

СОЗДАНИЕ СОВРЕМЕННОЙ ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ХРАНЕНИЯ, КАТАЛОГИЗИРОВАНИЯ И ПОИСКА НАУЧНОЙ ИНФОРМАЦИИ

Лысенко А.И.

*Лысенко Алексей Иванович – студент,
специальность: медицинская биофизика,
медико-профилактического факультета,
Новосибирский государственный медицинский университет, г. Новосибирск*

Аннотация: развитие сети Интернет предоставило создателям интеллектуальных информационных систем новые возможности, связанные с одновременным доступом к множеству разнородных источников данных, что открывает широкие перспективы в развитии более совершенных технологий получения знаний. Однако многие современные исследования в области интеллектуального поиска опираются на неявное предположение о возможности широкого распространения более или менее подробной стандартизации представления информации. Реализация подобных проектов, прежде всего, концепции *Semantic Web* консорциума W3, позволила бы вывести работу с информацией на качественно новый уровень, но одна из основных особенностей сети Интернет как феномена цивилизации заключается в том, что развитие сети изначально носит децентрализованный характер, поэтому многие веб-сайты, содержащие важную информацию из той или иной предметной области, не соответствуют рекомендациям консорциума W3. В данной статье приводится обзор решений проблем с выбором платформ для создания интеллектуальной информационной системы для хранения, обработки и разделения на каталоги существующих публикаций в медицинской отрасли, а также предлагаются системы автоматизации основных этапов научно-информационного процесса (включая создание глобальных тезаурусов и онтологий).

Ключевые слова: интеллектуальная система, научные базы данных, интеграция, платформа, стандартизация, онтология, хранение, создание каталогов, поиск информации.

На сегодняшний день образовательная система в области медицины и центры здравоохранения достаточно компьютеризированы и снабжены современным сетевым оборудованием.

Во всем мире интеллектуальные информационные комплексы научных организаций максимально сосредоточены на создании единой системы хранения, поиска, архивирования и каталогизирования научной информации как современный библиотечно-информационный центр, который обладает собственными электронными ресурсами и обеспечивает доступ к мировым информационным порталам, с помощью глобальной сети Интернет.

Кроме того, развитие науки сопровождается непрерывным накоплением важной информации, которую необходимо сохранить для дальнейшего применения.

Сегодня специализированные поисковые системы, которые осуществляют функцию – поиска необходимой научной информации непосредственно в Интернете, исчерпали свой ресурс, они не обеспечивают достаточного уровня семантического анализа документов, предоставляя информацию в виде «подборок», заставляя пользователя системы анализировать и выбирать важный материал самому, что является достаточно трудоемким и не практичным процессом.

Создание интеллектуальной информационной системы, в которой наряду с традиционной информационной системой, присутствует и рассуждающая информационная система (формализующая правила логического вывода), а также интеллектуальный интерфейс (диалог, графика и т.п.) является важнейшей целью для развития библиотечно-информационных сетевых центров. Поэтому изучение и проведение анализа существующих интеллектуальных информационных систем для хранения и поиска научной информации, позволяет выявлять существующие проблемы актуализации необходимого материала, разрабатывать новейшие единые интегрированные программные комплексы, которые обеспечат минимальное участие человека в процессе контроля над качеством обработки информации, [1].

Создание интеллектуальной системы обработки и хранения научного материала возможно за счет современных информационных технологий, которые смогут интегрировать защищенную единую и безопасную среду с помощью технико-технологических средств путем разработок единого информационного пространства.

Разработка интеллектуальных информационных смоделированных сетей на базе новейшей автоматизированной системы позволят не только разграничить информацию по каталогам, но и смогут обеспечить постоянное обновление научно-информационного процесса. Этот аспект расширит доступ к научной информации, создаст единую систему из разнородных блоков информации.

