



Открытые связанные данные в здравоохранении



Российский офис Консорциума W3C

при поддержке

**Центра семантических технологий НИУ «ВШЭ»
и Российского национального исследовательского
медицинского университета имени Н. И. Пирогова**



Содержание

1	Введение	3
2	Современные проблемы эффективного использования информации.....	5
3	Предлагаемое решение	6
3.1	Международный опыт	6
4	Примеры и перспективы использования технологии Semantic Web и LOD в медицине и здравоохранении.....	8
4.1	Использование семантических технологий в государственных программах здравоохранения	8
4.2	Трансляционная медицина и исследования в области биохимии и биофизики.....	10
4.2.1	Примеры использования технологий семантического поиска информации в области трансляционной медицины и биомедицинских исследований	11
4.2.1.1	Проект LODD - банк открытой информации по лекарственным средствам	11
4.2.1.2	Проект TMO (Translational Medicine Ontology).....	13
4.2.1.3	Проект BioRDF - интеграция баз данных в области нейробиологии	14
4.2.1.4	База знаний молекулярной биологии NAR.....	15
4.2.1.5	NCBI - Национальный центр биотехнологической информации	16
4.2.2	Перспективы использования технологий семантического поиска в области научных и клинических исследований	16
4.3	Практическая медицина.....	17
4.3.1	Использование семантических технологий в интересах врача.....	20
4.3.2	Использование семантических технологий в интересах пациента	21
4.4	Использование семантических технологий в интересах бизнеса.....	21
4.5	Синергетический эффект LOD.....	22
5	Внедрение семантических технологий в сфере здравоохранения России.....	22
5.1	Предпосылки внедрения технологии Semantic Web в России	22
5.2	Преимущества использования технологии Semantic Web и LOD	23
5.3	Этапы запуска процесса внедрения семантических технологий в российском здравоохранении	24
6	Перспективы использования и развития ССМЗ в России	25
	Приложение 1. Краткое описание технологии	26
	Приложение 2. Пример использования семантических технологий в интересах врача и пациента.....	36
	Приложение 3. Перечень нормативно-справочной информации	38
	Используемые термины и сокращения.....	39

1 Введение

В настоящее время в России используется большое количество разрозненных медицинских информационных систем (МИС), формирующих разнообразные базы данных: регистр заболеваний и медицинских услуг, оказанных населению; описание лекарственных препаратов; результаты научных исследований; научные статьи и труды в области медицины и здравоохранения и т.п. Эти данные хранятся в специализированных электронных библиотеках в различных форматах. Однако, эффективных механизмов извлечения из таких источников знаний, хранения и предоставления к ним широкого доступа, не существует. И даже если некоторые данные открываются государственными органами для общественного доступа, то их чаще всего невозможно связать с данными из других источников в автоматическом режиме, приходится производить анализ и связывание информации вручную.

Аналитическая записка посвящена новой «прорывной» технологии эффективного использования знаний из различных источников информации, основанной на семантическом связывании¹ открытых данных – Linked Open Data (LOD), а также актуальности ее применения в активно модернизируемой в настоящее время отрасли – здравоохранении.

Одна из главных возможностей использования LOD в государственных интересах организации здравоохранения - оценка дисбаланса между инвестициями в исследования отдельных заболеваний и их распространенностью. В России этот дисбаланс частично компенсируется государственным заказом. Однако отсутствие счетных метрик, формируемых из взаимосвязанных данных из различных источников, мешает точному определению приоритетов, начиная с биомедицинских исследований и заканчивая государственным заказом на развертывание и производство фармпрепаратов.

На web-сайтах, созданных правительственными усилиями, содержится важная информация по здравоохранению, но эта информация в основном опубликована в текстовом виде, либо только в частично- машиночитаемом формате. Хотя прилагаются усилия объединить информационные ресурсы с целью отображения статистики в виде графиков или карт, данные не связаны с другими источниками данных. Они также не адресуют научные исследования к дисбалансам здравоохранения, что могло бы помочь оценить глубину этого несоответствия для конкретных условий, например, для клинических случаев, или для продолжения исследований в данной области.

Расчет дисбаланса в здравоохранении позволил бы руководителям определять соответствие стратегий исследований, установленных Национальными Инновационными Системами с национальными потребностями здравоохранения.

LOD – это данные, опубликованные в стандартах Semantic Web международного Консорциума W3C и связанные друг с другом. Иными словами, это своего рода идеология, согласно которой, открытые данные должны быть не просто открыты и пригодны для машинной обработки, этого недостаточно. Настоящую ценность данные обретут, когда

¹ Семантическое связывание – выявление смысловых связей между данными (объектами в тексте) и отображение их на графе с занесением в базу знаний.

они будут связаны между собой, объединены в облако – пространство знаний о некоторой предметной области или многих предметных областях.

Данная технология активно разрабатывается и реализуется в международном проекте W3C SWEO Linking Open Data (LOD) и позволяет в области здравоохранения:

- расширять возможности и увеличивать скорость поиска и анализа медицинской информации врачам и пациентам,
- повышать оперативность и качество принятия решений руководителям здравоохранения разного уровня,
- повышать эффективность научных исследований научным работникам,
- сокращать сроки разработки и внедрения новых продуктов (лекарственных средств, медицинских приборов и т.п.) производителям.

Одним из важнейших достоинств технологии LOD является возможность объединения в общую сеть распределенных семантических хранилищ, созданных различными организациями (органы управления здравоохранением, медицинские организации, научно-исследовательские центры, профессиональные сообщества, ВУЗы и др.) на основе единых открытых стандартов. Как показывает международный опыт, это позволяет системе саморазвиваться, постоянно пополняя ее данными и повышая их качество:

- По мере публикации дополнительных наборов выстраиваются новые цепочки пользователей:
Биохимики → Фармакологи → Врачи-исследователи → Практикующие врачи → Клиницисты → Медстатистики → Менеджеры здравоохранения
- Формируется политика, какие данные публиковать, в какую очередь (в зависимости от того, нужно ли для государства объединить информационные ресурсы смежных отраслей и можно ли построить аналитику на этих наборах) - управляемость
- Информационные ресурсы интегрируются с мировыми ресурсами открытых данных LOD
- Сервисы на открываемых данных начинают создавать бизнес – привлечение бизнеса в соинвестирование проектов

Целью данной аналитической записки является инициирование процессов создания семантической сети² знаний в сфере российского здравоохранения (ССМЗ) на основе опыта и результатов международного проекта LOD и внедрения семантических технологий автоматизированной интеллектуальной обработки и объединения знаний, накопленных различными медицинскими, государственными, научно-исследовательскими, производственными организациями. ССМЗ будет представлять собой открытый информационный ресурс, предлагающий сервисы поиска и анализа информации на основе семантического медицинского портала.

Необходимая компетенция инициативной группы, представленной авторами аналитической записки, подтверждена многосторонним опытом ее участников и описана в **Error! Reference source not found..**

² Семантическая сеть — информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними.

2 Современные проблемы эффективного использования информации

В условиях стремительного роста темпа жизни возрастает важность принятия правильных решений в сжатые сроки. Наряду с увеличением объемов доступной для использования информации, сокращается время на ее сбор и структурированную обработку.

Для поиска информации и получения ответов на возникающие вопросы все чаще используются Интернет-технологии. Огромное количество информации, имеющей отношение к здравоохранению, публикуется во Всемирной сети, а также хранится в ведомственных и корпоративных системах по таким направлениям:

- Государственные программы охраны здоровья и их финансирование;
- Нормативно-справочная информация в здравоохранении;
- Современные достижения медицины в области диагностики и лечения;
- Информация о лечебных учреждениях и их возможностях;
- Научные публикации и результаты клинических испытаний;
- Электронные персональные медицинские документы;
- Достижения естественных наук (химия, биология и др.);
- Исследования в области фармакология;
- Медицинская статистика;
- и др.

В настоящее время Интернет превратился из «всемирной библиотеки знаний» во «всемирную свалку», где много недостоверной, несущественной и многократно дублированной информации. Поэтому поиск интересующей информации в этом источнике сталкивается с рядом существенных проблем:

- Как эффективнее просмотреть все выданные при поиске ссылки? – проблема скорости поиска и дефицита времени.
- Вся ли существенная информация получена и рассмотрена? – проблема полноты данных.
- Как сгруппировать выдаваемые данные и получать по ним статистику? – проблема структурирования больших объемов неструктурированных данных.
- Как выбрать оптимальный вариант из полученных данных? – проблема автоматического анализа информации с целью поддержки принятия решений.
- Как при необходимости смены параметров запроса на поиск, не повторять трудоемкую процедуру «ручной» обработки всего объема информации сначала? – проблема отсутствия автоматической обработки данных.

Данные, хранимые в различных ведомственных и корпоративных базах данных и файловых системах, чаще всего недоступны сторонним пользователям или организациям, а если даже и открываются для общественности на соответствующих Интернет-сайтах, то формат этих данных разнородный (тексты, графические структуры, электронные таблицы, файлы, базы данных) и не предоставляет возможности автоматической обработки, анализа и, тем более, автоматического связывания с данными из других открытых источников информации.

Так же, как и в других областях, в сфере российского здравоохранения по мере накопления новых знаний, проведения исследований, осуществления реформ, разработки новых, современных технологий, лекарственных средств постоянно возрастает количество вопросов – от «бытовых» до профессиональных (врачебных, управленческих,

исследовательских), требующих быстрого получения качественной информации. Налицо быстро развивающееся противоречие между растущими потребностями в информации, ее объемами и невозможностью эффективного использования.

Назрела необходимость усилий по унификации представления открываемой информации, ее смысловому связыванию и интеграции в единую сеть знаний в сфере российского здравоохранения.

3 Предлагаемое решение

3.1 Международный опыт

Во всем мире сталкиваются с изложенными выше проблемами. Для их решения разработан и активно применяется международным сообществом стандарт публикации данных в Web – стандарт Связанных Открытых Данных (Linked Open Data - LOD), созданный и активно развиваемый Консорциумом всемирной сети (W3C) как технология семантической паутины (Semantic Web).

Основные принципы LOD:

- Информация, публикуемая в Web, читается не только людьми, но и компьютерами;
- Данные и стандарт их публикации в Web используется как глобальная обобщенная база данных;

В формат LOD преобразуются любые данные – структурированные (базы данных, таблицы) и неструктурированные (интернет-страницы, документы различных форматов);

- Простота интеграции данных из разных источников позволяет подключаться разным производителям данных;
- Связывание данных из различных источников повышает ценность предоставляемой информации.

Краткое описание технологии приведено в Приложение .

