Лекция №4

ГИПЕРТЕКСТОВАЯ ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ (ГИТ)

ГИТ

Текст является универсальным средством представления, накопления и передачи знаний в человеческом обществе.

Гипертекст (ГТ) – одна из фундаментальных моделей представления знаний, выраженных в текстовом виде.

Обычный (одномерный) текст рассматривается как длинная строка символов, читаемая в одном направлении.

Многомерный текст (ГТ) включает **точки ветвления**, в которых чтение можно продолжать в нескольких направлениях в зависимости от информационных потребностей читателя.

Современные гипертекстовые системы позволяют пользователю самостоятельно формировать альтернативные траектории навигации по ГТ, максимально отвечающие его текущим интересам.

ОСНОВНЫЕ ПОНЯТИЯ ГИТ

В основе ГТ лежат следующие основные идеи.

1. Текст разбивается на фрагменты, представляющие его **семантические единицы (сеты)**. Между ними устанавливаются связи, которые могут наделяться именами.

2. В отличие от обычного текста, который читается последовательно, ГТ можно читать, **двигаясь по разным траекториям**, образованным связанными сетами.

3. Активируемые переходы выбираются читателем (пользователем). Имена (типы) связей облегчают решение задачи выбора перехода.

ГТ документ может быть как **электронным**, так и **бумажным**. Однако в полной мере функциональность ГТ реализуется лишь в электронных гипертекстовых документах.

В ГТ документе может быть представлено **несколько уровней детализации материала**. Такие документы моделируются деревьями или сетями.

ОСНОВНЫЕ ПОНЯТИЯ ГИТ

Таким образом, ГТ как информационная модель интегрирует положительные стороны **энциклопедий**, **монографий** и **тезаурусов**.

От **энциклопедий** ГТ наследует:

• детальное представление понятий;

• быстрый просмотр материала;

• алфавитный поиск.

От **монографий**:

• представление материала с разной степенью глубины и детальности;

• поиск по оглавлению.

От **тезаурусов** раскрытие объема и содержания понятий, а также связей между понятиями.

Проблемы и задачи, связанные с ГИТ

Гипертекстовая информационная технология (ГИТ) — технология обработки семантической информации, основанная на использовании ГТ. Она относится к проблематике ИИ, так как ее содержанием является представление, поиск и обработка семантической информации, выраженной в текстах.

Проблемы и задачи, связанные с ГИТ:

- модели ГТ (формализованная и условно-типовая);
- инструментальные средства для создания ГТ;
- гипертекстовые информационно-поисковые системы (ГИПС);
- методы извлечения знаний для гипертекстовых систем;
- автоматизация построения ГТ;
- место ГИТ среди технологий ИИ.

Области применения ГИТ

- информационные ресурсы и технологии Internet;
- гипертекстовые информационно-поисковые системы;
- гипертекстовые информационные модели экономических систем;
- базы данных с гипертекстовой организацией;
- представление электронной документации (в том числе, контекстно-зависимой и ситуативно-зависимой справки по программным средствам);
- электронные записные книжки;
- электронные картотеки, словари, энциклопедии, справочники;
- обучающие системы;
- экспертные системы;
- организация пользовательского интерфейса и др.

ФОРМАЛИЗОВАННАЯ МОДЕЛЬ ГИПЕРТЕКСТА

В основе моделей ГТ лежит понятие **информационно-справочной статьи (ИСС)**, выступающей в качестве информационной единицы ГТ. В конкретных технологиях **ИСС называют** по-разному: **страница**, **статья**, **тема**.

Элементам ИСС могут быть присвоены метки, уникальные в рамках ИСС. Кроме того, эти элементы могут наделяться интерактивным поведением. Такие элементы называются **гиперссылками**, которые могут быть локальными и глобальными.

С точки зрения программной реализации формализованная модель ГТ состоит из двух слоев.

Первый слой представляет отображаемое на экране содержимое документа, а адреса переходов хранятся во втором, скрытом слое модели.

