

М 1329

МИНИСТЕРСТВО ПО РАЗВИТИЮ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ И КОММУНИКАЦИЙ РЕСПУБЛИКИ УЗБЕКИСТАН
ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ ИМЕНИ МУХАММАДА АЛЬ-ХОРЕЗМИ

КАФЕДРА «ИНФОРМАЦИОННО-БИБЛИОТЕЧНЫЕ СИСТЕМЫ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ
ПРАКТИЧЕСКИХ И ЛАБОРАТОРНЫХ РАБОТ
по предмету:

«ИНФОРМАЦИОННО - ПОИСКОВЫЕ СИСТЕМЫ»

ТАШКЕНТ - 2022

Составитель: Ануфриева Ю.Г., ассистент кафедры
«Информационно – библиотечные системы»

Методические указания по выполнению практических и лабораторных работ по предмету «Информационно – поисковые системы» были обсуждены и рекомендованы к утверждению на заседании кафедры «Информационно-библиотечные системы» по направлению «Библиотечно – информационная деятельность», Протокол № 11 от 10 января 2022 года.

Методические указания по выполнению практических и лабораторных работ по предмету «Информационно – поисковые системы» были рассмотрены и утверждены на методическом совете факультета «Профессиональное образование в сфере ИКТ», Протокол № 17(5) от 14 января 2022 года.

Методические указания по выполнению практических и лабораторных работ по предмету «Информационно – поисковые системы» были рассмотрены и утверждены на Учебно-методическом совете ТУИТ 30 марта 2022 года Протокол № 8 (154).

ВВЕДЕНИЕ

Предмет «Информационно – поисковые системы» является предметом на выбор, среди учебных дисциплин информационно-библиотечного цикла, которые изучаются в 7 семестре студенты бакалавра направления 5350600 - «Библиотечно – информационная деятельность». Его целью является обеспечение профессиональной подготовки в области информационно-библиотечной деятельности, в частности, применение и использование студентами практических навыков и умений поиска информации в автоматизированных информационно-библиотечных системах и поисковых системах сети Интернет, на основе выполнения практических заданий

Содержание изучаемого предмета включает: предоставление информации о концепции и организации поиска информации, поиск информации в традиционных и электронных каталогах информационно – библиотечных учреждений, характеристики и классификации информационно-поисковых систем, основы веб-поиска и языков поиска.

Целью практических и лабораторных занятий является - закрепление теоретических знаний студентов по материалам лекционных занятий и формированию устойчивых умений и навыков при выполнении практических и лабораторных работ.

Практические (лабораторные) задания выполняются письменно согласно структуре, подготовительный этап может обсуждаться в малой группе, далее задания сдаются студентом в виде реферата по соответствующей теме с использованием конспектов лекций, раздаточных материалов, рекомендуемой литературы и Интернет-ресурсов.

1. СТРУКТУРА ПРАКТИЧЕСКОЙ (ЛАБОРАТОРНОЙ) РАБОТЫ

1.1. Практическая (лабораторная) работа должна быть оформлена согласно требованиям и состоять из следующих частей:

- Титульный лист;
- Содержание;
- Введение;
- Основная часть: главы, параграфы, схемы, структуры, диаграммы, анализ тематики;
- Заключение;
- Список использованной литературы и ресурсов;
- Приложения (если имеются)

1.2. Титульный лист практической (лабораторной) работы является первым листом и выполняется в установленном порядке. После титульного листа располагается содержание.

1.3. Содержание включает название основных разделов.

1.4. Введение состоит из следующих частей:

- Актуальность темы;
- Цель и задачи данной работы;
- Методологическая база;
- Практическая значимость и применение.

1.5. Разделы в основной части работы включают:

- Анализ, современные тенденции и перспективы развития;
- Общая постановка проблемы и пути её решения;
- Недостатки и преимущества практического решения рассматриваемых задач.

1.6. В заключение Практической (лабораторной) работы необходимо представить рекомендации и методы решения поставленных задач. Рассмотреть перспективы развития и возможность применения на практике инновационных технологий по рассматриваемым вопросам. Каждый студент в итоге должен иметь представление о цели, задачах, методах решения и достигнутых ожидаемых результатов.

1.7. В список использованной литературы входят все источники, использованные студентом в лабораторной работе. Библиографическое описание составляется согласно

существующим правилам описания по типу издания, в алфавитном, хронологическом, систематическом порядке.

2. ОСНОВНЫЕ ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ ПРАКТИЧЕСКОЙ (ЛАБОРАТОРНОЙ) РАБОТЫ

2.1. Содержание практической (лабораторной) работы должно соответствовать основной цели и поставленным задачам.

2.2. Текст должен быть набран на компьютере (печатный и электронный варианты).

2.3. Объём практической (лабораторной) работы должен быть от 3-х страниц и более. Текст должен быть напечатан с использованием шрифта Times New Roman, кегль 14, интервал - 1,5; поля: слева - 30 мм, справа, сверху, снизу-20 мм, печатный вариант - формат А4, электронный вариант собирается на общий групповой диск

2.4. Практическая (лабораторная) работа сдаётся в печатном варианте и защищается на текущем лабораторном занятии.

Защита практической (лабораторной) работы с использованием презентации в программе Power Point приветствуется.