В этом формате за рубежом создана мощная сеть связанных баз данных по различным областям знаний, которые формируются и эксплуатируются государственными и коммерческими структурами, университетами и исследовательскими центрами, библиотеками и музеями, новостными агентствами и социальными сетями и другими заинтересованными организациями и лицами. Все больше различных организаций и государственных структур (в Великобритании, США, Канаде, Германии, Швеции, Венгрии и др.) принимают решение выкладывать свои данные в семантическую сеть в данном стандарте. Наглядное представление темпов роста публикаций открытых данных в мире представлено на Рис. 1. Каждый кружок на рисунке – это публикация данных какой-либо организацией в стандарте LOD, цвет отражает область, к которой относятся данные.

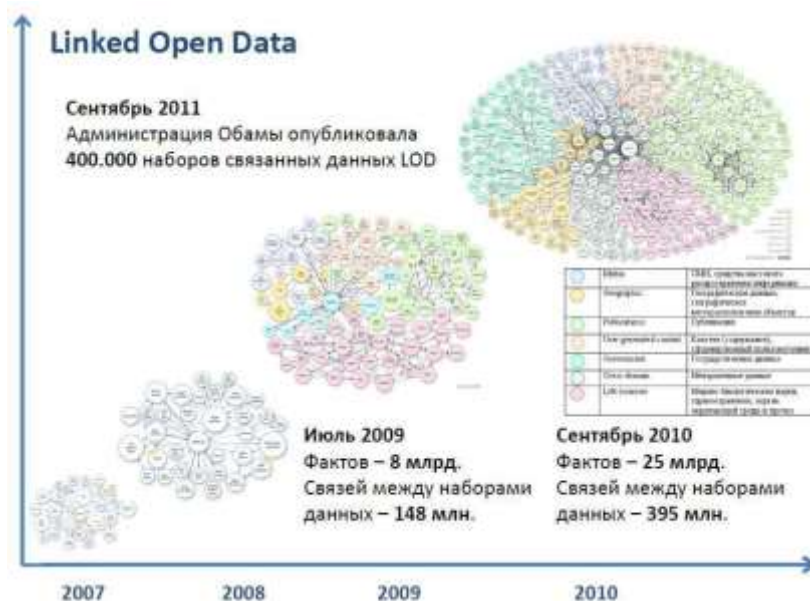


Рис. 1 – Рост публикации открытых данных в мире.

Источник: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

Схематичное представление международной сети семантических баз знаний в различных областях приведено на Рис. 2.

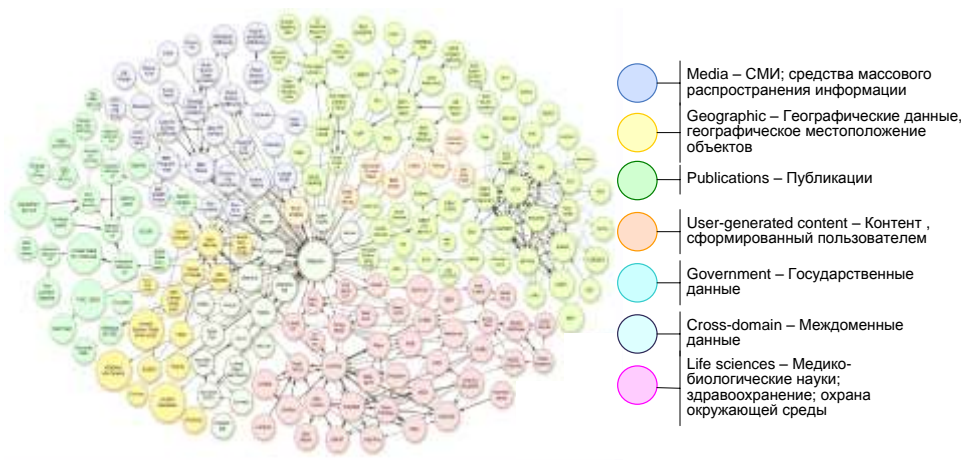


Рис. 2 – Международная сеть семантических баз знаний LOD.

Источник: http://richard.cyganiak.de/2007/10/lod/lod-datasets_2010-09-22_colored.pdf

Сегмент, относящийся к области естественных наук и здравоохранения (Life sciences) представлен на Рис. 3. Как видно, он является весьма значительным в этой структуре.

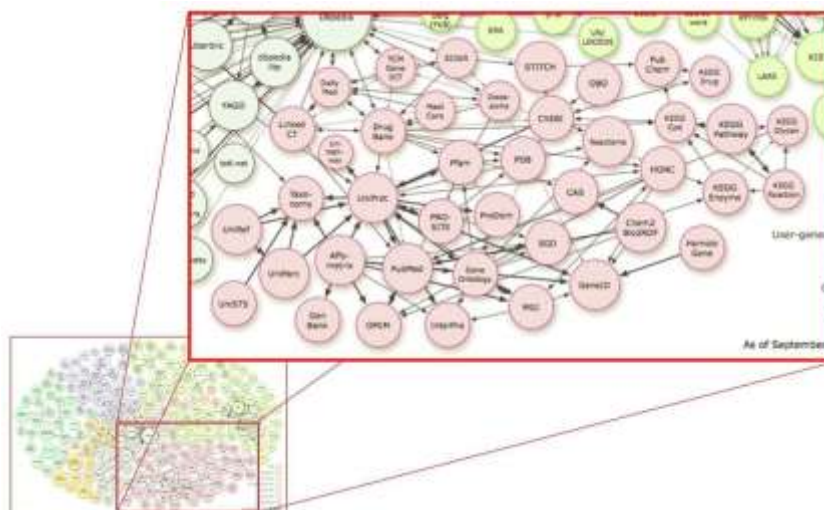


Рис. 3 – Сегмент LOD Life sciences за рубежом

В настоящее время за рубежом известны следующие направления Открытых Связанных Данных (LOD) из области Life sciences:

- Государственные программы здравоохранения;
- Биологические науки;
- Трансляционная медицина;
- Данные о разработках новых лекарственных средств фармакологическими компаниями;
- Владение данными пациентов;
- Поиск патентов и литературы;
- Установленные правила при проведении биоисследований и клинических испытаний;
- Совместная научно-исследовательская деятельность;
- Структурированные экспериментальные результаты.

4 Примеры и перспективы использования технологии Semantic Web и LOD в медицине и здравоохранении

4.1 Использование семантических технологий в государственных программах здравоохранения

Одна из существенных возможностей использования LOD в государственных интересах организации здравоохранения в настоящее время – оценка дисбаланса между инвестициями в исследования отдельных заболеваний и их распространенностью. Ярким примером такого дисбаланса служат данные Всемирной организации здравоохранения (ВОЗ) о том, что более одного миллиарда человек (то есть одна шестая часть мирового населения) страдают от одного или нескольких забытых тропических болезней, а финансирование этого направления практически полностью отсутствует. Хотя большая часть такого значительного несоответствия может быть связана с рыночными аспектами биомедицинских исследований, проводимыми фармацевтическими компаниями, это также вызвано и отсутствием знаний в отношении текущего состояния и интенсивности

этого дисбаланса. На сегодняшний день отсутствуют взаимосвязанные данные, позволяющие точно описать степень несоответствия между усилиями в области биомедицинских исследований и глобальными потребностями в медицинской помощи.

Существует множество web-сайтов, созданных правительственными усилиями в разных странах, которые содержат важную информацию по здравоохранению, но эта информация в основном опубликована в текстовом виде и либо вовсе не-, либо только частично-машиночитаемом формате. Хотя прилагаются усилия объединить связанные информационные ресурсы с целью, например, получения выдержек о дисбалансах в здравоохранении, или с целью отображения статистики в виде графиков или карт, эти данные не связаны с другими источниками данных. Можно привести следующие примеры публикаций дисбалансов:

1) Агентство по здравоохранению и качеству исследований (AHRQ) в США публикует два ежегодных отчета о дисбалансах Национального здравоохранения. Тем не менее, эти отчеты сосредотачиваются лишь на нескольких демографических группах населения и ограниченном числе клинических условий. Даже при наличии свидетельства, что на протяжении последних десятилетий инвестиции в медицинские исследования были экономически эффективными, Комиссия по медицинским исследованиям и развитию отмечает большой дисбаланс между этими инвестициями и общей заболеваемостью в 1990 г. Правительство США также сообщает об отсутствии хорошей корреляции между заболеваемостью и суммой средств, выделяемых для борьбы с болезнями. Гораздо более серьезная проблема существует в развивающихся странах, угрожая малоимущим, детям и женщинам.

2) Исследование (Girosi F., et al) о том, какое количество смертей можно предотвратить при определенном годовом бюджетном расходе на иммунизацию и скрининг населения по ряду распространенных заболеваний (например, общий расход в 90 млн. долл. США в год на вакцинацию от пневмонии позволяет предотвратить от 15000 до 27000 жизней в год). Ставится вопрос о внедрении систем напоминаний по иммунизации и скринингу в МИС лечебных учреждений. В данном случае дисбаланс проявляется в следующем – очевидна выгода такой системы с точки зрения общества и совсем неочевидна выгода с финансовой точки зрения для лечебного учреждения, которое оплачивает МИС и ведение предупреждений (т.к. улучшается здоровье людей, уменьшается количество оказанных услуг).

В России дисбаланс частично компенсируется государственным заказом. Однако отсутствие счетных метрик, формируемых из взаимосвязанных данных из различных источников, мешает точному определению приоритетов, начиная с биомедицинских исследований и заканчивая государственным заказом на развертывание и производство фармакологических препаратов.

Одним из способов измерения дисбаланса является выполнение кроссекционных исследований, сравнивающих оценки финансирования на исследования конкретных заболеваний с данными по различным показателям заболеваемости, взятыми из открытых государственных деперсонифицированных статистических ресурсов. Существующие методы для количественной оценки дисбаланса, которые используют различные статистические показатели на выборочных данных, требуют больших затрат по усилиям и времени и используют только ограниченные наборы данных для анализа. В

противоположность этому, подход с семантическим поиском из Связанных онтологий (LOD) обеспечит в режиме реального времени обзор дисбаланса исследований за счет публикуемых в них “глобальных данных”. В этом случае возможен анализ в различных разрезах, методах, средствах и способах оценки дисбаланса. Более того, предоставление данных в машиночитаемой форматах и с возможностью взаимосвязей с другими наборами данных, делает расчеты менее трудоемкими и более надежными.

Расчет дисбаланса в здравоохранении позволил бы руководителям определять соответствие стратегий исследований, установленных Национальными Инновационными Системами, с национальными потребностями здравоохранения.

Анализируя возможности связанных семантических технологии, можно определить общие перспективы их использования в сфере управления и организации здравоохранения:

- контроль за принимаемыми управленческими решениями в отрасли и анализ их влияния;
- отслеживание изменений и появления новых нормативных и правовых документов;
- выявление тенденций в динамике показателей по направлениям медицины и здравоохранения на основе собранной статистики.
- отслеживание качества работы медицинских учреждений и организаций по отзывам специалистов и населения;
- знакомство с опытом других регионов/ стран при решении проблем;
- отслеживание проектов и программ, осуществляемых в регионах по различным клиническим направлениям, оценку их влияния и эффективности реализации.

