**Базы данных научной информации**

*База данных* - набор записей, который состоит из нескольких полей и для каталога неструктурированных источников информации является идеальной структурой. Прежде чем использовать определенную базу данных научной информации, конечно же, необходимо знать её организацию.

В данной работе под *базой данных* будем понимать «упорядоченную совокупность информационных соединений (библиографических описаний опубликованных/неопубликованных источников, полных текстов и т.д.), относящихся к определенной тематике и представленных на машинных (машиночитаемых) носителях: диски, магнитные ленты, компактные оптические (магнитооптические) диски, дискеты, в виде организованного файла или группы файлов.



*К основным характеристикам баз данных научной информации относятся:*

**1. Охватываемая сфера:** тема, типы источников, полнота источников, временный период;

**2. Структура и содержание записи, возможность обращения к отдельным полям:** из каких полей состоит запись; как организованы; по каким можно проводить поиск;

**3. Размер:** количество записей в базе данных, средний размер одной записи.

**4. Современность и обновление информации:** промежуток времени между публикацией статьи и появлением соответствующей записи в базе данных, частота добавления новых записей в базу данных;

**5. Качество:** ответственный за базу данных и контроль её качества;

**6. Способы поиска:** произвольный текст или упорядоченная индексирующая система (управляемый словарь);

**7. Рекомендации по методике поиска, документация, оперативная справка:** в электронном/печатном виде (изготовитель или поставщик).

**8. Стоимость:** зависит от условий доступа.

Для удовлетворения потребностей пользователей ресурсов к информационным системам, а также содержащим базы данных научной информации в настоящее время предъявляются следующие *требования:*

**1. Охват всех уместных информационных ресурсов.** Сбор информации должен осуществляться при помощи ввода данных (интерактивный ввод (пользователи), путём специализированных сетевых «пауков», осуществляя обмен данных с другой информационной системой,

**2. Релевантность документов.** Глобальная сеть Интернет предоставляет возможность накопления информации (для данной информационной системы) - нерелевантной или малорелевантной (при автоматическом сборе информации).

*Способы решения данной проблемы:* используя интерактивные средства создавать профессиональные Web-страницы или метаописания; разделение информационных ресурсов на собранные - экспертами/пользователями и «пауком», указание степени достоверности информации (в зависимости от её источника); указание поисковым средствам пространства поиска и анализа информации, критериев качественности собранной информации; введение экспертами и пользователями схем классификации ресурсов.

**3. Актуальность, полнота, достоверность происхождения документов** (интерактивный ввод (аутентифицированные пользователи), автоматизированная система сбора (ограничение области действия, собирающего информацию «паука»), ввод интеграций с другими системами (установление точных фильтров на импортируемые информационные ресурсы), для всех систем (проверка и классификация введенной информации).

**4. Наличие интеллектуальных служб обслуживания запросов пользователя** (поиск по атрибутам с поддержкой полнотекстового поиска, просмотр ресурсов по категориям, семантический поиск).

**5. Поддержка не централизованных архитектур информационной системы.** Данное требование необходимо для раскрытия полноты, актуальности и аутентичности информации.

**6. Структурированность информационного пространства.** Недостаточно хранить только полнотекстовые описания для поддержки сложных функций поиска и классификации информации;

**7. Предоставление информации пользователю в том виде, который выбрал пользователь;**

**8. Историчность информации** (сохранение описания жизненного цикла информационных ресурсов, а также возможность восстановить состояние ресурса на любой момент времени).

**9. Поддержка различных уровней абстракции для представления информации (**позволяет ускорять поиск информации пользователем без потерь в качестве)

**10. Архив** (хранения информационных ресурсов длительное время с возможностью их восстановления)



На сегодняшний день в мире отмечается достаточно устойчивый рост числа производимых в мире баз данных научной информации. Их количество составляет 8 тысяч, а годовой прирост - превышает 10 %. С одной это говорит о появлении на рынке производителей новой мощной индустрии - индустрии баз данных. Однако, с другой стороны, обостряет проблему типологии баз данных. Ведь без знания типа базы данных качественные характеристики, отличительные особенности и применимость базы данных (или группы баз данных) для решения конкретной задачи (задач) пользователя не могут быть оценены.

Классификация баз данных уже существующих на информационном рынке имеет сложности, связанные с отсутствием исследований по проблемам построения типологии баз данных. А также зарубежная терминологическая практика далеко не всегда адекватно отражает истинную типологию, и к тому же производители баз данных не всегда следуют данной практике.

Рассмотрим классификацию баз данных научной информации в зависимости от типа хранимой в них информации:

**1. Цифровая (числовая)** – база данных со структурированной информацией числового характера.

**2. Тексточисловая** – база, в которой помимо числовой информации содержится и текстовая.

**3. Полнотекстовая** – база, содержащая полные тексты информационных источников (например, полные тексты законодательных актов, патентов, монографий и т.д.).