ФОРМАЛИЗОВАННАЯ МОДЕЛЬ ГИПЕРТЕКСТА

$$\text{ФМГТ} = (x_0, x_1, \ldots, x_{11}), \qquad (1)$$

где x_0 — имя ИСС;

x_1 — заголовок ИСС;

x_2 — аннотация ИСС;

x_3 — точка входа в ИСС;

x_4 — множество текстовых фрагментов, входящих в ИСС;

x_5 — множество цифровых информационных объектов, входящих в ИСС (изображения, видео и т.д.);

x_6 — множество программных объектов, входящих в ИСС;

x_7 — справка по ИСС;

x_8 — признак ускоренного просмотра ИСС;

x_9 — признак детального просмотра ИСС;

x_{10} — список гиперссылок внутри ИСС;

x_{11} — список гиперссылок между ИСС.

ФОРМАЛИЗОВАННАЯ МОДЕЛЬ ГИПЕРТЕКСТА

В ИСС обязательными являются точка входа, имя, заголовок и аннотация. Остальные элементы являются необязательными.

Имя служит формальным идентификатором ИСС и используется для ее адресации программными средствами. В рамках ГТ все ИСС должны иметь уникальные имена.

Заголовок представляет содержательное название ИСС.

Если на ИСС не указывают гиперссылки из других ИСС, то она становится **главной темой** и включается в **список главных тем ГТ**.

Если ИСС не имеет исходящих внешних ссылок, то на текущий момент времени эта ИСС заканчивает один или множество путей навигации по ГТ.

Деление основных элементов содержимого ИСС на три группы (x_4, x_5, x_6) обусловлено удобствами программной реализации гипертекстовых редакторов и скрыто от пользователей.

ФОРМАЛИЗОВАННАЯ МОДЕЛЬ ГИПЕРТЕКСТА

Ускоренный просмотр помогает пользователю оперативно ознакомиться с ИСС. Часто линию ускоренного просмотра ИСС образуют элементы x_1 и x_2.

Активация признака детального просмотра обеспечивает представление всего содержимого ИСС.

В данном режиме пользователь может пройти по любому пути, включающему элементы x_4, x_5, x_6 и x_7.

Поскольку объем ИСС в принципе не ограничивается, предусмотрена справка x_7, которая представляет дополнительную информацию, связанную с содержанием ИСС.

Элементы x_7, x_8, x_9, x_{10} и x_{11} реализуются через интерактивные компоненты пользовательского интерфейса, обеспечивающие навигацию по ГТ.

УСЛОВНО-ТИПОВАЯ МОДЕЛЬ ГИПЕРТЕКСТА

Один из недостатков ФМГТ связан с отсутствием в ней возможности явного определения типов гиперссылок. **В УТМГТ все гиперссылки имеют явно указанный тип.**

Данная модель ГТ включает **тезаурус, список главных тем и совокупность указателей**. Обязательным компонентом является **тезаурус ПрО**, к которой относится информационная система.

1) **Тезаурус** — упорядоченный перечень терминов, в котором отражены семантические отношения между ними.

2) **Тезаурус** — автоматизированный словарь, отображающий семантические отношения между лексическими единицами дескрипторного информационно-поискового языка и предназначенный для поиска слов по их смысловому содержанию.

Каждый термин в тезаурусе снабжается его текстовой характеристикой (статьей). Тезаурус позволяет пользователю ГТ уточнять как содержание (смысл), так и объем интересующего его термина.

УСЛОВНО-ТИПОВАЯ МОДЕЛЬ ГИПЕРТЕКСТА

Для упрощения работы с ГТ, а также повышения эффективности поиска по нему в УТМГТ включаются список главных тем и указатели.

Список главных тем делит ГТ на сегменты, соответствующие более или менее независимым частям (срезам или аспектам) ПрО. Таким образом, он отражает самое общее представление о тематике ГТ.

Указателем называется упорядоченная установленным образом последовательность информационных объектов (понятий, выражений, обозначений и т.п.), ссылающихся на ИСС, в которых эти объекты упоминаются.

В лингвистике выделено около **200 семантических типов отношений**.

УСЛОВНО-ТИПОВАЯ МОДЕЛЬ ГИПЕРТЕКСТА

Наиболее часто употребляются **10 типов**, используемых в УТМГТ и приведенных в таблице.