4.2 Трансляционная медицина и исследования в области биохимии и биофизики.

Одной из самых успешных в мире областей использования технологий семантического поиска информации является Трансляционная медицина – это новая междисциплинарная область знаний, интегрирующая элементы клинической медицины и биотехнологические подходы к разработке новых терапевтических и диагностических средств, и направленная на перенос открытий, сделанных в результате фундаментальных исследований в лабораториях в сферу практического применения в медицине. Развитие данного направления позволит объединить "под одной крышей" усилия клиницистов и ученых биологической направленности, способствовать совершенствованию процесса, в котором базовые исследования представляют информацию клиническим, а клинические – базовым. Именно трансляционной медицине будет принадлежать ведущая роль в развитии биомедицины на протяжении ближайших десятилетий. Приоритетными задачами развития трансляционной медицины являются: разработка инновационных методов молекулярной диагностики, создание новых медицинских устройств и искусственных органов (бионика), тканевая инженерия, геновая и клеточная терапия. Во многих странах создаются институты трансляционной медицины (например, NIH – Национальные институты здоровья США). Платформу для взаимодействия институтов и ассоциаций, расширение границ такого взаимодействия за пределы одной страны, представление возможности обратиться к накопленным знаниям для их практического использования врачом-клиницистом и представляют семантические технологии работы с информацией.

В рамках консорциума W3C сформирована рабочая группа HCLS IG (Health Care and Life Sciences Interest Group), в настоящее время уже включающая более 100 индустриальных

членов. Целями группы HCLS являются разработка, продвижение и поддержка использования семантических веб-технологий в сфере здравоохранения и естественных наук, с акцентом на биологические науки, клинические исследования и трансляционную медицину. Использование семантических веб-технологий будет способствовать интеграции разнородных данных для решения вопросов эффективной поддержки принятия решений, так как они зависят от информационного взаимодействия между многими областями и процессами.

В настоящее время членами группы HCLS активно ведутся проекты:

- Интегрированная база знаний по неврологии (BioRDF) – Йельский Университет (Yale University, Kei Cheung).
- Взаимодействие клинических наблюдений (Clinical Observations Interoperability) – набор пациентов для исследований - Cigna Healthcare (Vipul Kashyap)
- Банк открытой информации по лекарственным средствам (LODD) – Объединение информации по лекарственным средствам из веб-источников – Свободный университет Берлина (Free University Berlin, Chris Bizer)
- Онтология Трансляционной медицины (ТМО) – высокоуровневая ориентированная на пациента онтология – фармацевтическая компания США Эли Лилли (Eli Lilly, Christi Denney)
- Научный форум (Scientific Discourse) – создание сообществ посредством компьютерных сетей – Гарвардский Университет (Harvard University, Tim Clark)
- Терминология (Terminology) – семантическое представление данных из существующих источников – Университет Дьюка (Duke University, John Madden).

4.2.1 Примеры использования технологий семантического поиска информации в области трансляционной медицины и биомедицинских исследований

4.2.1.1 Проект LODD - банк открытой информации по лекарственным средствам

В качестве примера объединения различных наборов данных можно рассмотреть проект LODD (Linking Open Drug Data) – объединенный банк открытой информации по лекарственным средствам, схема которого приведена на Рис. 4, а описание онтологий, полученных из различных достоверных интернет-источников – в Табл. 1

Источник - <http://www.w3.org/wiki/HCLSIG/LODD>

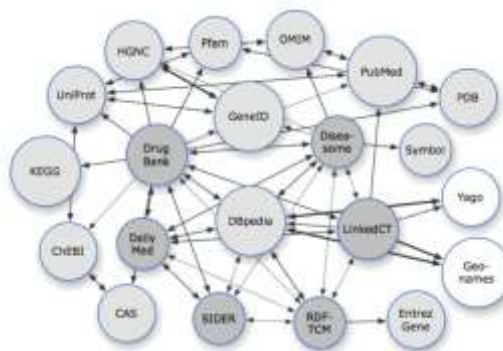


Рис. 4 – Наборы данных, входящих в Linking Open Drug Data – LODD

Табл. 1 – Наборы данных, их характеристика и источники в рамках проекта LODD:

Название	Тема	Краткое Описание	Размер и покрытие	Статус / Активность
DrugBank	Медикаменты (препараты)	Ресурс Drugbank.ca предоставляет комплексную информацию по препаратам (включая химические, фармакологические и фармацевтические аспекты)	766,920 триплов; 4,800 препаратов, 2,500 последовательно стей протеинов	Регулярно обновляется
LinkedCT	Клинические испытания лекарственных веществ	Источник связанных (объединенных) знаний по клиническим испытаниям от ClinicalTrials.gov	~25 миллионов триплов, 106,000 испытаний (на Апрель 2011)	Постоянно автоматически обновляется.
DailyMed	Медикаменты	Ресурс dailymed.nlm.nih.gov предоставляет информацию по разрешенным для отпуска по рецептам лекарствам, включая утвержденные FDA этикетки (вкладыши в упаковках)	164,276 триплов; 4,039 препаратов	регулярно обновляется
DBpedia	Препараты/ заболевания/ белки	RDF-данные о 2.49 миллионах объектов, извлеченных из Wikipedia.	218 миллионов RDF триплов; 2,300 препаратов, 2,200 протеинов	обновляется раз в 3 месяца
Diseasome	Заболевания/ Гены	Описывает характеристики генетических расстройств и заболеваний, связанных известными ассоциациями расстройство-ген	91,182 триплов; 2,600 генов	-
RDF-TCM	Гены/ Заболевания/ Лекарство/ Ингредиенты	Набор ассоциативных данных по традиционной китайской медицине, генам и заболеваниям, а также набор ссылок, маппирующий символы генов TCM для идентификаторов Extrez Gene, созданных Neurocommons.	117,643	обновлялся в августе 2009 (стабилен)
RxNorm	Препараты	Связанная версия набора данных NLM's RxNorm, которая связывает рецептурные препараты, ингредиенты и NDC через понятие уникального идентификатора RXCUI. RxNorm является продуктом, разработанным Национальной Библиотекой медицины NIH. В настоящее время она связывает 12 различных медикаментных словарей вокруг уникального идентификатора. В связи с лицензиями, только шесть из 12 словарей доступны как часть LODD облака.	более 7.7 миллионов триплов; 165,806 RXCUI Уникальных препаратов и ингредиентов; 332,754 RXAUI источников терминов	Базируется на выпуске RxNorm от 3/2010. Последнее обновление – 5/2010
SIDER	Заболевания/Побочные эффекты	Содержит информацию по выпущенным на рынок препаратам и о побочных эффектах, которые они могут вызывать.	192,515 триплов; 63,000 отчетов о неблагоприятных эффектах, 1,737 генов	Обновлялся в 2009
STITCH	Химические вещества/ Протеины	Содержит информацию по химическим веществам, протеинам и их взаимодействию.	7,500,000 химических элементов;	Обновлялся в июле 2009

			500,000 протеинов; 370 органических веществ	
ChEMBL	Химические вещества/ Анализы (протеины, микрофлора)/ Документы	Содержит информацию по тестовым препаратам с данными о деятельности по объектам, не ограничиваясь протеинами. Все сохранено и соединено с соответствующей литературой.	~24М триплов	Обновлялся 01.2009
WHO Global Health Observatory	Инфекционные болезни / Демография / Социально- экономические условия / Фактор ы окружающей среды	Данные и статистика по инфекционным болезням на уровне страны, региона и мира	354.300 триплов	Обновлялся 09.2010

4.2.1.2 Проект ТМО (Translational Medicine Ontology).

Современные исследования выявляют все больше ассоциаций между результатами лечения и генно-обусловленными факторами. Последние достижения в области тонкого биологического понимания процессов дают возможность фармацевтическим компаниям начинать разрабатывать специализированную индивидуальную терапию, что позволяет подбирать пациентам наиболее подходящее ему лекарство, в правильной дозировке, и в надлежащее время. Для того, чтобы разработать такие подходы к лечению, необходима возможность связывать лабораторные данные с клиническими исследованиями (доведение результатов исследования до использования в лечении). Источник – <http://www.w3.org/wiki/HCLSIG/PharmaOntology>.

Проект нацелен на создание моделей трансляционной медицины, чтобы специалисты здравоохранения и исследователи могли получить доступ к этой информации. Работа направлена на объединение фармакологических исследований с клиническими знаниями для поддержки принятия решений с целью оценки эффективности препарата и безопасности его назначения. Примеры возможных направлений этой деятельности включают: интеграцию медицинских данных для идентификации сигналов безопасности, агрегирование данных клинических испытаний для выявления интересующих исследований, и создание панелей мониторинга, показывающих, как разнородные и разрозненные данные могут быть интегрированы для поддержки принятия решений различных пользователей (Табл. 2).

Табл. 2 – Описание возможных интересов пользователей при работе с LOD в области ТМО.

Категория	Пользователь	Интерес
Исследование	Биолог	Исследование, поиск молекул, белков, ответственных за процессы, происходящие в организме (например, выделение белка, ДНК и т.п.)
	Биоинформатик	Управления знаниями на биологическом уровне, клеточное моделирование
	Иммунолог	Защитные механизмы организма

	Биохимик	Предсказание структуры молекул, лекарств Генная инженерия Использование нанотехнологий в медицине
	Фармаколог	Исследование эффективности препаратов
	Фармаколог-исследователь	Побочные эффекты лекарственных средств Определение терапевтической и минимальной токсической дозы.
Клиника	Специалист по клин. испытаниям	Дизайн клинического испытания, подбор группы пациентов для проведения исследования
	Медицинский кибернетик	Анализ данных, поиск тренда, Поддержка принятия решений
	Врач терапевт	Стандартное лечение
	Врач специалист	Специализированное лечение
Бизнес	Специалист Продажи и Маркетинга	Получение прибыли
	Стратегический менеджер (бизнес-планировщик исследования), инвестор	Оценка рынка
	Менеджер проекта	Выбор приоритетов ресурсов и деятельности
	Страховая компания	План покрытия лечения страховкой

Основное внимание в проекте ТМО уделяется развитию высокоуровневых онтологий (моделей предметной области), ориентированных на пациента, для фармацевтической промышленности. Эти модели призваны интегрировать исследовательские данные, системы управления гипотезами, данные эмпирических исследований, данные по соединениям, составам, разработке лекарственных средств, размерам рынков, конкурентные данные, данные по популяции населения и т.д. Такая интеграция данных поможет:

- исследователям получить ответы на существующие научные вопросы и поставить новые проблемы.
- фармацевтическим компаниям получать информацию, которая необходима для создания лекарств, позволяющую индивидуализировать подход к лечению, а также для раннего обнаружения соединений, которые имеют побочные эффекты.

