

Особенности построения информационно-поисковых систем

**Информационно-поисковая система -
совокупность информационно поискового
правил перевода с естественного языка на
информационно - поисковый и обратного
перевода, а также критерия соответствия,
предназначенная для осуществления
информационного поиска.**

Два основных способа поиска документа:

- использование иерархической модели
- использование гипертекстовой модели

Типовая схема ИПС



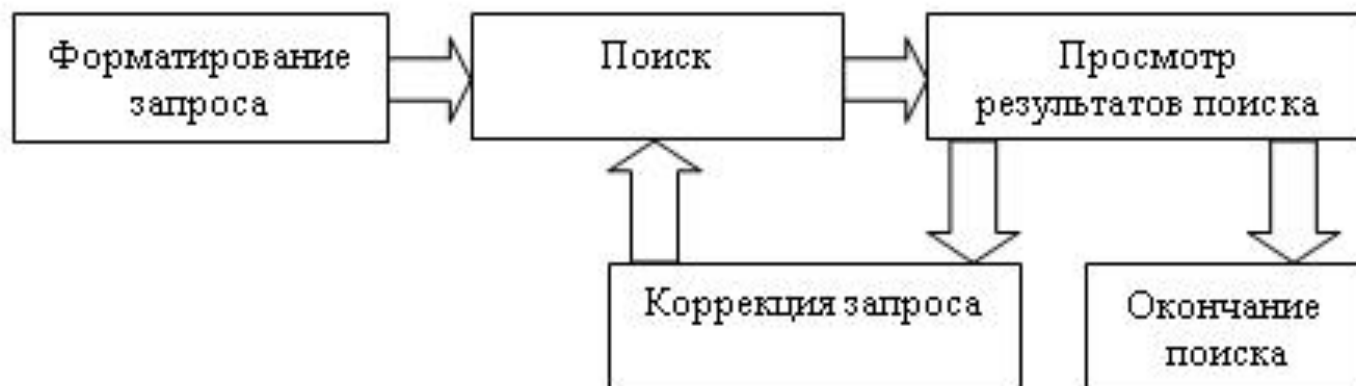
Информационный поиск –
процесс отыскания в некотором
множестве текстов (документов)
всех таких, которые посвящены
указанной в запросе теме (предмету)
или содержат нужные потребителю
факты, сведения.

Информационно-поисковый язык –
знаковая система, предназначенная для
описания (путём индексирования)
основного смыслового содержания
текстов (документов) или их частей, а
также для выражения смыслового
содержания информационных запросов с
целью реализации информационного
поиска.

В состав информационно-поискового языка входят:

1. Словарь индексационных терминов - множество терминов индексирования.
2. Кодовый словарь - множество кодовых терминов.
3. Словарь входов - множество входных терминов.
4. Вспомогательные средства языка индексирования - средства, используемые совместно с индексационными терминами для расширения или сужения определенных понятий.
5. Правила использования языка индексирования.