**4. Реферативно-библиографическая (библиографическая)** – база данных, содержащая описание публикуемых (статьи, книги, тезисы, доклады) и непубликуемых (диссертации, патенты) документов, которым обычно дается полное библиографическое описание (название, автор, том, источник, выпуск, год издания, страницы), набор ключевых слов и рубрикационных шифров, раскрывающих тему и содержание документа, а также приводятся рефераты или аннотации работ.

**5. Адресно-справочная (адресная или справочная)**  - база, содержащая адресную и справочную информацию.

**6. Программная** – база данных (или база программ) являются базы, содержащие программы для ЭВМ в различных видах (программы статистической обработки данных).

**7. Лексикографическая** - база, содержащие различные машиночитаемые словарные массивы (объект описания - лексические единицы).

Все типы баз данных условно можно разделить на 2 вида: первичные (числовые, тексточисловые, полнотекстовые базы данных и базы по физико-химическим свойствам) – содержат результаты научных разработок и исследований, и вторичные (библиографические, адресные, справочные) - содержат результаты логической и аналитико-синтетической переработки одного/нескольких первичных документов/сведения о них.

Если же подразделять по классам, то можно выделить 2 класса баз: документальные (полнотекстовые, библиографические, справочные) и фактографические (тексточисловые и справочные базы, числовые).

Для поиска пользователем баз данных научной информации существует специальный режим (расширенный или expand-поиск). Он дает возможность пользователю обратиться к общему словарному файлу, задав один/несколько связанных основополагающих термина и получив в ответ перечень баз данных с указанием количества отвечающих заданным терминам документов (записей) в каждой базе.

Затем пользователь может осуществить выбор наиболее "представительных" баз данных и искать уже в них, используя все штатные средства поиска и вывода найденной информации. Осуществлять поиск пользователь может, не переключаясь с базы на базу последовательно, а сразу в выбранном тематическом подмножестве баз данных (кластере).

Для этого пользователю предоставляются следующие механизмы поиска: **навигация** во всем доступном информационном пространстве (наглядное предоставление логической структуры информационного пространства), **лексический поиск** (информационный поиск по свободной лексике русского языка и языков, использующих латинский алфавит), **символьный поиск** (в качестве лексической единицы выступает определенная последовательность допустимых символов), **атрибутный поиск** (поиск объектов по значениям их характеристик - "автор", "название", "место издания" и т.п.).

Результат поиска - совокупность информационных объектов, значения характеристик которых удовлетворяют условиям поискового предписания. Результаты поиска, как и другие списки, формируемые в системе, должны допускать сортировку по значениям любого поля, определенного как ключевое.

Просмотр содержания информационного объекта и его структуры: последовательный (например, страница за страницей) и выборочный (переход на любую заданную страницу или на любой элемент, отраженный в структуре). Текст и структура произведения должны быть синхронизированы (любое изменение положения в структуре вызывает соответствующее изменение положения в тексте, и наоборот).

Многооконный режим работы позволяет пользователю видеть на экране одновременно два окна (как минимум) с разным содержанием.

 Одновременный вывод на экран (в разных окнах) текстов двух разных произведений или двух экземпляров одного произведения.

Поддержка аппарата гипермедийных и гипертекстовых связей обеспечивает пользователю оперативный переход от объекта/некоторого его элемента к другому взаимосвязанному с ним объекту/его элементу. При этом допускается существование связей типа "один-к-одному" и "один-ко-многим".

Протоколирование сеанса работы пользователя с системой даёт возможность перехода в любое из ранее существовавших состояний системы.

Установка закладок в тексте информационных объектов и возможность оперативного перехода на них.

Экспорт информации из системы с указанием на источник (собственно ЭИ или ЭБ и информационный объект, из которого копируется информация).

Однако, на практике зачастую бывает так, что для обеспечения максимальной полноты информации (особенно при многоаспектном поиске) ресурсов одной службы недостаточно, и для получения даже относительно небольшой доли дополнительной информации, необходимо привлечение ресурсов. В связи с этим возникают сложности с освоением не одной, а нескольких информационно-поисковых систем, переформулированием запросов и т.п. Поэтому на информационном рынке появились интеллектуальные шлюзы и универсальные информационные, представляющие собой специальные программные системы. Информационные шлюзы облегчают доступ пользователя к ресурсам (базам данных) различных информационных служб и не предусматривают развитых средств помощи пользователю в выборе стратегии поиска и оценке результатов поиска. В интеллектуальных же шлюзах реализуются задачи формирования поисковых предписаний, алгоритмы выбора баз данных, а также ввод запросов на соответствующем языке, для чего используются методы искусственного интеллекта и элементы экспертных систем. Различные шлюзовые службы могут охватывать ресурсы двух и более центров обработки данных, а службы Easy Net, Easylink, Searchline и InfoMaster являются службами с мировой сферой деятельности.