4.2.1.3 Проект BioRDF - интеграция баз данных в области нейробиологии

Проект BioRDF рабочей группы HCLS направлен на организацию данных и интеграцию в области нейробиологии (см. http://www.w3.org/wiki/HCLSIG_BioRDF_Subgroup). В Neuroscience Database Gateway (NDG -- динамический портал, единый интерфейс поиска, предоставляющий доступ к данным из области нейробиологии и обладающий обратной связью) указано, что есть около 200 баз данных по нейробиологии, которые доступны через Интернет и это не исчерпывающий список. Кроме того, имеются веб-ресурсы, которые не являются базами данных (например, печатные материалы). В проекте BioRDF решается задача интеграции данных из всех источников с использованием набора

семантических веб-технологий. Вместо сбора данных из различных БД и их централизованного хранения, запросы к любым источникам осуществляется благодаря их интеграции на различных уровнях, поддерживая комплексные исследования в нейробиологии.

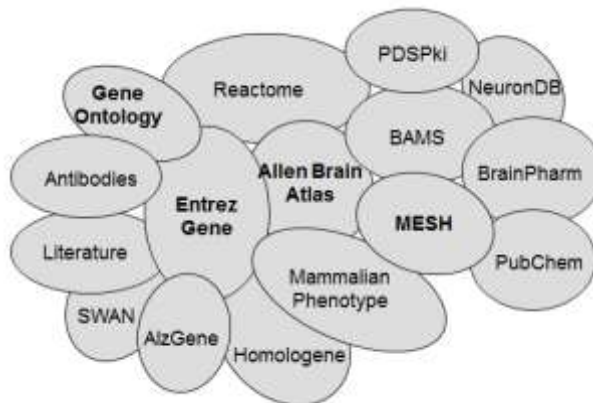


Рис. 5 – BioRDF: Интеграция разнородных данных в области нейробиологии

4.2.1.4 База знаний молекулярной биологии NAR

База знаний NAR (Nucleic Acid Research) включает 1170 существующих источников данных по молекулярной биологии (<http://nar.oxfordjournals.org/>). Онтология Linked Life Data (LLD) интегрирует наборы знаний, которые описывают отношения генов, протеинов, взаимодействия, путей, целей, лекарств, заболеваний и пациента. Существующее структурированное знание обогащено информацией из базы знаний PubMed, которая включает в себя более 19 миллионов цитат из биомедицинских статей MEDLINE и журналов, посвященных естественным наукам. Кроме того, LLD также включает информацию из UMLS metha-thesaurus, которая семантически интегрирует множество словарей в области биомедицины. Набор знаний The Linked Life Data состоит из более, чем 5 миллиардов RDF отчетов.

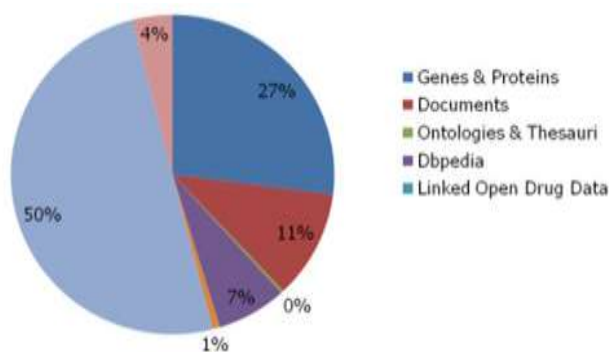


Рис. 6 – База знаний NAR

Набор знаний связывает более 20 полных источников информации и помогает охватить «глобальную картину» проблем исследований путем объединения прежде несвязанных данных из разнородных наборов источников информации. Чтобы сделать использование LLD-облака эффективным, созданы модели выравнивания, которые замещают пустующие отношения между информационными единицами. В качестве финального шага, было сгенерировано огромное количество семантических аннотаций (оптимизированных под

высокую точность и полноту), соединяющих экземпляры семантических данных и неструктурированную информацию.

4.2.1.5 NCBI - Национальный центр биотехнологической информации

Национальный центр биотехнологической информации США (NCBI) основан как центральный институт обработки и хранения данных молекулярной биологии и является частью Национальной медицинской библиотеки США (<http://www.ncbi.nlm.nih.gov>). NCBI предоставляет информацию о базах данных белковых доменов, ДНК (GenBank) и РНК, базах данных статей научной литературы (PubMed) и таксономической информации (TaxBrowser), обеспечивает поиск данных о конкретном биологическом виде (Taxonomy).

Мета-тезаурус "Онкология" содержит 50000+ классов объектов, примерно 8 сотрудников поддерживают его работоспособность целый рабочий день, онтология ежемесячно обновляется. Ресурс **обязателен для использования** в финансируемых Национальным Институтом Здравоохранения (National Institutes of Health) исследованиях рака и интенсивно используется в медицинском сообществе в качестве очень «достоверного» информационного источника.

4.2.2 Перспективы использования технологий семантического поиска в области научных и клинических исследований

В общем случае использование технологий семантического поиска в области научных и клинических исследований позволит осуществлять полный и постоянный охват и анализ информации по следующим направлениям:

- наличие данных по интересующей проблеме в различных аспектах (по темам, заболеваниям, авторам исследований, организациям, разработкам);
- существующие аналоги по исследуемому направлению;
- связь со смежными темами и областями.

Например, если в научной медицинской деятельности выделить информационные объекты «Исследователи», «Публикации», «Источники исследований» и др. и описать взаимосвязи между этими объектами, то можно при поиске с помощью технологий LOD выявлять связи представителей каждого информационного объекта:

- научные источники информации по искомой теме и их авторов;
- исследователей, занимающихся сходной проблематикой, их специализации, квалификации и уровни в научном мире;
- учреждения, организации и школы, которые представляют конкретные исследователи,
- работы, которые они публикуют, их проекты и открытия, клинические испытания, которые они проводят и т.п.

Цепочку этих связей можно продолжать, отвечая на следующие вопросы:

- «какие еще проблемы рассматриваются в данном учреждении/ организации»;
- «какими еще темами интересуется данный исследователь»;
- «какие научные представители относятся к одной группе/ организации», и т.п.

С целью оценки текущей ситуации в научной области, прогнозирования тенденций, составления отчетов можно получать статистику по всем перечисленным параметрам.

Важным является тот факт, что при наличии договоренностей между государственными органами здравоохранения, разработчиками МИС для медицинских учреждений и организаций, фармацевтическими компаниями, которые накапливают данные в своих базах (БД), можно организовать возможность совместного использования этих БД с целью получения новых знаний или разработки систем поддержки принятия решений и прогнозирования в исследуемой области.

Для достижения этой цели уже сейчас на Западе и в США проводятся мероприятия по стандартизации и унификации сбора и обмена медицинскими данными совместно с органами стандартизации, такими как CDISC, Health Level Seven (HL7), SNOMED, BRIDG, что позволяет использовать эти стандарты с семантическими веб-технологиями.

4.3 Практическая медицина

Примером использования семантических технологий в области практической медицины может послужить описанная в Journal of Biomedical Semantics 2011, 2 (<http://www.jbiomedsem.com/content/2/S2/S1>) последовательность действий врача при индивидуальном подборе терапии пациенту с тяжелым, трудно поддающимся лечению, заболеванием.

Пациента и его родственников беспокоят проблемы, связанные с расстройством памяти.

1. Во время приема пациент сообщает врачу-терапевту свои симптомы. Вся информация вводится врачом в электронную медицинскую карту пациента (ЭМК). С помощью LOD-проекта ТМО (Онтологии трансляционной медицины, см. выше) все необходимые данные автоматически переводятся в специальный формат, понятный для семантического анализа.

2. Врач составляет список вероятных диагнозов, с Болезнью Альцгеймера в качестве рабочего.

3. Врач назначает пациенту необходимое диагностическое обследование, чтобы иметь основные биохимические, гематологические, и цитогенетические параметры. Данные из биохимии, гематологии, и лаборатории цитогенетики передаются непосредственно в ЭМК пациента. Предварительный цитогенетический профиль и данные генетического анализа напрямую поступают в научно-исследовательскую семантическую сеть НИИ (Национальные институты здоровья, см. выше) по направлению Фармакогенетика (PGRN - Pharmacogenomics Research Network – Научно-исследовательская сеть, нацеленная на понимании того, как гены человека влияют на него или его реакцию на лекарства).

4. Результаты исследований / анализов в сочетании с описанием клинической информации добавляется в ЭМК пациента. Из перечисленных диагнозов на основании всей имеющейся информации выбирается Болезнь Альцгеймера как наиболее вероятная, подтвержденная поведенческой оценкой пациента, когнитивными тестами, а также, при необходимости, предоставленным изображением сканирования мозга. Диагноз также вводится в ЭМК.

5. Врачу необходимо выбрать наиболее подходящий медикамент для лечения Болезни Альцгеймера на основании клинического протокола медицинских записей пациента, учитывая тяжесть заболевания, цитогенетический профиль пациента, индекс массы тела, и одновременные медикаментозные назначения по поводу других заболеваний, а также возможность покрытия данного препарата медицинской страховкой (например, Medicare D).

Поиск в соответствующих онтологиях ответит на основные вопросы:

- *Какие лекарственные вещества рекомендуются?*
- *Какие медикаменты доступны для получения по рецепту, и какие разрешены для лечения болезни Альцгеймера?*

• *Что показывает цитогенетический профиль пациента? Такую возможность дает система поиска в онтологии ADME (Absorption, Distribution, Metabolism, and Excretion – позволяет оценить эффективность лекарственных веществ, наличие или отсутствие побочных эффектов на основе наличия или отсутствия рецепторов на эти в-ва).*

Лекарственные вещества сопоставляются с онтологией PGRN (о которой сказано выше) для определения:

– *Будут ли они эффективными? Являются ли рецепторы болезни положительными?*

– *Будут ли они вредны? Есть ли токсические метаболиты? Воздействует ли это в-во на CYP 450 (цитохром 450) или обладает ли ацетилирующим эффектом.*

• *Покрывается ли предыдущее прогностическое генетическое SNP-тестирование страховой компанией пациента? Покрывается ли имеющимся у пациента страхованием назначение лекарственных средств?*

6. *Врач консультируется с фармацевтом или просматривает информацию в литературе о лекарствах, чтобы избежать возможных взаимодействий с другими препаратами. И назначает, например, Aricept (Донепезил), поскольку он удовлетворяет критериям, перечисленным выше. Он описан как безопасный, эффективный, доступный, не имеющий проблем при взаимодействии с другими препаратами, и он покрывается страховкой.*

7. *При дальнейшем наблюдении пациент сообщает о тошноте от приема Донепезила. Врач знает об этом часто встречаемом побочном эффекте препарата (имеются и другие побочные эффекты: брадикардия, диарея, анорексия, боли в животе, и яркие сны и т.д.) и вновь обращается к литературе, чтобы понять, какое лечение является приемлемым и показанным пациенту. Побочный эффект документируется для будущего пост-маркетингового исследования. Врач меняет лечение, если необходимо, или добавляет другой препарат для облегчения побочных эффектов.*

8. *Онтология TCM может быть использована для поиска средств растительного происхождения, которые способствуют борьбе с болезнью Альцгеймера и их предполагаемых эффектах. Онтологии SIDER и DailyMed совместно могут быть использованы для поиска информации о побочных эффектах ингредиентов, входящих в данные лекарственные средства.*

9. *Врач предполагает разместить пациента в группу клинических испытаний для точного подбора медикамента и доз. Он исследует информацию по всем (местным, национальным и международным) исследованиям, связанным с болезнью Альцгеймера. Такие данные могут быть размещены в онтологиях FDA, ВОЗ, ClinicalTrials.gov, CitelineTrialTrove и т.д. Академические медицинские институты или фармацевтические компании могут также приглашать пациентов на исследование.*

• *Врач принимает решение о:*

– *записи пациента на клинические испытания, если одно из исследуемых веществ кажется подходящим для лечения и может принести пользу пациенту, или потому, что пациент сам заинтересован в участии в исследовании;*

– обеспечить пациента всей необходимой информацией об этом клиническом испытании: насколько оно подходит ему, есть ли вероятность быть включенным в это испытание и др.