Часто употребляемые типы семантических отношений

Тип связи	Обозначение
Синоним	СН
род—вид	РВ
вид—род	ВР
Часть—целое (укрупнение)	ЧЦ
Целое—часть (декомпозиция)	ЦЧ
процесс—надпроцесс	ПН
процесс—подпроцесс	ПП
причина—следствие	ПС
следствие—причина	СП
Ассоциация	АС

Графовой интерпретацией УТМГТ является **семантическая сеть**. В рамках УТМГТ ИСС включает имя, заголовок, собственно текст (содержимое) и список ссылок на ИСС, связанные с данной ИСС различными типами отношений. Такой список ссылок образует **локальный справочный аппарат ИСС**.

Он может быть организован тремя способами: **в виде списка**; **внедрение ссылок в текст**; **комбинированным**.

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ДЛЯ СОЗДАНИЯ ГИПЕРТЕКСТА

Существует большое число инструментальных средств для создания ГТ и различных форматов, включая PDF (Portable Document Format), RTF (Rich Text Format), DOC (Document Word) и WinHelp (Windows Help), CHM (Compiled HTML), а также целое семейство языков гипертекстовой разметки, самыми популярными из которых можно считать HTML (Hypertext Markup Language) и XML (eXtensible Markup Language). Благодаря широкому использованию ГТ в ИС практически любой инструментарий разработки ИС включает функции для построения ГТ. В частности, данные функции реализуются в средствах разработки электронной документации (например, Adobe Acrobat), авторских системах, редакторах презентаций, издательских системах, редакторах web-страниц и др.

Создание гипертекстового справочника

Создание гипертекстового справочника по программному продукту состоит из следующих этапов:

1. Определение структуры справочника и его разделов;
2. Подготовка текста и графических иллюстраций справочника;
3. Создание файла проекта справочника;
4. Компиляция исходных файлов тем (*topics*, ИСС), графических файлов и файла проекта с формированием файла справочника;
5. Программная реализация модуля приложения, обеспечивающего доступ к справочнику;
6. Тестирование и отладка справочника.

Создание гипертекстового справочника

Первый этап является трудно формализуемым и сложным. В рамках него специфицируются:

- назначение продукта, для которого создается справочник;
- категории пользователей продукта;
- рыночный сектор, на который ориентирован продукт;
- функции и характеристики продукта, представляемые в справочнике;
- основные разделы справочника и их примерное содержание;
- соглашения, фиксирующие стиль, дизайн и оформление справочника.

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ДЛЯ СОЗДАНИЯ ГИПЕРТЕКСТА

WinHelp и ***HTML Help*** представляют собой стандартные технологии построения и работы с гипертекстовыми справочниками для платформы ***Windows***.

Гипертекст в формате *WinHelp* реализуется в виде файла с расширением ***HLP*** (***help***-файла). *HLP*-файл формируется на основе файлов с текстом в формате *RTF* с помощью специального компилятора.

Для вызова справочника из приложения служит функция *Windows API WinHelp*().

Гипертекст в формате *HTML Help* реализуется в виде файла с расширением *CHM*. Представление и взаимодействие со справочником обеспечивают программные компоненты браузера *Internet Explorer*.

CHM-файл формируется на основе файлов в формате *HTML* с помощью специального компилятора.

Для вызова справочника из приложения служит функция *HTML Help API HtmlHelp*().

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ДЛЯ СОЗДАНИЯ ГИПЕРТЕКСТА

Гипертекст может быть разработан с помощью различных инструментальных средств. Наиболее популярными из них являются:

- ***HTML Help Workshop*** фирмы ***Microsoft***;
- ***HelpScribble***;
- ***KeyTools*** фирмы ***KeyWorks Software***;
- Система ***AnetHelp Tool*** российской фирмы ***Anet Soft***;
- ***RoboHelp***;
- Подключаемый модуль ***Mif2GO***;
- Система ***Help & Manual***.

ГИПЕРТЕКСТОВЫЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ (ГИПС)

ГИПС

ГИТ используется при организации больших массивов текстовых документов и реализации методов поиска информации в них.

Информационный поиск — совокупность операций, методов и процедур, направленных на отбор данных, хранящихся в ИС и соответствующих заданным условиям.

Информационно-поисковые системы (ИПС) подразделяются на три класса:

- **документальные;**
- **фактографические;**
- **гипертекстовые (ГИПС).**

ГИПС

Признаки документа, отражающие его содержание в ИПС, называют **поисковым образом**, а признаки запроса к ИПС — **поисковым предписанием**.

Процедура перевода документа и запроса в форму представления, принятую в ИПС, называется **индексированием**.