Процедура поиска

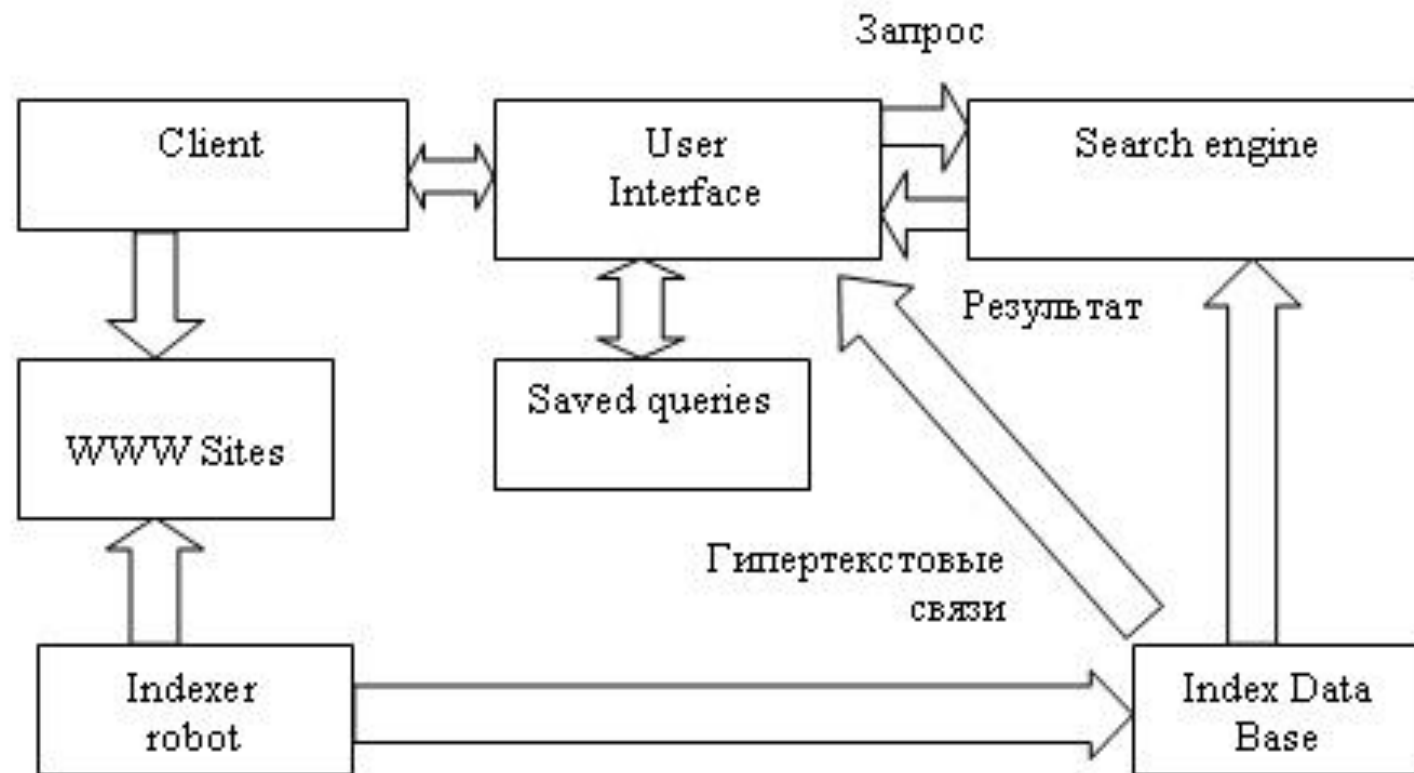


По форме диалога, способу задания условия отбора и механизму поиска ИПС можно разделить на два класса:

системы рубрикационного типа; (реализуются интерфейсом в виде иерархических последовательно раскрывающихся списков, через которые обеспечивается доступ к тематически связанным группам документов. Раскрывая очередную рубрику и перемещаясь, таким образом, по тематической иерархии, пользователь уточняет предметную область и увеличивает (усредненно) степень точности соответствия выдаваемых документов и информационной потребности.)

структурно-логические системы. (используются для работы с базами данных структурированной информации, когда каждый документ состоит из многих информационных полей, возможно, разного типа. Критерий отбора в этом случае строится как логическая комбинация простых, сводящихся к проверке условия присутствия или отсутствия в документе, слов (имен собственных или имен понятий, определяющих предмет поиска).)

ИПС для Интернета



Client (клиент) - это программа просмотра конкретного информационного ресурса. Такая программа обеспечивает просмотр документов WWW, Gopher, Wais, FTP-архивов, почтовых списков рассылки и групп новостей Usenet.

User interface (пользовательский интерфейс) - способ общения пользователя с поисковым аппаратом: системой формирования запросов и просмотра результатов поиска.

Search engine (поисковая машина) - служит для трансляции запроса на информационно-поисковом языке, в формальный запрос системы, поиска ссылок на информационные ресурсы Сети и выдачи результатов этого поиска пользователю.

Index database (индекс базы данных) - индекс, который является основным массивом данных ИПС и служит для поиска адреса информационного ресурса.

Queries (запросы пользователя) - сохраняются в его (пользователя) личной базе данных.

Index robot (робот индексирования) - служит для просмотра данных в Интернете и поддержания базы данных индекса в актуальном состоянии. Эта программа является основным источником информации о состоянии информационных ресурсов сети.

WWW sites - это весь Интернет, или, точнее, информационные ресурсы, просмотр которых обеспечивается программами просмотра.

Источником информации о состоянии информационных ресурсов сети является робот-индексировщик - это программа, которая по определенному алгоритму "заходит" на различные страницы, "читает" их и индексирует.

Банк данных. Основные понятия

$$\text{БнД} = \text{БД} + \text{СУБД}$$

$$\text{БнД} = \text{АС}$$

Банк данных (БнД) – это АИС, включающая в свой состав комплекс специальных методов и средств для поддержания динамической информационной модели предметной области с целью обеспечения обработки информационных запросов пользователя.

Требования, предъявляемые к БНД:

1. Удовлетворять актуальным информационным потребностям внешних пользователей, обеспечивать возможность хранения и модификации больших объемов многоаспектной информации
2. Обеспечивать заданный уровень достоверности хранимой информации
3. Обеспечивать доступ к данным только пользователям с соответствующими полномочиями
4. Обеспечивать возможность поиска информации по произвольной группе признаков
5. Удовлетворять заданным требованиям по производительности при обработке запросов
6. Иметь возможность реорганизации и расширения при изменении границ ПО
7. Обеспечивать выдачу информации пользователю в различной форме
8. Обеспечивать возможность одновременного обслуживания большого числа внешних пользователей.

Преимущества централизации управления данными:

1. Сокращение избыточности хранимых данных
2. Устранение противоречивости хранимых данных
3. Многоаспектное использование данных
4. Комплексная оптимизация
5. Обеспечение возможности стандартизации
6. Обеспечение возможности санкционированного доступа

БнД

Вычислительные
системы

↓
ОС

↓
Технические
средства

БД

↓
Словарь
данных

СУБД

↓
Администратор банка
данных

↓
Обслуживающий
персонал