Осуществление широкой доступности для пользователей к информационным ресурсам обеспечивается созданием web-сайтов, представляющие открытую библиотеку в единой мировой информационной системе, используя сеть Интернет. Для полноты картины о существующих информационных электронных библиотеках научной медицинской информации были собраны данные о размещении информационных ресурсов на различных сайтах медицинских высших образовательных учреждений. Из 50 медицинских вузов в открытом доступе имеются web-сайты 38 электронных библиотек. В таблице 1 представлены информационные ресурсы библиотек этих вузов.

Таблица 1. Информационные ресурсы библиотек вузов

№ п/п	Предоставленные ресурсы/информационные системы	Количество существующих библиотек	Количество библиотек, в %
1	Электронный каталог фонда библиотеки	18	44,5%
2	Региональный корпоративный межвузовский электронный каталог	4	11,5%
3	Информационные ресурсы на CD-дисках	5	14%
4	Интернет-ресурсы	16	41,5%
5	Предоставление доступа к внешним БД (НЭИКОН, EBSCO)	14	37%
6	Электронная библиотека собственных документов вуза	9	24%
7	Библиографические справки/Виртуальные выставки	10	5%

Оптимальной формой предоставления библиотечной информации и обеспечения функционального процесса интеллектуальной системы и открытого доступа являются электронные каталоги библиотек, составляющие основную долю информационного потенциала сети Интернет.

Важно отметить, что на сегодняшний день, электронные каталоги являются наиболее ценным и практичным продуктом интеллектуальной информационной системы хранения, поиска и каталогизирования материала. Кроме того, в широком доступе и удобном формате представлены публичные библиотеки книгохранилищ различных отраслей, как технических, так и медицинских публикаций, как университетских, так и прочих научных организаций.

В электронных библиотеках медицинских учреждений каталогизирование ведется в различных системах – «ИРБИС», «OPAC-Global», АБИС, которые поддерживают форматы RUSMARC (UNIMARC) [2].

В последние годы на рынок автоматизации обмена медицинскими данными и потоком научной информации были внедрены ведущие западные производители, такие как: InterSystems, Microsoft, IBM.

Реализация механизмов единого информационного пространства и предоставление широкодоступного обмена медицинским научным материалом основана на разработке качественной, функциональной, специализированной платформы.

Сегодня, как было сказано выше, имеется большое количество различных серверных платформ для реализации проектов по созданию интеллектуальных информационных систем хранения, поиска и каталогизирования данных.

Одним из развитых набором функций имеет пакет Oracle, который представлен для работы с языком Java и доступа к серверам с помощью сети Интернет. Данная платформа автоматизации работы с информацией обладает системой оптимизации широкого и одновременного доступа. Недостаток системы – это сложность администрирования. Хотя, разработка интеллектуальной системы на этой платформе позволяет окупаться проектам за короткий срок, обеспечивая надежную и эффективную работу. Основными свойствами такой системы являются [3]:

- 1) Высокая надежная и эффективная работа;
- 2) Каталогизирование информации по разделам и профилям, что позволяет максимально быстро и эффективно контролировать огромный поток гигабайтов данных;
- 3) Использование универсального средства защиты единого пространства;
- 4) Применение максимизации обработки запросов;
- 5) Обработка индексации в битовом отображении;

- 6) Наличие свободных таблиц (в других системах все таблицы создаются и заполняются только при создании системы данных);
- 7) Применение параллельных способов разделения потока запросов;
- 8) Наличие огромного функционального интерфейса мониторинга, разработки, администрирования;
- 9) Ориентация на облачные технологии.

Проектные решения, которые не уступают платформе Oracle, находятся в применении систем DB2 (IBM) и в разработках на базе MySQL.

Для информационных концепций решения по созданию интеллектуальных систем хранения данных на базе MySQL, имеются ряд преимуществ и характеристик:

- 1) Легкость в администрировании;
- 2) Возможность осуществления Web-подключений;
- 3) Наличие средств автоматических настроек и специальных шаблонов;
- 4) Наличие инструментария для синхронизации, анализа параметров и данных;
- 5) Возможность применения через удаленный доступ.

