

УДК 004.65

СОВРЕМЕННЫЕ НАУЧНЫЕ ИССЛЕДОВАНИЯ И СИСТЕМЫ ХРАНЕНИЯ БАЗ ДАННЫХ ДЛЯ ИХ ПРОВЕДЕНИЯ

Цицина Анна Сергеевна

аспирант

Научный руководитель: **Гайнутдинова Екатерина Валерьевна**

к.ф.н., доцент

«Астраханский государственный технический университет»

Аннотация: Научные исследования являются источником технологий, материалов и механизмов, при помощи которых становится возможным создавать продукты лучшего качества, меньшей стоимости, создавать методы лечения болезней, бороться со стихийными бедствиями и т.д. В работе рассматриваются современные методы хранения информации в информационных системах, обработка информации, а также её представления. Основное внимание уделяется современным документо-ориентированным базам данных, которые способны сохранить данные в информационной системе.

Ключевые слова: Исследования, исследовательские работы, научная информация, хранилище данных, база данных, информационные ресурсы, информационная система.

MODERN SCIENTIFIC RESEARCH AND DATABASE STORAGE SYSTEMS FOR THEIR IMPLEMENTATION

Tsitsina Anna Sergeevna

Gainutdinova Ekaterina Valeryevna

Abstract: Scientific research is a source of technologies, materials and mechanisms with the help of which it becomes possible to create products of better quality, lower cost, create methods of treating diseases, fight natural disasters, etc. The paper considers modern methods of information storage in information systems, information processing, as well as its presentation. The main attention is

paid to modern document-oriented databases that are able to store data in an information system.

Key words: Research, research papers, scientific information, data warehouse, database, information resources, information system.

В связи с тем, что формирование тематик и направление в разных фондах и агентствах ведется независимо друг от друга, часто возникает дублирование работ. При чем речь идет как о дублировании выполняемых работ, так и о дублировании результатов исследований. Также может, встречается дублирование выполняемых работ с работами, выполненными в период существования СССР, когда было произведено большое количество научных работ.

Результаты исследований оформляются техническими отчетами, актами и другой отчетной документацией, которая, как правило, хранится в печатном виде на бумаге в архивах заказчика и исполнителя. Для получения того или иного отчета необходимо проводить длительную переписку с исполнителем или заказчиком отчета, но, что важнее — информацию о том, что тот или иной отчет существует в большинстве случаев практически невозможно найти. Научные публикации по результатам исследований в профильных журналах выпускаются не всегда, а накопленное число исследований и широкий спектр различных изданий делает поиск данных, не опубликованных в сети Интернет, невероятно сложным.

Любая база данных является составной частью некой информационной системы (ИС), которая подразумевает не только хранение данных, но и их обработку. Поэтому, проектированию данных всегда сопутствует (а чаще предшествует) проектирование алгоритмов их использования [2, с.26-27].

Система проведения и организации научных работ, существующая в наше время, была заимствована от СССР и с момента образования не подверглась особым изменениям. На сегодняшний день существуют следующие аспекты модернизации системы выполнения научных работ [5, с.65-66]:

- Повсеместное использование персональных компьютеров и сети Интернет, для доступа к справочной информации;
- Большое количество накопленных научных отчетов, существующих в печатном виде;

– Использование достижений различных отраслей при создании инновационной техники;

– Развитый рынок материалов и услуг, позволяющий реализовать практически любой поисковый эксперимент с небольшими затратами, до открытия полномасштабной НИР (научно-исследовательской работы).

Можно принять следующие меры по повышению эффективности научных работ:

1) Создание единой формы «Результаты научного исследования», с обязательной публикацией в сети Интернет на специальном портале после выполнения НИР.

2) В техническом задании (ТЗ) на выполнение НИР описывать результат, который должен быть получен в ходе работы.

3) Внедрять оптимизированную структуру организации научно-исследовательских предприятий, основанную на функционировании трех подразделений: подразделения постановки проблем и вопросов, подразделения выдвижения научных теорий/гипотез и подразделения реализации экспериментов (технического подразделения).

4) Периодические выделения средств научным организациям на реализацию поисковых экспериментов.

Вопросы повышения эффективности научно-исследовательской работы в вузах сегодня в зоне особого внимания и профессионального сообщества, и государственных органов.