– проверяет, отвечает ли пациент критериям включения / исключения в исследование по запросам к ЭМК

10. Пациент проходит тщательное медицинское обследование (образ жизни, история болезни, геномика, протеомика, метаболизм, УЗИ), чтобы дополнить и обновить существующие данные.

11. Результаты данного обследования повлияют на размещение в определенную группу в данном клиническом испытании (например, группу плацебо или группу с определенной дозой лекарственного вещества; группы в соответствии с разной дозой). Состояния пациента обновляется в ЭМК.

(См. также Приложение 3).

Следующий пример (Табл. 3) демонстрирует пересечение интересов практической медицины, медицинской науки, клинических исследований и бизнеса при семантическом поиске из взаимосвязанных онтологий (онтологии ТМО – Translational Medicine Ontology).

Табл. 3 – Примеры запросов к ТМО

Вопрос к онтологиям	Результат поиска
Со стороны клиницистов	
Какие диагностические критерии для Болезни Альцгеймера?	Перечисляются 12 диагностических критериев включения и 9 критериев исключения.
Покрывает ли медицинское страхование Medicare D назначение препарата Donepezil?	Medicare D покрывает 2 торговых наименования брэнда Donepezil: Aricept и Aricept ODT.
Кто-нибудь из пациентов, страдающих Болезнью Альцгеймера, проходил лечение по другим неврологическим заболеваниям?	Было установлено, что Пациент 2, страдает от Болезни Альцгеймера и депрессии.
По клиническим испытаниям	
Пациент страдает от побочных эффектов препарата, которым лечат Болезнь Альцгеймера. Проводятся ли испытания препаратов, используемых для лечения Болезни Альцгеймера, с другим механизмом действия?	Найдено 438 препаратов связанных с испытаниями при Болезни Альцгеймера, только 58 находятся в активной стадии испытаний и только 2 из них (Doxogubicin и IL-2) имеют зарегистрированные механизмы действия. 78 препаратов связанных с БА имеют установленные механизмы действия.

Подбор пациентов с Болезнью Альцгеймера без APOE4 аллели. Такие пациенты будут необходимы для включения в клинические испытания препарата Варнеузумаб.	Из 4-х пациентов, только один не имеет APOE4 аллель, и может быть хорошим кандидатом для клинических испытаний.
Какие ведутся активные клинические испытания, в которые можно бы было включить Пациента 2 для подбора эффективной и безопасной терапии?	Существует 58 испытаний, в которых Болезнь Альцгеймера сочетается с другими патологиями: в 2-х случаях присутствуют умеренные когнитивные нарушения, у одного – гиперхолестеринемия, у 66 – инфаркт миокарда, 46 случаев фобий, и 126 – депрессии.
Исследования	
Какие гены связаны или являются причиной Болезни Альцгеймера?	Онтологии Diseasesome и PharmGKB указывают по меньшей мере на 97 генов имеющих некоторую связь с Болезнью Альцгеймера.
Какие SNPs последовательности (нуклеотидные последовательности) в геномной ДНК могут быть потенциальными биомаркерами Болезни Альцгеймера?	Онтология PharmGKB показывает 63 SNPs
Какие препараты на рынке лекарственных средств могут быть использованы в том числе и для лечения Болезни Альцгеймера, т.к. они модулируют причастные к этому заболеванию гены?	Найдено 57 лекарственных препаратов или классов препаратов, которые используются для лечения 45 заболеваний в том числе Болезни Альцгеймера, гипертонии/ гипотонии, сахарного диабета и ожирения.

Продолжение примеров использования семантических технологий в интересах врача и пациента можно посмотреть в Приложение .

4.3.1 Использование семантических технологий в интересах врача

Анализируя возможности семантических систем поиска, перспективы их использования **врачом** заключаются в получении оперативного доступа к медицинской литературе, научно-исследовательским материалам, базам данных различных проверенных источников информации для эффективного анализа и поиска возможных путей решения при:

- Постановке или уточнении диагноза в трудных случаях или редко встречающихся патологиях. В этом случае врач может описать найденные у пациента симптомы, синдромы, привести результаты обследования и по этим данным найти аналогичные случаи, описанные в медицинской литературе или хранящиеся в медицинских базах данных. Особый интерес для него представляют описанные случаи с подтвержденными и выверенными диагнозами (наличие патолого-анатомических заключений, протоколов хирургических операций и т.п.).

- Индивидуального подхода к лечению заболевания, вплоть до учета молекулярных процессов, что является залогом успеха и качества оказания медицинской помощи.
- Определении тактики ведения, включая прием лекарств и процедуры; поиске новых способов диагностики и лечения интересующего заболевания; научных школ и коллег, специализирующихся на лечении данного заболевания; клинических испытаний, куда можно разместить пациента для более углубленной работы с ним.
- Прогнозировании развития заболевания и разработки профилактических мероприятий.
- Поддержки принятия клинических решений для систем электронных медицинских карт (ЭМК). Задача заключается в выявлении передового практического опыта для разработки руководящих клинических принципов, основанных на стандартах и источниках знаний, и оптимальное их внедрение в работу с электронной медицинской картой пациента.

4.3.2 Использование семантических технологий в интересах пациента

При работе с семантическими системами поиска **пациент** получает возможности:

- Ведения своих собственных медицинских данных (различные взб-архивы, доступные для семантического поиска)
- Поиска сведений из значимых источников о заболеваниях, методах их профилактики, лечения и диагностики, ведении здорового образа жизни.
- Анализа назначенных лекарств на предмет противопоказаний, совместимости с другими лекарствами, наличие побочных эффектов. Поиска клинических испытаний интересующего препарата, критериев включения в группу.
- Поиска и получения достоверной актуальной и полной информации по аптекам, медицинским учреждениям, врачам, страховым медицинским организациям с учетом отзывов, наличия необходимых видов товаров и услуг, территориального расположения.
- Получения сведений о санитарно-эпидемиологической обстановке и поведении при ее ухудшении.
- Поиска новостной информации о современных достижениях и возможностях медицины.
- Ознакомления с нормативной и правовой информацией относительно получения медицинской помощи.

4.4 Использование семантических технологий в интересах бизнеса

Фармацевтические компании постоянно приобретают проекты, продукты, другие биотехнологические компании, а также и друг друга. Они постоянно формируют альянсы посредством совместных контактов, и подавляющее большинство исходной информации было бы намного проще интегрировать, если бы она хранилась в веб-форматах. Это работа значительно менее исследовательская, чем деятельность в области биомедицинских исследований, так как эти форматы данных уже хорошо разработаны, но экономическая необходимость снижения временных затрат, которые требуются либо для их приобретения, либо для доработки, очень высока.

4.5 Синергетический эффект LOD

Формирование и использование связанных данных LOD дает возможность намного более обосновано выработать стратегические направления в медицине и здравоохранении, постоянно мониторить достижения современной науки, четко выстраивать последовательность их внедрения в практику.

Биологи, химики, генетики, фармакологи-исследователи, специалисты в области биоинформатики, биохимии, биофизики накапливают обширные данные о биологических функциях, химических составах лекарственных веществ, болезнях, их лечении. В дополнение к этим знаниям разрабатываются специализированные хранилища данных о биомаркерах, каскадах реакций и перспективных биологических моделях. Все эти данные размещаются во взаимосвязанных хранилищах (Linked Data), которые работают как единое целое и обеспечивают более эффективное получение ответов на различные запросы уже со стороны практической медицины (**врачей-исследователей, фармакологов, врачей-клиницистов**), позволяя накапливать и публиковать данные о клинических испытаниях, способах диагностики и лечения заболеваний, медицинской статистике. Эти ресурсы, в свою очередь, позволяют эффективно анализировать состояние отрасли **менеджерам здравоохранения** с целью поиска наиболее перспективных решений.

5 Внедрение семантических технологий в сфере здравоохранения России

5.1 Предпосылки внедрения технологии Semantic Web в России

В настоящее время в России созданы все необходимые предпосылки для появления и использования семантических технологий их интеграции как с международными ресурсами LOD, так и с единым информационным пространством электронного здравоохранения РФ. Об этом свидетельствуют следующие факторы:

- достаточно велико и постоянно возрастает количество первичных данных в сфере здравоохранения в электронном виде (базы данных различных МИС, медицинская документация, научные статьи в электронном виде, информационно-поисковые системы и Интернет-источники (Приложение));
- повышается уровень доступа к сети Интернет для граждан и медицинских организаций, что является одним из приоритетов развития ИТ в России;
- повышается общая компьютерная грамотность населения, медицинских специалистов и научных работников;
- развивается оказание электронных медицинских услуг, как государственных, так и коммерческих;
- развивается и совершенствуется ИТ-инфраструктура государства вообще и здравоохранения в частности;
- иницируются и реализуются проекты создания и развития сети электронных библиотек, в том числе в области медицины, например, ФЭМБ (Федеральная электронная медицинская библиотека);
- активно организуются интернет-порталы для профессионального взаимодействия медицинских специалистов по различным вопросам;
- запущен процесс создания единой государственной информационной системы в сфере здравоохранения и социального развития России (Приказы Министерства

здравоохранения и социального развития РФ № 713 от 16.10.2006 «Об утверждении принципов создания единой информационной системы в сфере здравоохранения и социального развития» и № 364 от 28.04.2011 «Об утверждении концепции создания единой государственной информационной системы в сфере здравоохранения». Более расширенный список НСИ приведен в Приложение).