В настоящее время MySQL – это информационная платформа, которая представляет собой полнофункциональный пакет, подходящий для реализации интеллектуальной системы малых и средних медицинских учреждений. Важно отметить, что MySQL уступает системе Oracle по двум немаловажным свойствам: средства разработки и программируемость.

В связи с развитием «глобальной паутины» существенно возрастает интерес к решению проблем хранения и обработки уже имеющихся в учреждениях научных публикаций и материалов. В зарубежных странах и странах СНГ такие существенные проблемы решаются путем приобретения дорогостоящих систем централизации хранения данных, которые поступают от различных источников (систем структурных подразделений). Одним из ярких представителей такой системы является комплекс PI System (OSIsoft), которая создана на платформе Oracle. Но, завышенные стоимости таких систем является одним из недостатков применения ее в медицинских образовательных учреждениях, в организациях средних и крупных размеров. Важно отметить, что применение системы PI System требует установки отдельного сервера. Поэтому для этой системы эффективнее применение специальных приложений, которые сочетают в себе функции обработки, поддержки, хранения научной информации. Такую систему хранения разработала компания ICONICS, платформа имеет следующее название – Hyper Historian. Данное программное обеспечение для обработки данных, которое предназначено для особо важных приложений, требующие постоянный доступ и сбор информации,[3].

Система Hyper Historian является гибкой и обеспечивает высокий уровень надежности. Такой продукт позволит создать:

- 1) Простую одноузловую систему хранения данных;
- 2) Многоуровневую систему данных, которые возможно будет разделить на подуровни хранения, создавая каталоги различных профилей; (в том числе возможность перемещать информацию и данные между подуровнями системы);
- 3) Многоуровневую систему резервных копий файлов для аварийного восстановления работоспособности (причем гарантируется возможность непрерывности сбора информации и исключения их удаления или потери при отказе всей системы или серверов).

В Hyper Historian существует возможность интегрирования пакета ICONICS BizViz, позволяющая создавать систему отчетов, анализа, технологий других платформ или возможность применения облачных технологий.

Таким образом, стандартизированный интерфейс системы запросов SQL дает возможность интегрировать с системой Microsoft SQL уже имеющихся баз данных. Интеллектуальные системы хранения позволяют унифицировать и разделять на каталоги научную информацию и данные, применив расписание и триггеры. Такой системный подход к разработке информационного пространства позволяет не только проводить автоматический контроль над размещаемым материалом, но и экономит место на диске, а также дает возможность создавать резервные копии имеющихся файлов для длительного сохранения и их поиска.

По мнению ведущих экспертов в IT – технологиях, в настоящее время пользователи тратят до 80-90% бюджета на разработки и поддержание работоспособности и целостности системы хранения имеющейся информации. В развитых автоматизированных платформах интеллектуальных информационных системах необходимость расширения пространства является более чем актуальной задачей. Такая проблема позволила создать единую «облачную» систему ресурсов для осуществления организации вычисления, обмена и хранения данных. Эти системы позволяют заимствовать инфраструктуру и пространство, где требуется оплата за реальный объем использования ресурсов [4]. Важно отметить, что «облачные» технологии имеют и ряд недостатков: ограничение по применению пространства, проблем с обеспечением безопасности и защиты хранимой информации задержка производительности, дополнительные финансовые расходы на поддержание объема места хранения. Но популярность

«облачных» технологий и создания виртуального пространственного места хранения с взаимодействием уже имеющихся разработанных платформ будет расти. Самое главное в развитии отрасли интеллектуальных информационных систем – это развитие современных систем информатизации и сетевых технологий. Уже сегодня разработчики таких технологий поддерживают виртуализацию и работоспособность виртуального пространства на одном персональном устройстве и управляют информацией как единой системой [4].

Отдельным и немаловажным механизмом создания интеллектуальной системы является применение технологии порталов для хранения медицинской информации.