При сопоставлении поискового образа и поискового предписания используется тот или иной **критерий смыслового соответствия (релевантности)**.

Документальный поиск

Документальный поиск относится к числу сложных информационных процессов, поскольку он связан с проблемой оценивания смыслового соответствия документа и запроса.

Из-за субъективности и неоднозначности подобного оценивания этот вид поиска в принципе не может быть исчерпывающе точным и полным, в нем всегда будет присутствовать элемент нечеткости.

Развитием данного поиска является **полнотекстовый поиск**, реализуемый, например, в поисковых машинах Internet.

Фактографический поиск

В фактографических ИПС хранятся не документы, а собственно сведения (факты) об объектах ПрО.

Подобные ИПС реализуются, в частности, на основе реляционных БД.

С точки зрения обеспечения релевантности результатов поиска (выборки данных) запросу фактографический поиск в отличие от документального является точным и полным.

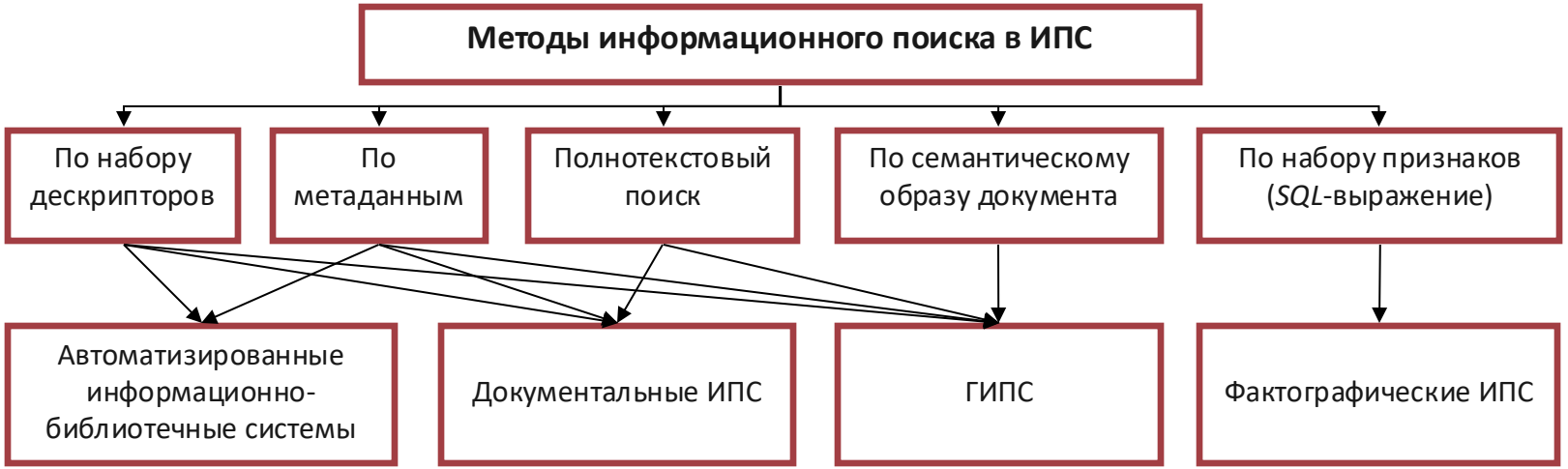
Гипертекстовый поиск

В гипертекстовых ИПС кроме содержимого документов отражается их семантическая структура.

Поэтому по глубине формализации ГИПС занимают промежуточное положение между документальными и фактографическими ИПС.

Главное различие между традиционными и гипертекстовыми ИПС заключается в том, что традиционные ИПС обычно формируются на основе структурированных данных, в то время как в ГИПС может быть представлена слабо формализованная совокупность текстов, иллюстраций, аудио и видео-документов и т.д.

Классификация методов информационного поиска в ИПС



Релевантность

Введем следующие обозначения: D — множество документов в информационном хранилище, $d_i \in D$ — i-й документ, $D_j \subseteq D$ — подмножество документов.

Зададим на D оценку смысловой близости пары документов $r(d_i, d_j) \geq 0$.

При $r = 0$ **документы** d_i **и** d_j **эквивалентны по смыслу**. Для **семантически несопоставимых документов** r **не определена**.