Как показывают глобальные политические и экономические процессы последнего времени, национальная безопасность может быть обеспечена только при условии максимальной независимости ее экономики от внешних факторов. В первую очередь это касается наукоемких отраслей, где необходим качественный инновационный прорыв, который бы позволил создать основу для конкурентоспособной высокотехнологичной экономики. Очевидно, что решение этой задачи требует наличия трех составляющих – кадровой, материально-технической и финансовой. Представляется, что именно университеты могут стать точками роста, обеспечивающими ее инновационное развитие. Действительно, специфика деятельности университетов позволяет эффективно интегрировать проблемы подготовки кадров, проведения научных исследований и коммерциализации результатов интеллектуальной деятельности. В этой связи было бы логично включить в

перечень критериев эффективности вузов и показатель инновационной активности (с учетом специфики вузов).

Было бы целесообразно возобновить государственные программы развития инновационной деятельности вузов. В ее рамках ежегодно проводить конкурс программ инновационной деятельности с выделением победителям финансовой поддержки для завершения коммерчески перспективных НИР и создания малых инновационных предприятий.

Следует продолжить работу по сокращению количества диссертационных советов, поощрять практику объединения сохранных или вновь создаваемых советов. Представляется, что диссертационные советы должны функционировать только на базе вузов, реально лидирующих в соответствующей области знаний. Они должны привлекать к своей работе и видных специалистов из других университетов и академических организаций.

Сегодня университеты способны осуществлять подготовку кадров для инновационной экономики. Они выращают новое поколение специалистов, способных создавать инновационные разработки и знающих пути их выхода на рынок, а при необходимости – готовых создавать инновационные предприятия. Во многом это возможно благодаря созданию в университетах условий для научно-исследовательской деятельности, нацеленной на производство необходимых для развития ведущих отраслей экономики инноваций.

При наличии большого количества накопленных в советский и постсоветский период научных отчетов, разобщенности фондов и исследовательских организаций, и повсеместном использовании сети интернет, рационально создать единый портал результатов научных исследований для осуществления удобного и быстрого поиска отчетов о выполненных работах, который был бы доступен как сотрудникам научных и исследовательских организаций, так и чиновникам, осуществляющим проверку актуальности той или иной работы [4, с.111-112].

Форму результата научного исследования рациональнее составить по трем пунктам:

- 1) На решение какой проблемы было направлено исследование;
- 2) Какая гипотеза была выдвинута;
- 3) Каким способом гипотеза была проверена.

На каждую проверенную гипотезу должна составляться своя индивидуальная форма (отдельный файл), которая, при этом, дополняется

сведениями об авторах исследования и организации, которую представляют авторы, ключевыми словами для быстрого и удобного поиска. При этом система будет позволять оставлять отзывы от других ученых о достоверности того или иного исследования и оценивать рейтинг авторов и организаций. Стоит повторить, что большое значение также будут представлять формы не подтвердившихся теорий, позволяя другим исследователям не идти по ложному пути.

Бланк справочного исследования, в котором проверялась не какая-то гипотеза, а «что мы получим» при заданных параметрах, должна иметь отличительную форму, отражающую какие количественные или качественные характеристики, были получены.

При создании данной системы большую роль будет играть стимулирование пополнения базы данных уже выполненными и сохранившимися в печатном виде отчетами. При этом формулы и модели, не подтвержденные экспериментальным исследованием, не представляют интереса для системы.

Дополнение такой базы исследованиями классиков физики и механики будет нести большое образовательное значение.

В настоящее время в мире наблюдается устойчивый рост числа производимых в мире баз данных научной информации. Сейчас количество баз данных, производимых в мире, оценивается цифрой 8 тысяч, а годовой прирост числа баз данных превышает 10 %. С одной стороны, это дает возможность констатировать появление на рынке производителей новой мощной индустрии - индустрии баз данных, а с другой, обостряет проблему типологии баз данных. В то же время качественные характеристики, отличительные особенности и применимость базы данных (или группы баз данных) для решения конкретной задачи (задач) пользователя не могут быть оценены без знания типа базы данных.

Однако задача классификации уже существующих на информационном рынке (и возможных в будущем) баз данных не так проста, какой она представляется на первый взгляд. Это связано с тем, что глубоких исследований по проблемам построения некоей общепринятой типологии баз данных не существует, а сложившаяся зарубежная терминологическая практика (в приложении к типам баз данных) далеко не всегда адекватно отражает истинную типологию, и к тому же производители баз данных не всегда следуют данной практике.

Идея хранилищ данных оказалась востребованной, так как во многих видах государственной, деловой, научной, социальной деятельности необходимы тематически объединенные и исторически очищенные совокупности данных, при этом постоянно возрастала потребность:

- в более дешевых данных;
- в точных и структурированных данных;
- в большей оперативности получения и обработки данных;
- в интегрированных данных.

Данные, помещаемые в хранилище, должны отвечать определенным требованиям - предметной ориентированности, интегрированности, поддержки хронологии и неизменяемости [1, с.214-115].