Правительством РФ обозначены следующие основные направления развития российского информационного общества («Стратегия развития информационного общества в Российской Федерации», утвержденной Распоряжением Президента Российской Федерации от 7 февраля 2008 г. № Пр-212.):

- В рамках РФ:
 - формирование единого информационного пространства;
 - повышение доступности для населения и организаций современных услуг в сфере информационных и телекоммуникационных технологий;
 - расширение использования информационных и телекоммуникационных технологий для развития новых форм и методов обучения, в том числе дистанционного образования;
 - внедрение новых методов оказания медицинской помощи населению, а также дистанционного обслуживания пациентов;
 - создание условий для коммерциализации и внедрения результатов научных исследований и экспериментальных разработок, а также расширение обмена научной информацией;
 - повышение качества подготовки специалистов и создание системы непрерывного обучения государственных служащих в области информационных и телекоммуникационных технологий.
- В части международного сотрудничества в области развития информационного общества:
 - участие в международном информационном обмене;
 - участие в международных исследовательских проектах по приоритетным направлениям развития науки, технологий и техники;
 - участие в разработке международных стандартов в сфере информационных и телекоммуникационных технологий, гармонизация национальной системы стандартов и сертификации в этой сфере с международной системой.

5.2 Преимущества использования технологии *Semantic Web* и *LOD*

Предлагаемая технология обладает следующим набором преимуществ по сравнению с использованием традиционных средств поиска информации:

- интеллектуальный поиск по совокупности запрашиваемых параметров с получением существенных данных и отсеиванием «шума», не относящегося к сути запроса, агрегирование (схлопывание, исключение дублирования) информации по идентичным объектам, который по экспертным данным составляет более 75%;
- автоматическая обработка информации за счет унификации ее представления и построение целостной картины интересующего объекта с его связями и атрибутами (характеристиками);

- сохранение в семантической базе знаний связей с первоначальными источниками данных и возможность их извлечения при необходимости;
- возможность работы с любыми значимыми исходными информационными материалами различных форматов, в том числе и неструктурированными, что по данным экспертов позволит дополнительно задействовать 500-600% информации;
- постоянная автоматическая актуализация данных с установлением степени достоверности рассматриваемых источников данных;
- возможность проведения статистического анализа полученных данных;
- возможность расширения предметной области семантической базы знаний простыми средствами (сродни настройке), которое в традиционных приложениях означало бы доработку или переработку структуры базы данных;
- интеграция различных сегментов в глобальную семантическую сеть Semantic Web и возможность использования сведений об идентичных объектах из других семантических баз знаний;
- создание на основе существующего информационного ресурса новых информационных продуктов, услуг и сервисов, в том числе аналитической, статистической и аудиторской направленности.

5.3 Этапы запуска процесса внедрения семантических технологий в российском здравоохранении

Внедрение технологии LD в России имеет определенные преимущества по сравнению с государствами-«первопроходцами» за счет использования опыта внедрения в зарубежных странах.

Для запуска процесса внедрения семантических технологий в российском здравоохранении требуется дать первоначальный импульс созданию сети семантических хранилищ и в дальнейшем обеспечить координацию и взаимодействие вновь образуемых и расширяемых элементов этой сети. В этих целях предлагается:

1. Создать пилотный образец для ограниченной предметной области (семантическое хранилище, пилотные сервисы).
2. Организовать национальный координационный центр, отвечающий за разработку и распространение методологии, стандартов, интеграцию и координацию элементов сети, адаптацию зарубежного опыта, в том числе семантических структур (онтологий).
3. Привлечь отраслевые профессиональные сообщества, органы управления здравоохранения, научные институты для обеспечения высокого уровня доверия к источникам данных.
4. Пропагандировать применение технологии и стимулировать развитие сети за счет привлечения новых участников – владельцев и пользователей знаний, расширения предметной области и сфер применения (новые хранилища, новые источники данных, новые сервисы).
5. Обеспечить обратную связь с конечными пользователями для постоянного улучшения качества предоставляемых знаний.
6. Интегрировать российскую сеть в международную, развиваемую в рамках проекта LOD, предоставив отечественным специалистам дополнительные

возможности и ресурсы за счет уже имеющихся знаний и наработок за рубежом.

6 Перспективы использования и развития ССМЗ в России

Позиционирование ССМЗ как общей расширяемой семантической сети баз знаний позволяет рассматривать ее использование и развитие в следующих направлениях.

1. Использование соответствующих сервисов ССМЗ в качестве справочно-экспертной системы и системы поддержки принятия решений при интеграции с МИС, АИС органов управления здравоохранением и другими системами.
2. Постоянная актуализация и пополнение базы знаний за счет:
 - клинической, научной и прочей информации из различных распределенных и локальных БД лечебных, учебных и научных учреждений;
 - неструктурированной информации, полученной из российских и зарубежных электронных библиотек (в первую очередь ФЭМБ) и других достоверных источников информации;
 - структурированной информации, хранящейся в аналогичных зарубежных базах знаний.
3. Построение сервисов интеллектуального поиска знаний, получения статистических и аналитических данных в интересах различных профессиональных сообществ, учебных заведений, органов управления отраслью, граждан.
4. Использование в качестве инструмента для анализа обратной связи (обращений, отзывов, комментариев), полученной вне официальных запросов и статистики.
5. Использование в качестве основы для создания публичного информационного портала о здравоохранении и здоровом образе жизни в свете современной Концепции информатизации создания единой государственной информационной системы в сфере здравоохранения.

Российский офис W3C
НИУ Высшая школа экономики
Виктор Клинецов

Российский национальный
исследовательский медицинский
университет им. Н.И. Пирогова
Татьяна Зарубина
д.м.н., проф.

Приложение 1. Краткое описание технологии

Semantic Web (Data Web, Web of Data, Web 3.0) – это направление развития Интернета, целью которого является представление информации в виде, пригодном для машинной обработки. Этот вид представления данных – формат RDF – стандартизирован Международным консорциумом всемирной сети ("W3C"). Ожидается, что благодаря созданию Semantic Web можно будет решить проблему обработки лавинообразно растущих объемов неструктурированных данных.

Linked Data – это данные, опубликованные в стандартах Semantic Web и связанные друг с другом. Связываемость данных – это одна из основных возможностей, предоставляемых Semantic Web. Связи между данными повышают удобство использования и ценность данных. Данные, интегрированные в единую семантическую сеть с помощью технологий Linked Data, представляют собой пространство знаний о некоторой предметной области.

Возможность и простота связывания данных в Semantic Web появляются благодаря использованию стандартного формата их представления Resource Description Framework (RDF). Этот формат в Semantic Web играет роль, аналогичную роли формата HTML в традиционном Интернете. Но, в отличие от HTML, формат RDF ориентирован на обеспечение обработки информации не только людьми, но и ЭВМ.

RDF представляет собой простой способ описания экземпляров данных в формате «субъект-отношение-объект», в котором в качестве любого элемента этой тройки (RDF-трипла) используются только идентификаторы ресурсов. В виде RDF-триплов может быть представлена любая информация, любые информационные объекты, которые можно выделить и идентифицировать в сети Интернет. Это могут быть страницы сайтов, документы, рисунки и т.п.

Технологии Semantic Web и LOD позволяет извлекать знания из различных видов источников – структурированных (БД, таблицы) и неструктурированных (интернет-страницы, RSS-ленты, документы форматов TXT, PDF, DOC, XLT) – и представлять их в виде информационных объектов с атрибутами и связями между объектами заданных типов, интересных для потенциальных пользователей информации.

Такое представление информации наглядно и быстро оценивается пользователем, а также способствует улучшению результатов поиска.

Извлеченные объекты и связи объединяются и сохраняются в семантическом хранилище знаний (RDF-хранилище).

Использование семантических хранилищ позволяет осуществлять более точный поиск информации, автоматическую обработку, логические выводы и формирование глобальных баз знаний.

Общая схема системы по преобразованию неструктурированной и структурированной информации в RDF-формат и ее сохранению в семантическом хранилище знаний для дальнейшего использования с применением семантических технологий представлена на Рис. 7.

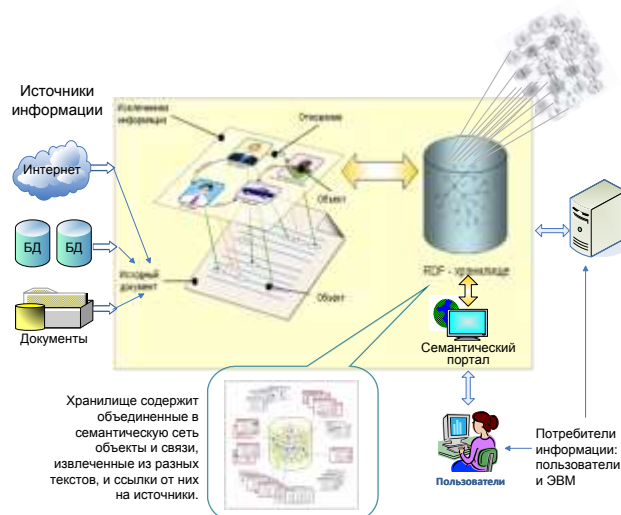


Рис. 7 – Общая схема системы

Примеры типов информационных объектов семантического хранилища медицинских знаний:

- «Лекарственные средства»,
- «Заболевания»,
- «Медицинское оборудование»,
- «Медицинские работники»,
- «Медицинские учреждения»,
- «Научно-исследовательские материалы»

Таблица примеров семантических объектов российского сегмента LOD в сфере здравоохранения приведена в Табл. 4, возможные источники информации, для формирования этих объектов приведены в Табл. 5. Примеры семантической взаимосвязи объектов приведены на Рисунках 8-11.

Накопление и структурирование информации по таким типам объектов создает достаточно полную картину об определенной области медицины и здравоохранения и позволяет проводить анализ информации о нем с использованием как традиционных, так и новых средств (OLAP, Business Intelligence, Data Mining, Text Mining и др.).



