**Модели процессов накопления информации. Основные принципы поиска информации. Информационно-поисковые системы.**

Хранение и накопление являются одними из основных действий, осуществляемых над информацией и главным средством обеспечения ее доступности в течение некоторого промежутка времени. В настоящее время определяющим направлением реализации этой операции является концепция базы данных, склада (хранилища) данных.

База данных может быть определена как совокупность взаимосвязанных данных, используемых несколькими пользователями и хранящихся с регулируемой избыточностью. Хранимые данные не зависят от программ пользователей, для модификации и внесения изменений применяется общий управляющий метод.

**Банк данных** — система, представляющая определенные услуги по хранению и поиску данных определенной группе пользователей по определенной тематике.

**Система баз данных** — совокупность управляющей системы, прикладного программного обеспечения, базы данных, операционной системы и технических средств, обеспечивающих информационное обслуживание пользователей.

**Хранилище данных** — это база, хранящая данные, агрегированные по многим измерениям. Основные отличия ХД от БД: агрегирование данных; данные из ХД никогда не удаляются; пополнение ХД происходит на периодической основе; формирование новых агрегатов данных, зависящих от старых — автоматическое; доступ к ХД осуществляется на основе многомерного куба или гиперкуба.

**ИПС** (информационно-поисковая система) - это система, обеспечивающая поиск и отбор необходимых данных в специальной базе с описаниями источников информации (индексе) на основе информационно-поискового языка и соответствующих правил поиска.

**Главной задачей** любой ИПС является поиск информации релевантной информационным потребностям пользователя. Очень важно в результате проведенного поиска ничего не потерять, то есть найти все документы, относящиеся к запросу, и не найти ничего лишнего. Поэтому вводится качественная характеристика процедуры поиска - релевантность.

**Релевантность** - это соответствие результатов поиска сформулированному запросу.

Основными показателями ИПС для WWW являются пространственный масштаб и специализация. По пространственному масштабу ИПС можно разделить на локальные, глобальные, региональные и специализированные. Локальные поисковые системы могут быть разработаны для быстрого поиска страниц в масштабе отдельного сервера. Региональные ИПС описывают информационные ресурсы определенного региона, например, русскоязычные страницы в Интернете. Глобальные поисковые системы в отличие от локальных стремятся объять необъятное - по возможности наиболее полно описать ресурсы всего информационного пространства сети Интернет.

В общем случае, можно выделить следующие поисковые инструменты для WWW: каталоги, поисковые системы, метапоисковые системы.

**1. Каталог**

**Каталог** - поисковая система с классифицированным по темам списком аннотаций со ссылками на web-ресурсы. Классификация, как правило, проводится людьми.

Поиск в каталоге очень удобен и проводится посредством последовательного уточнения тем. Тем не менее, каталоги поддерживают возможность быстрого поиска определенной категории или страницы по ключевым словам с помощью локальной поисковой машины. База данных ссылок (индекс) каталога обычно имеет ограниченный объем, заполняется вручную персоналом каталога. Некоторые каталоги используют автоматическое обновление индекса.

Результат поиска в каталоге представляется в виде списка, состоящего из краткого описания (аннотации) документов с гипертекстовой ссылкой на первоисточник.

Адреса популярных каталогов: Yahoo, Magellan, Weblist, Улитка

**2. Поисковая машина**

**Поисковая машина** - поисковая система с формируемой роботом базой данных, содержащей информацию об информационных ресурсах.

Отличительной чертой поисковых машин является тот факт, что база данных, содержащая информацию об Web-страницах, статьях Usenet и т.д., формируется программой-роботом.

Поиск в такой системе проводится по запросу, составляемому пользователем, состоящему из набора ключевых слов или фразы, заключенной в кавычки. Индекс формируется и поддерживается в актуальном состоянии роботами-индексировщиками.

В описании документа чаще всего содержится несколько первых предложений или выдержки из текста документа с выделением ключевых слов. Как правило, указана дата обновления (проверки) документа, его размер в килобайтах, некоторые системы определяют язык документа и его кодировку (для русскоязычных документов).

**3. Метапоисковая машина**

**Метапоисковые системы** (поисковые службы) - системы, способные послать запросы пользователя одновременно нескольким поисковым серверам, затем объединить полученные результаты и представить их пользователю в виде документа со ссылками.

Адреса известных метапоисковых систем: MetaCrawler, SavvySearch