так как четко (как казалось) определяла модель хранения и обработки информации. Медицинские информационные системы нового поколения, очень сложно укладываются в реляционную модель. Как только медицинская система начинает учитывать не только сам факт услуги, но и ее детализацию, то сложность системы резко возрастает, и только система на основе семантической сети может позволить обрабатывать такие данные естественным образом, существенно сокращая трудозатраты на развитие и сопровождение.

Ценность информации, введенной в систему, построенной на основе семантических технологий выше, так как позволяет генерировать новую информацию по определенным правилам, однако стоимость ввода в систему также высока. Возможно, низкая скорость и высокая стоимость ручного ввода информации в семантическую сеть определили малую долю таких медицинских информационных систем в сегменте. В свою очередь малая доля семантически структурированной медицинской информации мешает научным исследованиям в данной области. Так, медицинский персонал, на начальных этапах обучавший ИС IBM Watson потратил несколько месяцев на перевод электронного линейного текста в семантическую сеть системы (и это с учетом давней истории электронного медицинского документооборота в США).

В любом случае, опыт показывает, что перед созданием интеллектуальных экспертных систем на основе семантических технологий необходимо накопить достаточное количества знаний для принятия решений.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Целью диссертационной работы является разработка методов и средств построения медицинских информационных систем на основе открытой семантической технологии проектирования интеллектуальных систем, использующей в качестве формальной основы семантические сети с базовой теоретико-множественной интерпретацией.

Для достижения поставленной цели необходимо решить следующие задачи:

Произвести анализ существующих подходов к построению интеллектуальных медицинских информационных систем. Выделить их достоинства и недостатки.

Разработать принципиальную модель медицинской информационной системы.

Разработать средства взаимодействия медицинской информационной системы с базой знаний.

Разработать унифицированный язык представления знаний медицинской информационной системы.

Разработать модели основных медицинских классификаторов.

Разработать модель пользовательского интерфейса медицинской информационной системы.

Разработать модель хранения медицинских записей.

Реализовать тестовые примеры на основе полученных моделей.

Объектом исследования являются интеллектуальные медицинские информационные системы.

Предметом исследования являются технологии разработки интеллектуальных медицинских информационных систем.

Работа выполнялась в соответствии научно-техническими заданиями и планами работ кафедры интеллектуальных информационных технологий. Тема исследования соответствует пункту 5.1 перечня приоритетных направлений научных исследований Республики Беларусь на 2011–2015 годы, утвержденного Постановлением Совета Министров Республики Беларусь № 585 от 19.04.2010.

Основные положения диссертационной работы докладывались и обсуждались на V Международной научно-технической конференции Открытые семантические технологии проектирования интеллектуальных систем «OSTIS-2015» (Минск, 19–21 февраля 2015 года).

По теме диссертации опубликовано 3 печатных работы, из них 1 статья в сборнике научных статей IV Международной научно-технической

конференции Открытые семантические технологии проектирования интеллектуальных систем «OSTIS-2014», 1 статья в научно-практическом журнале «Электроника инфо», 1 статья в V Международной научно-технической конференции Открытые семантические технологии проектирования интеллектуальных систем «OSTIS-2015».

Библиотека БГУИР

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Диссертация состоит из введения, списка использованных сокращений, трех глав, заключения и списка использованных источников из 30 наименований, содержит 57 страниц машинописного текста, включая рисунки, 7 приложений на 14 страницах.

В первой главе проведен ретроспективный анализ медицинских систем, созданных на основе искусственного интеллекта. Определены особенности реализации современных медицинских информационных систем, основанных на знаниях. И была выбрана технология OSTIS для разработки моделей и средств интеллектуальной медицинской информационной системы.

В процессе ретроспективного анализа стало ясно что, несмотря на отсутствие каких-либо ограничений для развития и повсеместного распространения искусственного интеллекта в медицине, до настоящего времени сегмент разработки интеллектуальных медицинских систем находится в начальном состоянии. Отчасти это связано с особенностями применения, но в большей степени с особенностями разработки таких систем, заключающиеся в отсутствии семантически структурированных априорных и апостериорных знаний. Регулирующие государственные органы должны взять на себя не только юридические аспекты регулирования создания и внедрения медицинских информационных систем, но и приступить к созданию их компонентов: справочников, классификаторов, библиотек программных экспертных компонентов на основе унифицированных технологий. Однако такие процессы на территории стран бывшего СССР только начинаются, и здесь имеется уникальная возможность создать (адаптировать) уже имеющиеся стандарты медицинских информационных систем, и воспользоваться чужими Lessons Learned, и наконец, приступить к накоплению семантически структурированной информации в базах знаний.

При анализе подходов к построению современных медицинских информационных систем выяснилось, что, стек технологий как IBM Watson, так и Socmedica не несет полноценного решения для создания интеллектуальных медицинских информационных систем. Обе системы позиционируют себя не как интеллектуальная медицинская система, а как сервис поддержки принятия решений. Система IBM Watson от других интеллектуальных медицинских систем более отличается своей аппаратной составляющей, чем программной. С ростом возможностей вычислительной техники становится понятно, что в медицине решения должны приниматься не на основе математических алгоритмов, а на основе накопленных данных. Исторически, накопление происходило, однако даже возможностей

современной техники не хватает для формализации накопленной информации. Медицинские данные изначально необходимо хранить в формализованном, семантически структурированном виде. То есть хранить информацию медицинской информационной системы необходимо в семантическом хранилище – базе знаний.