Хранилище данных выполняет множество функций, но его основное предназначение - предоставление точных данных и информации в кратчайшие сроки и с минимумом затрат.

Понятие хранилище данных в первоначальном понимании было основано на понятии распределенной витрины данных (Distributed DataMart - DDM). Поэтому в классическом исполнении хранилище данных было прежде всего сквозной базой данных и информации предприятия [1, с.37-38].

Среда хранилища была предназначена только для чтения и состояла из детальных и агрегированных данных, которые полностью очищены и интегрированы; кроме того, в репозитории хранилась обширная и детальная история данных на уровне транзакций. С точки зрения архитектурного решения такое хранилище данных реализует свои функции через подмножество зависимых витрин данных

Достоинствами архитектуры классического хранилища данных являются:

- общая семантика;
- централизованная, управляемая среда;
- согласованный набор процессов извлечения и бизнес-логики использования;
- непротиворечивость содержащейся информации;
- легко создаваемые по шаблонам и наполняемые витрины данных;
- единый репозиторий метаданных;
- многообразие механизмов обработки и представления данных.

К недостаткам можно отнести большие затраты по реализации, высокую ресурсоемкость в масштабе всего предприятия, потребность в сложных сервисных системах, рискованный сценарий развития, когда все данные и метаданные находятся в одном репозитории и в неблагоприятном случае могут быть потеряны. Кроме того, при фильтрации, агрегировании и рафинировании «сырых» данных для такого хранилища обычно теряется очень много информации, которая может быть чрезвычайно полезной при бизнес - анализе.

Современные профессиональные базы данных.

 arXiv.org — крупнейший и старейший архив электронных статей и препринтов. Помимо публикаций по техническим и естественным наукам, в системе хранятся публикации по статистике, экономике и финансам.

 НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА
eLIBRARY.RU Доступ к полным текстам российских научных журналов. Доступ к РИНЦ (Российский индекс научного цитирования). Научная электронная библиотека eLIBRARY.RU предоставляет доступ к полным текстам российских научных журналов, а также Российскому индексу научного цитирования. Значительное количество материалов представлено в открытом доступе: например, все выходящие в издательстве «Наука» журналы. Московский кампус имеет подписку на коллекцию изданий.

 DIRECTORY OF OPEN ACCESS JOURNALS Directory of Open Access Journals — агрегатор информации о публикациях научных журналов открытого доступа. Платформа содержит информацию о более 5 миллионах статей из различных областей науки: права, истории, политологии, социальных наук и т.д. В системе присутствуют журналы ведущих мировых издательств: Springer, Wiley, Taylor & Francis и др.

 Dimensions Информационно-аналитическая платформа с научными публикациями, программными и аналитическими документами, информацией о грантах и патентах. Платформа Dimensions включает информацию о более 111 млн. публикаций из открытых источников (препринты, статьи из журналов с открытым доступом, главы из книг, монографии и др.), 5,5 млн. грантовых записей от 500+ фондов, более 40 млн. патентных записей, а также клинических исследованиях и отчетах ведущих научных организаций.

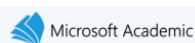
 Open
DOAR Directory of Open Access Repositories — международный агрегатор репозиторий открытого доступа. В представленных на платформе репозиториях можно найти научные статьи, книги, препринты, диссертации, массивы данных по различным дисциплинам: бизнес и экономика,

образование, политология, право и др. В системе есть и репозитории в том числе и российских научных и образовательных организаций.



Directory of open access books — европейский проект, собирающий информацию о книгах открытого доступа, выходящих по всему миру, то есть эти книги можно скачать и читать совершенно легально. На сайте можно найти более 25 000 книг самых разных научных издательств, в том числе ведущих мировых: Springer, Oxford University Press, Taylor & Francis и др. Языки публикаций: английский, французский, немецкий, испанский.

InCites InCites позволяет анализировать эффективность исследований и сопоставлять научную производительность по организациям, журналам, странами и т.п. Объектами изучения в InCites являются исследователи, организации, регионы, источники, области исследования, источники финансирования. Для того, чтобы воспользоваться инструментами InCites - необходимо пройти регистрацию.



Microsoft Academic - это бесплатная платформа для поиска академических публикаций, разработанная компанией Microsoft Research. В настоящий момент Microsoft Academic индексирует почти 250 млн публикаций разного типа: в их числе журнальные статьи, книги и главы книг, патенты, материалы конференций.