Рис. 8 – Семантические связи объекта «Лекарство»



Рис. 9 – Семантические связи объекта «Научная работа»



Рис. 10 – Семантические связи объекта «Заболевание»

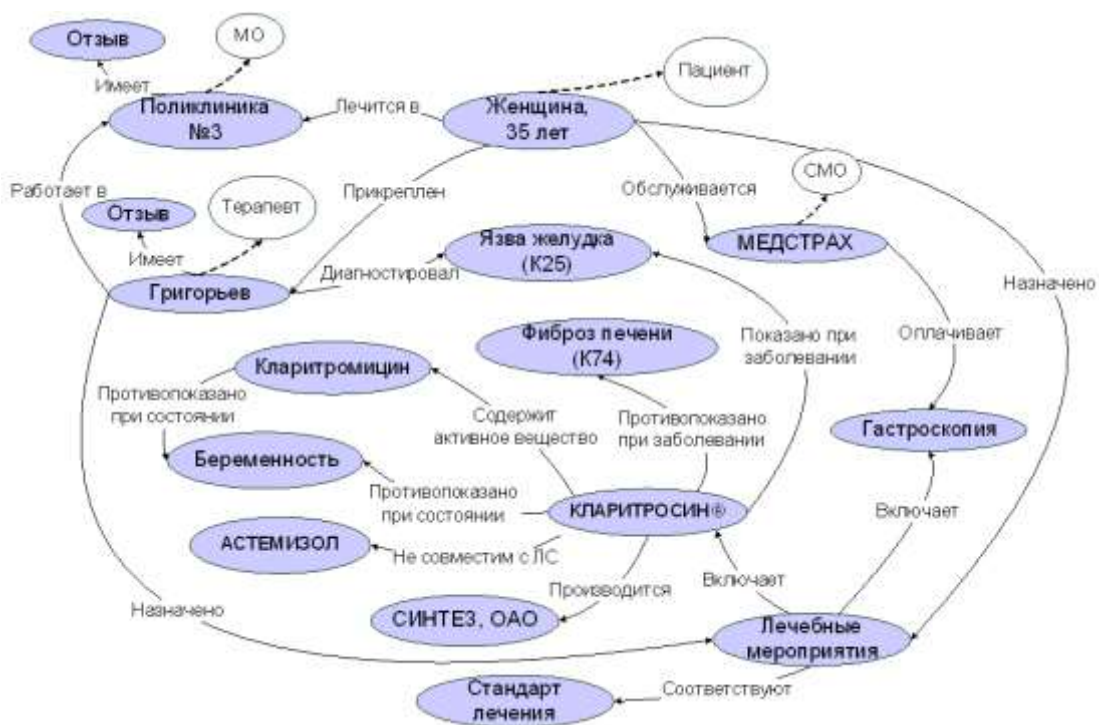


Рис. 11 – Пример семантической сети данных для конкретного случая

Табл. 4 - Примеры типов семантических объектов российского сегмента LOD и их взаимосвязей

№	Тип объекта	Атрибуты типа объекта	Примеры связей типов объектов	Типы источников информации
1.	Лекарства	<ul style="list-style-type: none"> лекарство (наименование) МНН АТХ группа (антибиотики, антигистаминные, гормоны и т.п.) отношение к ЖНВЛП активные вещества механизм действия фирма-производитель дозировка побочный эффект 	<p>Лекарство показано при <i>заболевании</i></p> <p>Лекарство противопоказано при <i>заболевании</i></p> <p>Лекарство является аналогом <i>лекарства</i></p> <p>Лекарство ослабляет действие <i>лекарства</i></p> <p>Активное вещество входит в состав <i>лекарства</i></p> <p>Фирма-производитель производит <i>лекарство</i></p> <p>Лекарство не совместимо с <i>лекарством</i></p> <p>Лекарство имеется в наличии в <i>аптеке</i></p> <p>Лекарство изъято из обращения <i>дата</i></p>	<ul style="list-style-type: none"> сайты производителей лекарств сайты справочников лекарств сайты аптек реестр аптек сайты поиска лекарств по аптекам справочники медицинские сайты медицинских учреждений материалы съездов, конгрессов, форумов и т.п.
2.	Заболевания	<ul style="list-style-type: none"> заболевание состояние симптом синдром лечение (включая оперативное вмешательство) последствия осложнения 	<p>Состояние является побочным эффектом <i>лекарства</i></p> <p>Состояние является эффектом передозировки <i>лекарства</i></p> <p>Заболевание приводит к <i>осложнению</i></p> <p>Заболевание характеризуется <i>симптомом</i></p> <p>Заболевание лечат в <i>медицинском учреждении</i></p> <p>Заболевание приводит к <i>последствиям</i></p> <p>Заболевание лечат <i>лекарством</i></p>	<ul style="list-style-type: none"> справочники медицинские исследования результаты диагностики на электронных носителях, ЭИБ МКБ 10 научные публикации, медицинские порталы и сайты (в т.ч. гос. органов здравоохранения) новости материалы съездов, конгрессов, форумов и т.п.
3.	Медицинское оборудование	<ul style="list-style-type: none"> оборудование тип оборудования (лечебное, диагностическое, для хирургич. лечения и др.) фирма-производитель, вид исследований (лучевое, радиоизотопное и т.п.) тип исследований (рентген, КТ, МРТ, УЗИ и т.п.) 	<p>Фирма-производитель производит <i>оборудование</i></p> <p>Оборудование выполняет <i>тип исследования</i></p> <p>Оборудование применяется при <i>заболевании</i></p> <p>Медицинское учреждение имеет в наличии <i>оборудование</i></p> <p>Оборудование исследует <i>орган</i></p>	<ul style="list-style-type: none"> сайты производителей оборудования справочники медицинские сайты мед. учреждений научные публикации выставки

№	Тип объекта	Атрибуты типа объекта	Примеры связей типов объектов	Типы источников информации
		<ul style="list-style-type: none"> заболевания органы специализация мед. работника медицинские учреждения 		
4.	Медицинские работники	<ul style="list-style-type: none"> работник (ФИО) вид (специализация) образование категория уч. степень уч. звание должность повышение квалификации (сертификаты) стаж работы 	<p>Работник работает в <i>медицинском учреждении</i></p> <p>Работник написал <i>научно-исследовательский материал</i></p> <p>Работник лечит <i>заболевание</i></p> <p>Работник имеет <i>отзыв</i></p>	<ul style="list-style-type: none"> реестр медработников научные публикации сайты мед. учреждений медицинские порталы форумы новости
5.	Медицинские учреждения, организации (МО)	<ul style="list-style-type: none"> МО тип учреждения (поликлиника, стационар, диагностич. центр, НИИ и т.д.) специализация территориальное расположение ведомственная принадлежность мощность форма собственности 	<p>МО лечит <i>заболевание</i></p> <p>МО имеет <i>отзыв</i></p>	<ul style="list-style-type: none"> медицинские порталы и сайты (в т.ч. гос. органов здравоохранения) сайты МО реестр МО паспорт МО форумы
6.	Научно-исследовательские материалы	<ul style="list-style-type: none"> заголовок (работа) Тема (лекарство, исследование, лечение, открытие, заболевание) виды публикаций (диссертация, монография, статья, тезис, обзор и т.д.) уровень публикации (учрежденческий, российский, международный) принадлежность ВАК или рецензируемым м/н источникам автор 	<p>Работа имеет <i>отзыв</i></p> <p>Работа цитирует <i>работу</i></p> <p>Медицинский работник работал над <i>работой</i></p> <p>Медицинский работник цитирует <i>работу</i></p> <p>Работа имеет <i>рецензию</i></p> <p>Работа исследует <i>предмет</i></p> <p>Работа посвящена открытию <i>лекарства</i></p> <p>Работа посвящена лечению <i>заболевания</i></p>	<ul style="list-style-type: none"> периодические медицинские издания (журналы, сборники) учебники, монографии библиотеки сайты медицинских учреждений медицинские порталы и сайты новости, обзоры блоги справочники медицинские форумы

№	Тип объекта	Атрибуты типа объекта	Примеры связей типов объектов	Типы источников информации
7.	Обучение	<ul style="list-style-type: none"> • предмет • специализация • квалификация • вид обучения • учебное учреждение (институты, курсы и т.д.) и его подразделение • объем обучения • цены 	<p>Учебное учреждение имеет в наличии <i>курсы</i></p> <p>Вид обучения стоит <i>цена</i></p>	<ul style="list-style-type: none"> • сайты медицинских учреждений • медицинские порталы и сайты • публикации ВАК
8.	Диагностика	<ul style="list-style-type: none"> • виды диагностики • методы диагностики • описание • показания • мед. работники • цены • аналоги • подготовка к исследованию 	<p>Вид диагностики показан при <i>диагнозе</i></p> <p>Вид диагностики противопоказан при <i>заболевании</i></p> <p>Вид диагностики противопоказан при <i>состоянии</i></p> <p>Вид диагностики производится в <i>МО</i></p> <p>Вид диагностики можно заменить <i>видом диагностики</i></p> <p>Вид диагностики противопоказан при приеме <i>лекарства</i></p> <p>Вид диагностики может привести к <i>осложнениям</i></p> <p>Вид диагностики применяется совместно с <i>лекарством</i></p> <p>Вид диагностики проводится видом медицинского работника</p>	<ul style="list-style-type: none"> • сайты медицинских учреждений и центров • справочники по диагностике • медицинские порталы и сайты (в т.ч. гос. органов здравоохранения) • научные медицинские публикации • новости
9.	Немедикаментоз-ные методы лечения	<ul style="list-style-type: none"> • метод • вид лечения • описание • механизм действия 	<p>МО использует <i>метод</i></p> <p>Метод лечит <i>заболевание</i></p> <p>Метод противопоказан при <i>заболевании</i></p> <p>Медицинский работник использует <i>метод</i></p> <p>Метод имеет аналог <i>метод</i></p> <p>Метод имеет аналог <i>лекарство</i></p> <p>Метод имеет <i>отзыв</i></p> <p>Метод вызывает <i>осложнения</i></p> <p>Метод стоит <i>цена</i></p> <p>Метод совместим с <i>методом</i></p>	<ul style="list-style-type: none"> • сайты медицинских учреждений и центров • справочники • научные медицинские публикации • новости

№	Тип объекта	Атрибуты типа объекта	Примеры связей типов объектов	Типы источников информации
			Метод совместим с <i>лекарством</i>	

Табл. 5 - Перечень возможных источников информации

№	Источники информации (сайты)	Комментарий
1.	Сайт Министерства здравоохранения и социального развития http://www.minzdravsoc.ru/	<ul style="list-style-type: none"> • Законодательная база • Новости здравоохранения России • Перечень ЖНВЛП
2.	Сайт Росздравнадзора http://www.roszdravnadzor.ru/	Законодательная база
3.	Сайт Федерального фонда обязательного медицинского страхования http://www.ffoms.ru/	Сведения о территориальных фондах ОМС
4.	Сайт Московского городского фонда обязательного медицинского страхования http://www.mgfoms.ru/	Сведения о СМО и МО г. Москвы
5.	Сайт Центрального НИИ организации и информатизации здравоохранения http://zo.mednet.ru/	<ul style="list-style-type: none"> • Данные статистики • Сведения о классификаторах ЗО РФ
6.	Портал по ведению государственного реестра лекарственных средств http://grls.rosminzdrav.ru/	<ul style="list-style-type: none"> • Государственный реестр лекарственных средств • Государственный реестр предельных отпускных цен
7.	РЛС.net – регистр лекарственных средств России http://www.rlsnet.ru/	<ul style="list-style-type: none"> • Справочник лекарств, действующие вещества, взаимодействие лекарств, справочник болезней (МКБ-10), АТХ-классификация, производители лекарств • энциклопедия БАД
8.	Справочник лекарственных препаратов Видаль http://www.vidal.ru/	<ul style="list-style-type: none"> • Справочник лекарств, действующие вещества • взаимодействие лекарств • АТХ-классификация • производители лекарств • БД «Врачи России»
9.	Сайт об обращении лекарственных средств http://www.regmed.ru/	<ul style="list-style-type: none"> • Реестр зарегистрированных ЛС и ТКФС • Реестр зарегистрированных цен на ЛС, реестр данных по изъятым препаратам
10.	Сайт о классификаторах http://www.classbase.ru	<ul style="list-style-type: none"> • Общероссийские классификаторы (ОКАТО и т.д.) • Медицинские классификаторы (МКБ-10, ОК ПМУ, ОК СКМУ)
11.	Электронная версия МКБ-10 http://www.mkb10.ru/	Коды болезней
12.	Российский фонд фундаментальных исследований http://www.rfbr.ru/	Статьи, новости науки, данные о новых исследованиях и разработках.
13.	Клиническое подразделение Российского Научного Медицинского Общества Терапевтов http://www.ingorts.ru/index.php/ru	Статьи, новости науки, данные о новых исследованиях и разработках.