В качестве семантической технологии для разработки модели медицинской информационной системы была выбрана технология OSTIS. Наличие семейства SC-языков, готовой реализацией sc-памяти, строгой теоретико-множественной трактовки делает проект OSTIS подходящей и зрелой технологией для проектирования интеллектуальных медицинских информационных систем. Не смотря на видимые достоинства OSTIS, для практического применения технологии в настоящее время существует несколько ограничений:

отсутствуют реализованные компоненты для поддержки многопользовательской авторизации;

отсутствуют реализованные компоненты для экспорта данных из базы знаний в линейные sc-языки;

обработка знаний, представленных в виде семантической сети, требует значительных системных ресурсов. Этот факт подтверждается и выбором аппаратной архитектуры при разработке IBM Watson.

Во второй главе представлены средства и методы построения медицинской информационной системы на основе технологии OSTIS. А в третьей главе описана практическая реализация и апробация предложенных средств и моделей.

Разработан унифицированный язык представления знаний медицинской информационной системы, который использует 3 инварианта отношений, но число классов (понятий) ограничено только предметной областью. Это позволяет создавать унифицированные агенты для работы с базой знаний, и постепенно расширять библиотеку интеллектуальных агентов не прибегая к переработке всей системы отношений. Однако такой подход порождает излишнюю вычислительную сложность и трудности при проектировании scs-кода. Поэтому крайне необходимы программные средства для автоматизации проектирования классов. Однако это невозможно, пока не будет создан инструмент для экспорта scs-кода из базы знаний. Экземпляры классов будут создаваться, и редактироваться посредством разработанного интерфейса, который так же будет храниться в виде базы знаний.

Созданы и апробированы модели и инструменты для транслирования в базу знаний МКБ-10 и Реестра лекарственных средств Республики Беларусь. Реляционную модель данных можно выразить в модели XML, а нереляционную

в модели RDF. И для этих моделей созданы и апробированы трансляторы в SCs код.

Предложена модель и созданы некоторые агенты для отображения интерфейса, ориентированного на документы, которая позволяет отойти от разделения ИС на рабочие места, и позволяет настроить доступ к информации более гибко. Основным недостатком предлагаемой модели интерфейса является ее сложность проектирования. Данная модель предполагает написание значительного объема sc-кода для каждого документа, что практически невозможно без создания дополнительных инструментов визуализации и верификации. Предложенная модель интерфейса является работоспособной, реализовав необходимое количество агентов для отображения элементов, можно создать инструменты для редактирования самого же пользовательского интерфейса.

Предложена модель хранения медицинских данных, которая поддерживает эволюционное развитие системы из хранилища данных в базу знаний, содержащей семантически структурированную информацию, пригодную для принятия решений, и это главное достоинство предлагаемой модели хранения медицинских данных в МИС. Апробация предложенной модели в настоящее время явно затруднена из-за отсутствия общепризнанных медицинских классификаторов (лабораторных исследований, клинических терминов, лекарственных средств).

Разработан sctp-клиент на языке высокого уровня C#. Разработанная библиотека является полностью работоспособной и актуальной, ее можно использовать для других проектов программистами, владеющими языком C#. Это можно рассматривать, как личный вклад автора в развитие технологии OSTIS. Однако в sctp-протоколе до настоящего времени не реализована поддержка идентификации пользователей, что явно ограничивает его применение до уровня прототипов приложений. Безусловно, можно реализовать собственную систему авторизации на языке C#, что усложнит общую структуру системы и приведет к проблемам подпроектной совместимости в рамках проекта OSTIS.

ЗАКЛЮЧЕНИЕ

Подводя итоги можно говорить о решении всех задач, перечисленных во введении к данной работе и успешном достижении поставленной цели.

Все предложенные модели и средства оказались работоспособны для разработки клинических медицинских информационных систем.

Более того, предложенные модели достаточно простые и подходят для разработки систем учета немедицинского назначения.

После незначительной доработки некоторых компонентов технологии OSTIS, таких как авторизация пользователей и экспорт содержимого базы знаний в линейный текст, можно начинать разработку интерфейсов и онтологий на основе предложенных моделей для накопления апостериорных медицинских знаний. Говорить о создании экспертных медицинских систем в условиях отсутствия этих знаний некорректно, а практическая их реализация невозможна.

Список опубликованных работ

Каешко, А.И. Нейросетевое прогнозирование заболеваемости военнослужащих органов пограничной службы Республики Беларусь. / А.И. Каешко, Г.А. Живень, О.А. Чеботарев, // Информационные технологии и системы 2013 (ИТС 2013): материалы международной научной конференции, БГУИР, Минск, Беларусь, 23 октября 2013 г.

Каешко, А.И. Принципы интеграции содержимого RDF-хранилищ в проект Ostis. / А.И. Каешко, Д.Г. Колб // Открытые семантические технологии интеллектуальных систем (OSTIS-2014): Материалы IV международной науч.-тех. конф (Минск, 20-22 февраля 2014г.) Минск: БГУИР, 2014. - с.447-452.

Каешко, А.И. Подход к трансляции содержимого rdf-хранилищ в семантические сети с базовой теоретико-множественной интерпретацией / А.И. Каешко, Д.Г. Колб // Журнал Электроника-инфо № 3 (105) 2014 г.

Каешко, А.И. Принципы построения клинической системы поддержки принятия решений на основе технологии OSTIS. / А.И. Каешко, Д.Г. Маргунов // Открытые семантические технологии интеллектуальных систем (OSTIS-2015): Материалы IV международной науч.-тех. конф (Минск, 18-21 февраля 2015г.) Минск: БГУИР, 2015. - с.149-156.