Также введем оценки ряда важных свойств документов: $S=(S_1, S_2, \ldots, S_k)$, $k >0$. Чем больше значение оценки, тем важнее для пользователя документ.

Поисковый запрос может рассматриваться как виртуальный документ z.
В идеальном случае ($r(z, d_i) = 0$) ему точно соответствует документ d_i.

Виды поиска

Используя введенные обозначения, определим следующие **виды поиска**:

1. Найти $(D_j \subseteq D) \mid r(z, d_i \in D_j) \rightarrow min$. Если $D_j = \varnothing$, то в D нет документов, релевантных запросу. При $|D_j| = 1$ есть единственный подходящий документ. Если же $|D_j| > 1$, то таких документов несколько;
2. Найти $(D_j \subseteq D) \mid r(z, d_i \in D_j) \leq \Delta$, где Δ — оценка наибольшего допустимого расхождения смыслов запроса и искомых документов;
3. Найти $(D_j \subseteq D) \mid S_f(d_i \in D_j) \rightarrow max$. Результатом поиска служит подмножество документов, которым приписана наибольшая оценка важности. Обобщением этого варианта является векторный поиск, учитывающий оценки нескольких свойств;
4. Комбинированный поиск: Найти $(D_j \subseteq D) \mid r(z, d_i \in D_j) \leq \Delta \;\&\; S_k(d_i \in D_j) \rightarrow max$.

Эффективность информационного поиска

Интеллектуальные возможности ИПС в части функций информационного поиска обусловлены способами задания и вычисления *r* и *S.*

Эффективность информационного поиска документов, обеспечиваемая ИПС, оценивается по двум показателям:

$k_п$ – коэффициент информационной полноты;

$k_ш$ – коэффициент информационного шума.

Коэффициенты $k_п$ и $k_ш$ принимают значения в интервале от 0 до 1.

В некоторых источниках эти коэффициенты выражают в процентах.

Эффективность информационного поиска

Определим коэффициенты полноты и шума:

$$k_{\text{п}} = \lim_{k \to m} \frac{1}{k} \sum_{i=1}^{k} \frac{|D_i \cap D_i^0|}{|D_i|};$$

$$k_{\text{п}} = \lim_{k \to m} \frac{1}{k} \sum_{i=1}^{k} \frac{|D_i^0 / D_i|}{|D_i^0|},$$

где m — достаточно большое число, чтобы по теореме о больших числах обеспечить требуемую достоверность результата эксперимента по определению $к_{\text{п}}$ и $к_{\text{ш}}$.

Эффективность информационного поиска E_1 выражается через коэффициенты $к_{\text{ш}}$ и $к_{\text{п}}$, что позволяет рассматривать ее в качестве **интегрального показателя эффективности информационного поиска ИПС**.

Эффективность информационного поиска

В литературе в функции $E_1(к_ш, к_п)$ вместо $к_ш$ принято использовать обратный ему показатель — **коэффициент точности $k_т$**.

$$k_т = 1 - k_ш$$

Таким образом, запишем данную функцию в виде:

$$E_1 = \frac{2k_т k_п}{k_т + k_п}$$

В теории информационного поиска предложен **обобщенный комплексный показатель эффективности $E_β$** (мера Ван Ризбергена), позволяющий учитывать предпочтение, отдаваемое пользователем ИПС точности или полноте:

$$E_β = \frac{(β^2 + 1)k_т k_п}{β^2 k_т + k_п},$$

где β — параметр, отражающий предпочтение пользователя ИПС одному из показателей эффективности, входящих в $E_β$ (точности, полноте), над другим.

Эффективность информационного поиска

В литературе в функции $E_1(к_ш, к_п)$ вместо $к_ш$ принято использовать обратный ему показатель — **коэффициент точности $k_т$**.

$$k_т = 1 - k_ш$$

Таким образом, запишем данную функцию в виде:

$$E_1 = \frac{2k_т k_п}{k_т + k_п}$$

В теории информационного поиска предложен **обобщенный комплексный показатель эффективности $E_β$** (мера Ван Ризбергена), позволяющий учитывать предпочтение, отдаваемое пользователем ИПС точности или полноте:

$$E_β = \frac{(β^2 + 1)k_т k_п}{β^2 k_т + k_п},$$

где β — параметр, отражающий предпочтение пользователя ИПС одному из показателей эффективности, входящих в $E_β$ (точности, полноте), над другим.