Проект «Гутенберг» — старейшая бесплатная электронная библиотека мира. Большинство загруженных в нее книг, а сейчас их больше 60 000, — это тексты открытого доступа на английском, французском, немецком и других языках. Книги на русском языке тоже есть. В библиотеке хранятся как художественная литература, так и научные и научно-популярные работы по истории, философии, праву, психологии и другим научным областям. Каждая книга из библиотеки имеет несколько форматов, в том числе есть возможность скачать и загрузить нужную литературу в электронную книгу или приложение для чтения книг.



ProQuest Dissertation & Theses Global (PQDT Global) - полнотекстовая коллекция включает 3,8 млн. диссертаций и дипломных работ из 30 стран мира, опубликованных с 1861 года, со всего мира. Возможен выбор русскоязычного интерфейса и перевод аннотаций на русский язык.

Ebook Central. Полные тексты книг зарубежных научных издательств. Доступ к полнотекстовой мультидисциплинарной коллекции избранной качественной научной литературы ведущих мировых издательств.

 Полные тексты журналов и справочников Handbooks издательства Elsevier. Полнотекстовая база научных журналов академического издательства Elsevier, а также полный доступ к справочникам Handbooks of numerical analysis, statistics, economics series, economics & finances.

 Реферативная база, являющаяся указателем научного цитирования. Позволяет в краткие сроки получить самую полную библиографию по интересующей теме. Не содержит полных текстов статей.

 Международный репозиторий научных статей и препринтов. SSRN — один из крупнейших репозиториев научных статей и препринтов. На платформе размещены более 800 тысяч полнотекстовых публикаций по более 50 предметным областям, в том числе по социальным и гуманитарным наукам.

 Реферативная база. Позволяет в краткие сроки получить самую полную библиографию по интересующей теме. Не содержит полных текстов статей. База данных научного цитирования. Содержит информацию как по зарубежным научным изданиям, так и по отечественным академическим журналам. Представляет доступ к спискам библиографических ссылок к каждой публикации.

 IEEE Xplore - Интернет библиотека с доступом к реферативным и полнотекстовым статьям и материалам конференций. Бессрочно без подписки, не все полнотексты открыты.

InspecAnalytisc – Аналитический модуль базы данных Inspec. InspecAnalytics позволяет визуализировать результаты поиска, сравнивать полученные результаты на уровне учреждений, авторов, тематик по количеству публикаций. Новые функциональные возможности позволяют пользователям получать новые идеи и выявлять тенденции и модели, которые ранее были недоступны.

Основными достоинствами использования баз данных научной информации являются:

1) более короткий срок поступления к пользователю по сравнению с журналами в печатной форме;

- 2) разнообразие доступа (на CD-ROM, и в режиме on-line);
- 3) возможность "перекачивания" полного текста статьи в память персонального компьютера пользователя;
- 4) отсутствие необходимости переплета и размещения в книгохранилище;
- 5) возможность одновременного использования одного и того же журнала несколькими читателями;
- 6) дополнительные услуги: выборка информации и навигации внутри одного документа или среди нескольких.

К важным характеристикам баз данных научной информации относятся:

- 1) Охватываемая сфера: тематика, типы источников, насколько полно представлен каждый источник, охватываемый период времени;
- 2) Структура и содержание записи, возможность обращения к отдельным полям: из каких полей состоит запись, как они организованы, по каким из них можно проводить поиск;
- 3) Размер: сколько записей в базе данных, средний размер одной записи.
- 4) Современность и обновление информации: сколько времени проходит между публикацией статьи и появлением соответствующей записи в базе данных, как часто добавляются новые записи в базу данных;
- 5) Качество: кто несет ответственность за базу данных и какой контроль ее качества предусмотрен;
- 6) Способы поиска: поиск при помощи произвольного текста или при помощи упорядоченной индексирующей системы (управляемый словарь);
- 7) Рекомендации по методике поиска, документация, оперативная справка: обычно обеспечивается в печатном или электронном виде изготовителем или поставщиком.
- 8) Стоимость: в зависимости от условий доступа.

В настоящее время для удовлетворения потребностей пользователей ресурсов к информационным системам, содержащим базы данных научной информации, предъявляются следующие требования:

- 1) Охват всех уместных информационных ресурсов. Для сбора информации должны быть созданы процедуры ввода данных. Возможные варианты ввода данных:
 - интерактивным вводом данных пользователями;

- сбором данных в сети посредством специализированных сетевых “пауков”;
- путем обмена данными с другой информационной системой;
- в ряде работ отмечается, что большинство информационных систем не смогут охватить всю информацию в одной центральной базе данных. Предполагается, что НИС кроме задач поиска информации в своей БД будут искать информацию и в других системах, а также служить порталами позволяющими выяснить, где же искать информацию, если нет в данной системе.