№	Источники информации (сайты)	Комментарий
14.	Всероссийское Научное Общество Кардиологов http://www.cardiosite.ru	Статьи, новости науки, данные о новых исследованиях и разработках.
15.	Национальная ассоциация по борьбе с инсультом http://www.nabi.ru/	Статьи, новости науки, данные о новых исследованиях и разработках.
16.	Международный союз по борьбе с инфекциями, передаваемыми половым путем http://www.iusti.ru/	Статьи, новости науки, данные о новых исследованиях и разработках.
17.	Национальное научно-практическое общество скорой медицинской помощи http://cito03.ru/	Статьи, новости науки, данные о новых исследованиях и разработках.
18.	Национальное научное Общество Кардиоваскулярная Профилактика и Реабилитация http://www.cardioprevent.ru/	Статьи, новости науки, данные о новых исследованиях и разработках.
19.	Российская ассоциация специалистов перинатальной медицины http://raspm.ru/	Статьи, новости науки, данные о новых исследованиях и разработках.
20.	Научно-исследовательский институт неотложной детской хирургии и травматологии http://doctor-roshal.ru/	Сведения о деятельности института
21.	Научный центр акушерства, гинекологии и перинатологии имени академика В.И. Кулакова http://www.ncagip.ru/	Сведения о деятельности центра
22.	Вся медицина в интернет http://www.medlinks.ru/	Книги, статьи, словари, каталоги, форумы, новости медицины
23.	Портал учреждений здравоохранения Российской Федерации http://uzrf.ru/	Законодательство, новости медицины, статьи, сайты ЛПУ
24.	Электронная медицинская библиотека http://medlib.ws/	Электронные версии книг по медицинской тематике
25.	Профессиональные медицинские порталы http://vitaportal.ru/ http://med-info.ru/ http://panacea.ru/ http://vrachirf.ru/	Информационно-поисковые порталы медицинских сообществ
26.	http://www.ros-med.info	Медицинская информационно-справочная сеть

Приложение 2. Пример использования семантических технологий в интересах врача и пациента

У Пациента диагностирована болезнь Альцгеймера, Назначенные препараты не приводят к ожидаемым результатам. Необходимы дополнительные способы решения проблемы:

- 1. Подбор альтернативных лекарственных средств, направленных на лечение болезни Альцгеймера. При поиске таких средств пациент хочет знать терапевтические и побочные эффекты, которые имеют эти альтернативные лекарства.*
 - Онтология TCM может быть использована для поиска средств растительного происхождения, которые способствуют борьбе с болезнью Альцгеймера и их предполагаемых эффектов.*
 - Онтологии SIDER и DailyMed совместно могут быть использованы для поиска информации о побочных эффектах ингредиентов, входящих в данные лекарственные средства. Так, совместно SIDER и TCM обеспечат дополняющую (друг друга) информацию об эффектах альтернативных медикаментов. Например, информация о побочных эффектах из SIDER:*
 - для ингредиента аденозин – гипертензия, боли в спине, онемение, нервозность, остановка сердца и т.д.*
 - для ингредиента тестостерон – гипертензия, дезориентация, боль в животе, головокружение и т.д.*
 - для ингредиента уксусная кислота – боль, инфекции, ацидоз.*
- 2. Получение информации по клиническим испытаниям этих, созданных на растительной основе, препаратов.*
 - Онтологии LinkedCT и DailyMed могут предоставить такую информацию.*
 - Например, в онтологии LinkedCT приводится отрицательное испытание о влиянии свойств растения Ginkgo на слабоумие при болезни Альцгеймера.*
 - Это испытание также упоминается в выдержке из PubMed.*
- 3. Выяснение связи “Gingko biloba” с болезнью Альцгеймера, чтобы увидеть входят ли гены (так называемые гены гинкго) в список тех, о которых докладывалось в исследовательском сообществе по борьбе с данным заболеванием. Если об этих генах не было заявлено в сообществе, как о главных генах, ассоциированных с болезнью Альцгеймера, пациент хочет узнать, ассоциировались ли вообще эти гены с болезнью Альцгеймера и с какими другими заболеваниями эти гены ассоциировались.*
 - Онтология TCM может быть использована для поиска генов, ассоциированных с “Gingko biloba” для болезни Альцгеймера. Эти гены -- MART, AGPS, APP, CASP3, CREB1, ADAMTS2, TTR, APLP2, ACHE, APOE, MARK1.*
 - Онтология Diseasesome может быть использована для того, чтобы выяснить, были ли вообще гены Gingko ассоциированы с болезнью Альцгеймера. Результаты поиска демонстрируют, что гены APP, TTR, ACHE, APOE отражены в Diseasesome и гены APP и APOE также отображены в базе данных Diseasesome как ассоциированные с болезнью Альцгеймера.*

- *Diseasome* может также быть использован для поиска других заболеваний, с которыми ассоциируются гены *Gingko*.
4. Уточнение информации об ингредиентах “*Gingko biloba*” (может быть использован TSM) с целью безопасности его назначения по следующим вопросам:
- наличия у пациента аллергии на какой-либо из ингредиентов,
 - противопоказаний,
 - неблагоприятного взаимодействия с другими лекарственными веществами.

Приложение 3. Перечень нормативно-справочной информации

1. Стратегия развития информационного общества в Российской Федерации. Утв. Распоряжением Президента Российской Федерации от 7 февраля 2008 г. № Пр-212.
2. Государственная программа Российской Федерации «Информационное общество (2011-2020 годы)». Утв. Распоряжением Правительства РФ от 20 октября 2010 г. №1815-р.
3. Распоряжение Правительства РФ от 25 апреля 2011 г. № 729-р «Перечень услуг, оказываемых государственными и муниципальными учреждениями и другими организациями, в которых размещается государственное задание (заказ) или муниципальное задание (заказ), подлежащих включению в реестры государственных или муниципальных услуг и предоставляемых в электронной форме».
4. Распоряжение Правительства РФ от 17 декабря 2009 г. N 1993-р «Сводный перечень первоочередных государственных и муниципальных услуг, предоставляемых органами исполнительной власти субъектов Российской Федерации и органами местного самоуправления в электронном виде, а также услуг, предоставляемых в электронном виде учреждениями субъектов Российской Федерации и муниципальными учреждениями».
5. Приказ Министерства здравоохранения и социального развития РФ от 16 октября 2006 г. № 713 «Об утверждении принципов создания единой информационной системы в сфере здравоохранения и социального развития».
6. Приказ Минздравсоцразвития РФ от 28 апреля 2011 г. № 364 «Об утверждении концепции создания единой государственной информационной системы в сфере здравоохранения».
7. Методические рекомендации по составу прикладных компонентов регионального уровня единой государственной информационной системы в сфере здравоохранения, а также функциональные требования к ним, обязательные для создания в 2011–2012 годах в рамках реализации региональных программ модернизации здравоохранения.
8. Единая система межведомственного электронного взаимодействия. Постановление Правительства РФ от 08.09.2010 г. № 697.
9. Перечень документов (сведений), обмен которыми между органами и организациями при оказании государственных услуг и исполнении государственных функций осуществляется в электронном виде. Распоряжение Правительства РФ от 17.03.2011 № 442-р.
10. Технические требования к универсальной электронной карте и федеральным электронным приложениям. Постановление правительства РФ от 24.03.2011 № 208.

Используемые термины и сокращения

Термин	Описание
Business Intelligence	Инструменты, используемые для преобразования, хранения, анализа, моделирования в ходе работы над задачами, связанными с принятием решений на основе фактических данных.
Data Mining	Интеллектуальный анализ данных – собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.
HCLS	Health Care and Life Sciences – Здоровоохранение и наука о жизни
LD	Linked Data – технология связывания данных.
LOD	Linking Open Data – открытые связанные данные.
LODD	Linking Open Drug Data – Банк открытой информации по лекарственным средствам
OLAP	Online analytical processing – аналитическая обработка в реальном времени – технология обработки информации, включающая составление и динамическую публикацию отчётов и документов.
RDF	Resource Description Framework – это разработанная консорциумом Всемирной паутины модель для представления данных, в особенности — метаданных.
Semantic Web	Семантическая паутина – направление развития Всемирной паутины, целью которого является представление информации в виде, пригодном для машинной обработки
SNPs	Single nucleotide polymorphisms - это однонуклеотидные позиции в геномной ДНК, для которых в некоторой популяции имеются различные варианты последовательностей (аллели)
Text Mining	Процесс извлечения информации из текстовых данных на основе обнаружения в них закономерностей
TMO	Translational Medicine Ontology – онтология трансляционной медицины
W3C	<i>World Wide Web Consortium</i> , W3C – Консорциум Всемирной паутины
АИС	Автоматизированная информационная система
АТХ	Анатомо-терапевтическо-химическая классификация лекарственных средств
БАД	Биологически активная добавка
БД	База данных
ВАК	Высшая аттестационная комиссия
ВУЗ	Высшее учебное заведение
ЖНВЛП	Жизненно необходимые и важнейшие лекарственные препараты
КТ	Компьютерная томография
ЛС	Лекарственное средство
МО	Медицинская организация
МИС	Медицинская информационная система
МКБ 10	Международная классификация болезней, 10 пересмотр

Термин	Описание
МНН	Международное непатентованное наименование лекарственного средства
МРТ	Магнитно-резонансная томография
НИИ	Научно-исследовательский институт
НЦК	Национальный центр компетенции и поддержки медицинских семантических технологий
ОКАТО	Общероссийский классификатор объектов административно-территориального деления
ОК ПМУ	Общероссийский классификатор простых медицинских услуг
ОК СКМУ	Общероссийский классификатор сложных и комплексных услуг
ОМС	Обязательное медицинское страхование
СМО	Страховая медицинская организация
ССМЗ	Семантическая сеть знаний в сфере здравоохранения, химико-биологических и медицинских исследований и разработок
РФ	Российская Федерация
ТКФС	Типовая клинико-фармакологическая статья
УЗИ	Ультразвуковое исследование
ФЭМБ	Федеральная электронная медицинская библиотека
ЭВМ	Электронная вычислительная машина
ЭИБ	Электронная история болезни