При β = 1 точность и полнота одинаково важны.

На интервале β ∈ [0; 1[приоритет имеет точность, а на интервале β ∈]1; ∞[— полнота.

КОМПЬЮТЕРНЫЕ МЕТОДЫ ПОИСКА В ТЕКСТЕ

Методы поиска в тексте, используемые человеком:

- поиск «сверху» (по оглавлению с аннотациями глав и, возможно, менее крупных разделов);
- поиск «снизу» (с помощью различных указателей);
- поиск с помощью ГТ связей (перекрестных ссылок);
- полнотекстовый поиск путем просмотра всего текста.

В информационно-поисковых системах применяются следующие методы поиска:

- индексирование текстов и поиск по ключевым словам (по индексу);
- поиск, включающий морфологический разбор и отождествление различных грамматических форм слов;
- поиск с ранжированием документов по степени релевантности запросу;
- использование формальных поисковых языков;
- комплексные методы.

КОМПЬЮТЕРНЫЕ МЕТОДЫ ПОИСКА В ТЕКСТЕ

В технологиях БД и БЗ наряду с перечисленными применяются следующие методы поиска:

- использование формальных языков запросов, позволяющих описывать условия совместного вхождения ключевых слов в документ (это направление представляют *SQL*-подобные языки);
- методы семантического анализа текста.

Средства автоматического извлечения знаний из текстовых ресурсов *Internet* реализуются в поисковых машинах. При этом различают:

1. методы итеративного поиска;
2. методы поиска по выборке;
3. методы, использующие каталоги (рубрикаторы и классификаторы);
4. семантические методы поиска, использующие подходы ИИ.

ПОИСКОВЫЕ СРЕДСТВА

Для поиска информации в *Internet* служат различные классы поисковых средств:

- каталоги (*directories*);
- подборки ссылок (*bookmarks*);
- поисковые машины (*search engines*);
- БД адресов электронной почты (*email addresses databases*);
- средства поиска в архивах *Gopher* (*Gopher archives*);
- системы поиска файлов (*FTP search*);
- системы поиска новостей (*usenet news*) и др.

КОМПЬЮТЕРНЫЕ МЕТОДЫ ПОИСКА В ТЕКСТЕ

Каталог ресурсов *Internet* — постоянно обновляемая и пополняемая система ссылок на ресурсы, распределенные по иерархической структуре категорий. Каталоги облегчают поиск за счет упорядоченности ссылок на ресурсы. Все интеллектуальные функции остаются за человеком.

Подборки ссылок на информационные ресурсы Internet представляют собой отсортированные по темам адреса ресурсов.

Формирование и актуализация каталогов и подборок ссылок выполняются вручную персоналом соответствующих ИС. Подобная работа требует высокой квалификации и достаточно трудоемка.

Наряду с универсальными существуют и специализированные каталоги, систематизирующие сведения о ресурсах Internet, имеющих определенную тематическую направленность.

ПОИСКОВЫЕ МАШИНЫ

Поисковые машины (или **поисковые системы**) позволяют находить ресурсы *Internet* непосредственно по их текстовому содержимому.

Функционирование поисковой машины включает два базовых процесса:

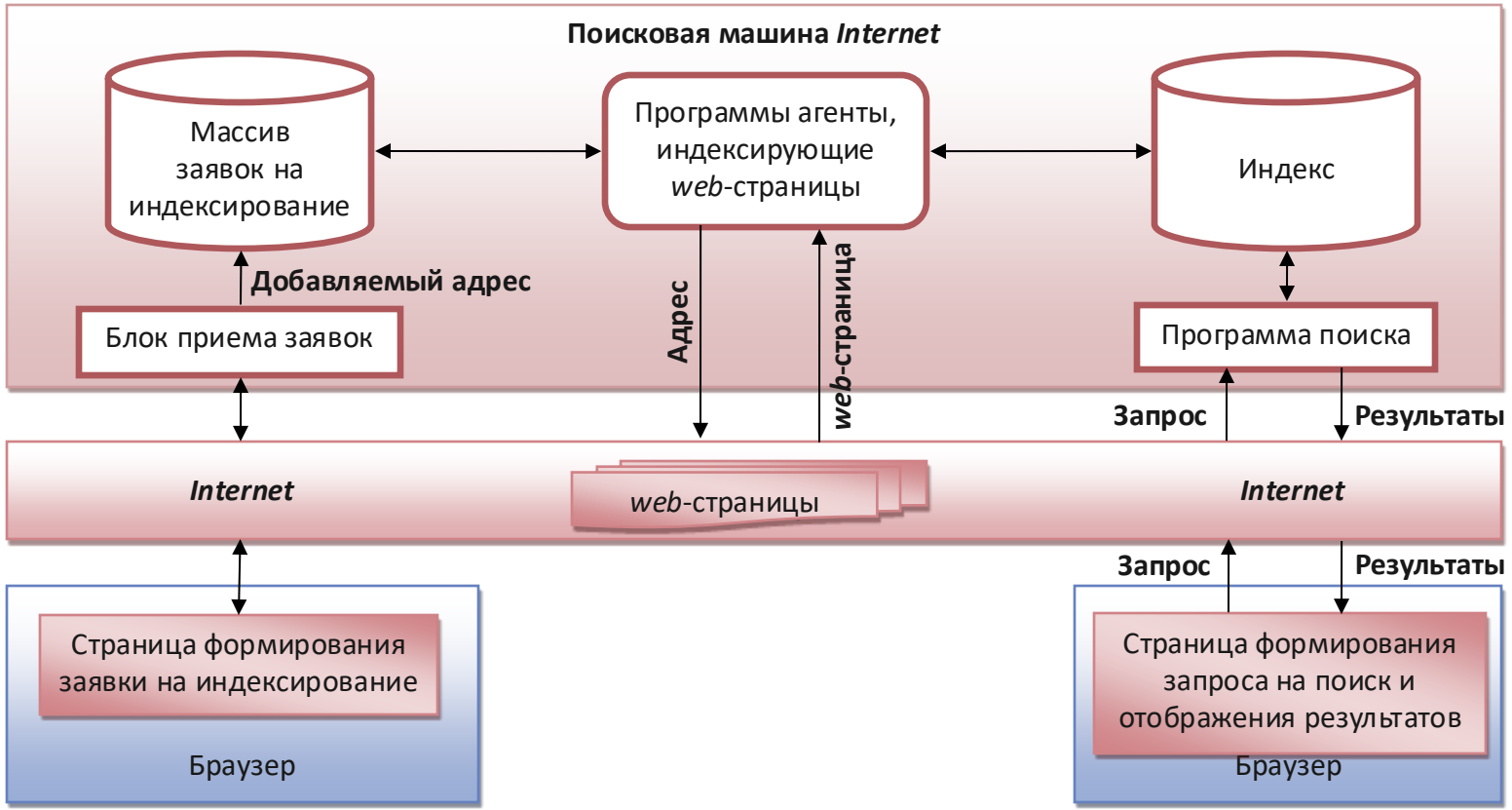
1) индексирование ресурсов *Internet* (автоматическое построение и обновление индекса);

2) поиск по индексу по запросам пользователей.

Главными компонентами типовой поисковой машины являются:

1. **программный агент**, «перемещающийся» по сети и индексирующий ресурсы (*web*-страницы);
2. **БД (индекс)**, содержащая информацию, собираемую агентом;
3. **программа поиска**, применяемая пользователями для поиска информации в БД.

Упрощенная структура типовой поисковой машины



ПОИСКОВЫЕ МАШИНЫ

Агент – самый интеллектуальный из компонентов поисковой машины. Он обладает автономностью, имеет блоки навигации, управляющие «перемещением» по сети, и механизмы индексации, основанные на некоторой базе правил.

Одной из проблем является реализация алгоритма перемещения (навигации) по сети. Предпочтительным для индексирования web-ресурсов принят метод, который осуществляет сначала навигацию вширь, а затем вглубь (это подтверждается статистикой работы поисковых машин).

Разновидностями агентов являются:

- **Кроулеры (*crawlers*)** просматривают заголовки страниц и возвращают поисковой машине только первую найденную ссылку.
- **«Роботы»** проходят по ссылкам различной глубины и вложенности.
- **«Пауки» (*spiders*)** сообщают о содержании найденного документа, индексируют его и пересылают извлеченную информацию в БД поисковой машины.