Портал – это информационная система, которая обеспечивает организацию (учреждение) необходимыми для решения задач проекта информационными разделами и функциональными модулями путем интегрирования сайтов с применением набора типовых приложений. Такие порталные технологии появились в 1997 году, пройдя путь эволюции и несколько поколений развития, [5]. Одним из ведущих и важнейших решений порталных технологий является продукт OracleAS Portal. Эта платформа выступает в роли образующей системы, обеспечивающей участникам авторизированный и многоканальный широкополосный доступ ко всем имеющимся приложениям, которые участвуют в хранении и обработки медицинской информации.

OracleAS Portal интегрирован с приложениями на базе Oracle Application Server 10g. Такая система позволяет обеспечить максимальный уровень доступности, масштабируемости, защиты информации и администрирования [2].

Одной из перспективных концепций стандартов широкого поиска информации является концепция Semantic Web, которая сегодня является наиболее прогрессирующей. Данная концепция способна развить поисковую систему сети Интернет как универсальную модель сетевого обмена онтологических данных. Данная концепция интернета была разработана и предложена в реализацию в 2001 году Т. Бернерс-Ли, Дж. Хендлер и О. Лассила [3]. Концепция характеризуется тем, что содержание страниц вносится конкретизированная структура, позволяющая распознавать и анализировать смысл обработанных данных, в результате выполнять запросы пользователей более точно.

Semantic Web характерен именно для естественных научных предметных областей, так как изучение этих дисциплин форматизировано и общедоступно, а также более удобно для онтологического анализа. Таким образом, применение Semantic Web открывает возможность интегрированию медицинских научных материалов за счет конкретизированной спецификации из разнородных источников. Одной из особенностей развития Semantic Web явилась интеграция на публикацию открытой связанной информации.

Таким образом, архитектура семантического Веба (Semantic Web) основана на следующем принципе: разработка языка, на котором будет выражаться собранная информация по предметной области, перевод всех сайтов и ресурсов на созданный язык программирования, разработка программ (обработка запросов, логический вывод информации). Вышесказанное позволяет сделать вывод о том, что цель Semantic Web – это создание языковой модели, на которой будут описываться все существующие данные (информация) по предметной области. Т. Бернерс-Ли разграничил синтаксис и семантику языка описания: RDF (Resource Description Framework) – язык синтаксиса Semantic Web с применением онтологий для распознавания смысла слов; OWL (Ontology Web Language) – Язык онтологий.

В результате, Semantic Web – это многоуровневая архитектура с использованием стандарта кодирования символов.

При переходе к практическому применению онтологии концепции Semantic Web, важно отметить, что создание «информационного фильтра» позволяет существенно сократить время на поиск необходимой научной информации, а разделение и распознавание смыслового обозначения и интегрированных онтологий обеспечивает достоверное получение информации. На базе такой концепции построена вся система сети Интернет, а уже применение и развития этого алгоритма, дает перспективу на будущее для создания интеллектуальной оптимальной информационной системы для хранения, обработки и использования научного материала в различных областях.

Применение системного подхода Semantic Web к разработке портала информационной базы данных предоставит возможность не только сконцентрировать информационные публикации, но и актуализировать информацию, которая хранится в различных приложениях и источниках, с разработкой каталогизированной структуры научных данных.

Список литературы

1. Добко Т.В. Система электронных библиотек // Библиография, 2014. № 2.
2. Костюк К. Рынок электронных ресурсов в настоящем и будущем // Университетская книга, 2015 № 2.

3. OracleAS Portal Transaction Base: Создание информационной системы медицинского высшего образовательного учреждения. Oracle. Россия, 2016.
4. Мешечак Н.А. Модель комплексного доступа к медицинским информационным ресурсам// Научная и технические библиотеки, 2013. № 6.
5. Ботуз С. Интеллектуальные интерактивные системы и технологии управления удаленным доступом: Учебное пособие / С. Ботуз. М.: Солон-пресс, 2014. 340 с.