

Контрольная работа № 3

По теме № 3: «Естественный язык как средство человеко-машинного общения и как объект компьютерной обработки»

ИНФОРМАЦИОННЫЙ ПОИСК

Поиск информации является одной из основных составляющих человеческой деятельности. С ним мы сталкиваемся ежедневно: изучая театральную афишу, чтобы выбрать интересный спектакль; подбирая удобный авиарейс; листая телефонную книгу.

Человеку, в силу своей профессии или увлечений часто сталкивающемуся с подбором и поиском какой-либо тематической информации, рано или поздно (с возрастанием ее объема) приходится применять некоторые принципы систематизации и классификации имеющихся данных, обеспечивающие более удобный и эффективный поиск.

Так, в библиотеках составляют картотеку: сведения о книге по определенной схеме записываются на карточку, туда же помещается шифр – несколько букв и цифр, по которым можно определить местоположение книги (хранилище, стеллаж, полку); карточки расставляются в алфавитном или тематическом порядке. Применение ЭВМ дает более широкие возможности для работы с большими массивами информации.

Очевидно, что работу с большими информационными массивами можно организовать с помощью Баз Данных. Однако еще до создания и широкого распространения БД появился класс прикладных программных систем для поиска информации – **информационно-поисковые системы**.

Основные определения

Информационно-поисковая система (ИПС) - программная система для хранения, поиска и выдачи интересующей пользователя (абонента) информации.

Абонент обращается к ИПС с **информационным запросом** – текстом, отражающим **информационную потребность** данного абонента, например, его желание найти список книг по программированию на языке С++ или список аптек, в которых можно купить нужное лекарство.

Поиск информации ведется в **поисковом массиве**, который формируется (и по мере необходимости обновляется) разработчиками или администраторами системы. Элементы поискового массива вводятся в информационно-поисковую систему на естественном (или близком к нему) языке, а затем обычно подвергаются **индексированию**, т.е. переводу на **формальный информационно-поисковый язык**.

Индексирование – описание документа, факта или запроса на информационно-поисковом языке¹.

По характеру поискового массива и выдаваемой информации ИПС подразделяют на **документальные** и **фактографические**.

Документальная ИПС предназначена для отыскания документов (статей, книг, отчетов, описаний к авторским свидетельствам и патентам, веб-сайтов), содержащих необходимую информацию. Поисковый массив такой ИПС состоит из поисковых образов документов (т.е. элементов, каждый из которых передает основное содержание документа) или из самих документов. В ответ на предъявляемый информационный запрос ИПС выдает некоторое множество документов (или адреса их хранения), содержащих искомую информацию. **Документом** называют любой осмысленный текст, который обладает определенной логической завершенностью и содержит сведения о его источнике и/или создателе.

Фактографическая ИПС обеспечивает выдачу непосредственно фактических сведений, затребованных потребителем в информационном запросе. Поисковый массив состоит из **фактографических записей**, т.е. из описаний фактов, извлеченных из документов и представленных на некотором формальном языке.

¹ Слова "индексирование" или "индексация" используются и в несколько ином (близком) значении. Для текстовой базы составляется список встречающихся в ней терминов, и каждому из них ставится в соответствие некоторый **индекс** (координаты в текстовой базе); чаще всего это номер документа и номер слова в документе. При поступлении запроса слово сначала ищется в этом списке, и по найденным координатам выдаются нужные документы. Если слов в запросе несколько, над их координатами производится операция пересечения. Именно так организован поиск статей, включающих заданное слово, в подсистемах помощи Windows.

Например, документальная ИПС в ответ на запрос «Боинг-777» могла бы выдать ссылки на книги, статьи, инструкции и т.п., касающиеся производства, устройства или использования этих типов самолетов. Фактографическая ИПС в ответ на тот же запрос может выдать список свойств самолетов данного типа: количество двигателей, скорость, дальность полета и др.

В настоящее время фактографические ИПС (как специальный класс поисковых систем) практически не разрабатываются, выполняемые ими действия реализуются с помощью штатных баз данных (БД) и Систем управления базами данных (СУБД). Далее, говоря ИПС, будем иметь в виду документальную информационно-поисковую систему.

Одним из популярных способов перевода документа на информационно-поисковый язык является **координатное индексирование** – индексирование с помощью списка ключевых слов/словосочетаний, определяющих содержание этого документа.