ПОИСКОВЫЕ МАШИНЫ

Системой правил для всего этого сообщества автономных программ управляют **администраторы поисковых машин**. Они же устанавливают параметры алгоритмов определения степени релевантности документа и запроса.

Обычно в этих алгоритмах учитываются:

- количество слов запроса в текстовом содержимом документа (т.е. в *HTML*-коде);
- теги, в которых эти слова встречаются;
- местоположение искомых слов в документе;
- удельный вес слов, относительно которых определяется релевантность, в общем количестве слов документа;
- время существования *web*-сайта;
- индекс цитируемости *web*-сайта и др.

МЕТОДЫ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ДЛЯ ПОСТРОЕНИЯ ГИПЕРТЕКСТА

МЕТОДЫ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ДЛЯ ПОСТРОЕНИЯ ГИПЕРТЕКСТА

Существуют два класса **источников знаний**:

1) **эксперты** (специалисты в ПрО, для которой формируется ГТ);
2) **текстовые документы на ЕЯ.**

Соответственно **методы извлечения знаний подразделяются на два больших класса:**

1) **приобретение знаний от экспертов (коммуникативные методы);**
2) **обработка документов (текстологические методы).**

МЕТОДЫ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ДЛЯ ПОСТРОЕНИЯ ГИПЕРТЕКСТА

Первый класс методов извлечения знаний имеет следующую структуру.

1.1. Пассивные методы.
- 1.1.1. Наблюдение за работой эксперта.
- 1.1.2. Запись и анализ лекций.
- 1.1.3. Запись и анализ вербальных отчетов.

1.2. Активные методы.
- 1.2.1. Работа с группой экспертов.
  - 1.2.1.1. Метод «мозгового штурма».
  - 1.2.1.2. Метод «круглого стола».
  - 1.2.1.3. Ролевые игры.
- 1.2.2. Индивидуальная работа с экспертом.
  - 1.2.2.1. Анкетирование.
  - 1.2.2.2. Интервьюирование.
  - 1.2.2.3. Свободный диалог.
  - 1.2.2.4. Исследовательская игра с одним экспертом.

МЕТОДЫ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ДЛЯ ПОСТРОЕНИЯ ГИПЕРТЕКСТА

Структура **второго класса методов извлечения знаний** приведена ниже.

2.1. Обработка текстов на ОЕЯ.

2.1.1. Анализ специализированной документации.

2.1.2. Анализ специализированных инструктивных и нормативных материалов (должностных и производственных инструкций, методик и др.).

2.2. Обработка текстов на ЕЯ.

2.2.1. Анализ учебной литературы.

2.2.2. Анализ научной и научно-практической литературы.

2.2.3. Анализ периодических изданий.

2.2.4. Анализ технической документации.

АВТОМАТИЗАЦИЯ ПОСТРОЕНИЯ ГИПЕРТЕКСТА

Ручное формирование ГТ на основе объемного текстового материала весьма трудоемкий процесс.

Для упрощения формирование ГТ служат средства, позволяющие:

- автоматически определять позиции, в которых нужно устанавливать гиперссылки;
- автоматически выявлять связи между документами.

Среди российских программных продуктов можно отметить следующие средства автоматизации построения ГТ:

- авторскую систему ***HyperMethod*** (разработчик — компания «ГиперМетод»), включающую компонент *HyperText Assistant*, выполняющий автоматическую расстановку гиперссылок в формируемом электронном издании на основе системы настраиваемых правил;
- комплексную систему анализа текстов ***TextAnalyst*** (разработчик — научно-производственный инновационный центр «Микросистемы»).

МЕСТО ГИТ СРЕДИ ТЕХНОЛОГИЙ ИИ

ГИТ базируется на основных парадигмах ИИ:

- **использовании БЗ;**
- **логическом выводе;**
- **общении с пользователем на ОЕЯ.**

Гипертекст расширяет возможности человека, связанные с поиском и обработкой информации, за счет установления ассоциаций, построения обобщений, формирования целостного представления о содержании документа и т. д.

В настоящее время существует тенденция интеграции гипертекстовых ИС со специализированными пакетами прикладных программ. При этом возникают гибридные ИС, предназначенные для решения различных классов трудноформализуемых задач. В ряде источников гипертекстовые ИС рассматриваются как представители систем, доставляющих знания.