2) Релевантность документов. При автоматическом сборе информации в сети Интернет возможно накопление информации нерелевантной или малорелевантной для данной информационной системы. Это проблема может решаться следующими способами:

- созданием подробных форматов представления метаданных о ресурсах и сильно структурированных справочников для тематической классификации ресурсов, требованием от создателей Web - страниц четко и точно следовать форматам метаданных и вкладывать описания в метаданных на Web-страницы. Проблематично требовать от пользователей записи в их Web - страницы метаданных в каком-либо формате, это требует дополнительной работы и знакомства с форматами метаданных. Способ решить эту проблему: создание профессиональных Web - страниц или метаописаний интерактивными средствами;

- разделением всех информационных ресурсов на собранные экспертами/пользователями и “пауком”, и указанием степени достоверности информации в зависимости от ее источника;

- точным указанием поисковым средствам пространства поиска и анализа информации, а также критериев качества собранной информации;

- введением соответствующих потребностям пользователей схем классификации ресурсов и классификации экспертами ресурсов согласно этим схемам.

3) Актуальность, полнота, достоверность происхождения документов. Проблемы актуальности и полноты решаются способами аналогичными способам проблемы решения охвата документов. Проблема достоверности происхождения информации решается:

- для интерактивного ввода - ограничением ввода только аутентифицированными пользователями;
- для автоматизированных систем сбора - путем ограничения области действия, собирающего информацию “паука”;
- для ввода интеграцией с другими системами – путем установления точных фильтров на импортируемые информационные ресурсы;
- для всех систем - проверкой и классификацией введенной информации.

4) Наличие интеллектуальных служб обслуживания запросов пользователя. Службы обслуживания запросов пользователей должны поддерживать поиск по атрибутам с поддержкой полнотекстового поиска, просмотр ресурсов по категориям. В ряде проектов определено, что службы поиска должны поддерживать семантический поиск.

Кроме того, к информационным системам, работающим со многими типами информационных ресурсов, предъявляются следующие требования:

5) Поддержка не централизованных архитектур информационной системы. Это требование является необходимым условием для полноты, аутентичности и актуальности информации. Опыт эксплуатации информационных систем показал, что трудно реализуемо, во многих случаях даже невозможно, создание централизованных научных систем, которые охватывают научную информацию в какой-то области науки, или в какой-то стране. Опыт создания таких систем завершился удачно лишь в Дании и Исландии;

6) Структурированность информационного пространства. Для поддержки сложных функций поиска, классификации информации недостаточно хранить только полнотекстовые описания;

7) Предоставление информации пользователю в виде, выбранном пользователем;

8) Историчность информации. Научная информация специфична достаточно коротким временем жизни и актуальности. Для многих типов информационных ресурсов важно хранить описание жизненного цикла этих ресурсов и иметь возможность восстановить состояние ресурса на любой момент времени.

9) Поддержка различных уровней абстракции для представления информации.

10) Архив. Выше было отмечено, что большая часть научной информации быстро устаревает. Но существуют информационные ресурсы, которые могут быть доступны длительное время. К таковым, например, относятся документы, имеющие длительную юридическую силу, патенты или мультимедийная информация об исторических событиях, которая может быть востребована через любой период времени. Кроме того, научные отчеты институтов, речи ученых могут также иметь огромную историческую ценность, становясь только еще ценнее со временем. Поэтому системы должны поддерживать возможность длительного хранения информационных ресурсов с возможностью восстановления их.

Список литературы

1. Кириллов, В.В. Основы проектирования реляционных баз данных [Электронный ресурс] / В.В. Кириллов – М.: Центр информационных технологий - Режим доступа: <http://www.citforum.ru/database/dbguide/index.shtml>. (дата обращения: 10.01.2021).
2. Основы научных исследований : учеб. пособие / Ф. В. Гречников, В. Р. Каргин. – Самара : Изд-во СГАУ, 2015. – 111 с.
3. Интернет-технологии в деятельности издательств [Электронный ресурс]. – Режим доступа: <http://www.hi-edu.ru/e-books/xbook737/01/part002.htm/> (дата обращения: 10.01.2021).
4. Гасанов Р.Ф., Голубев С.С. и др. Методология организации научных исследований. М: ЦНИИ «Центр», 2017. - 118 с. Режим доступа: <https://elibrary.ru/item.asp?id=30011299>.
5. Теория и методология научного исследования: монография / В.А. Песоцкий и др.; под ред. В.А. Песоцкого. – М.: ИИУ МГОУ, 2017. – 200 с. Режим доступа: <https://elibrary.ru/item.asp?id=32746409>.