3. ТЕМЫ ДЛЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

№	темы практических занятий	КОЛ-ВО ЧАСОВ
1.	Обеспечение поиска информации в традиционной библиотеке	2
2.	Организация поиска информации в современной библиотеке	2
3.	Роль электронных каталогов в автоматизированных библиотечных системах	2
4.	Модуль поиска информации в автоматизированных библиотечных системах	2
5.	Алгоритмы поиска информации в массивах	2

6.	Принципы веб-поиска	2
7.	Структура и принципы работы национальной поисковой системы WWW.Uz.	2
8.	Индексирование. Статистический анализ текста (закон Ципфа).	2
9.	Изучение основы классифицированного ИПС	2
10.	История и принципы работы поисковой системы Google	2
11.	История и принципы работы поисковой системы Яндекс	2
12.	Содержание HTML-документов	2
13.	Алгоритмы поиска XML. Программирование поиска XML	2
14.	Булева алгебра в поиске. Принципы поиска векторной модели	2
15.	Ранжирование результатов поиска и важность терминов	2
Всего		30

3.1. ТЕМЫ, РЕКОМЕНДОВАННЫЕ ДЛЯ ПРОВЕДЕНИЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ ПО ПРЕДМЕТУ “ИНФОРМАЦИОННО – ПОИСКОВЫЕ СИСТЕМЫ”

Практическая работа № 1.

Обеспечение поиска информации в традиционной библиотеке.

Цель задания: Изучить основы поиска информации в традиционной библиотеке, по каким параметрам определяется поиск.

Содержание задания: Библиотечный каталог – совокупность расположенных по определенным правилам библиографических записей на документы, раскрывающая состав и содержание фонда библиотеки. Каталоги могут иметь сложную структуру и

подразделяться на части и ряды. Библиотечные каталоги возникли в глубокой древности. Свидетельством наличия каталогов являются дошедшие до нас фрагменты каталогов. С тех пор каталоги неоднократно претерпели изменения во внешнем виде и в принципах организации. В библиотеках хранятся огромные объемы информации различных форматов и на различных носителях. Поиск по такому разнообразию материалов часто требует участия специалиста. Поиск требуемых материалов усложняется еще больше, если пользователь затрудняется сформировать четкий запрос, а лишь обрисовывает тематику.

Традиционно библиотеки предоставляют возможность поиска по электронному каталогу и подписным электронным ресурсам. В единичных случаях предоставляется доступ к оцифрованным электронным фондам, не попадающим под ограничения авторского права. Сложность поиска информации заключается в том, что для каждого электронного ресурса существует своя поисковая система.

Технология выполнения задания: Ознакомится с системой поиска в традиционной библиотеке, на примере библиотечного учреждения, дать необходимую характеристику библиотечным каталогам.

Практическая работа № 2.

Организация поиска информации в современной библиотеке.

Цель задания: Изучить основы поиска информации в библиотечных учреждениях.

Содержание задания: **Справочно-поисковый аппарат библиотеки** - совокупность традиционных и электронных библиотечных каталогов и картотек, справочных и библиографических изданий, библиографических и фактографических баз данных, используемых при обслуживании читателей для поиска необходимой библиографической и фактографической информации.

Состав СПА:

- система библиотечных каталогов и картотек, традиционных и электронных, баз данных;
- справочно-библиографический фонд;
- фонд (архив) выполненных библиографических справок;
- приобретённые эл. ресурсы, доступные читателям в стенах б-ки.

Одним из обязательных элементов СПА являются каталоги, перечень библиогр. записей документов, имеющихся в фонде одной или нескольких библиотек, раскрывающий состав или содержание библиотечных фондов.

Так как основой поиска информации в библиотеке является её **Справочно-поисковый аппарат (СПА)**, который включает в себя:

Каталоги библиотеки:

- алфавитный (АК),
- систематический (СК),
- электронный (ЭК);

Картотеки:

- систематическая картотека статей (СКС),
- тематические картотеки,
- специальные картотеки;

Фонд справочных изданий:

- энциклопедии (*универсальные, отраслевые, популярные*),
- словари (*энциклопедические, отраслевые, специальные*),
- справочники ,
- библиографические указатели (*научно-вспомогательные, рекомендательные*).

Библиотечный каталог – это перечень *библиографических описаний* тех документов, которые имеются в фонде библиотеки.

Технология выполнения задания: Определить основные особенности библиотечного каталога, возможности различных каталогов и картотек.

Практическая работа № 3.

Роль электронных каталогов в автоматизированных библиотечных системах.

Цель задания: Изучить роль каталогов в автоматизированных библиотечных системах.

Содержание задания: Электронный каталог (ЭК) - это библиотечный каталог в машиночитаемой форме, работающий в реальном режиме времени и представленный в распоряжение читателей библиотеки. ЭК обеспечивает доступ к ресурсам каталога не только пользователям библиотеки, но и удаленным пользователям.

Электронный каталог отвечает современным требованиям, который обеспечивает:

- ввод библиографических записей в диалоговом и пакетном режиме с удаленных или локальных терминалов;
- поиск информации в диалоговом или пакетном режимах по разовым запросам и избирательное распространение информации;
- печатание библиографических данных в виде каталожных карточек, библиографических указателей, отчетов по управленческой информации и др.;
- возможность обмена информацией с другими библиографическими базами данных, в том числе и подключение к сети Интернет;
- редактирование библиографических записей и при необходимости удаление их из ЭК.

Технология выполнения задания: Дать характеристику электронному каталогу в автоматизированной библиотечной системе (система на выбор студента).

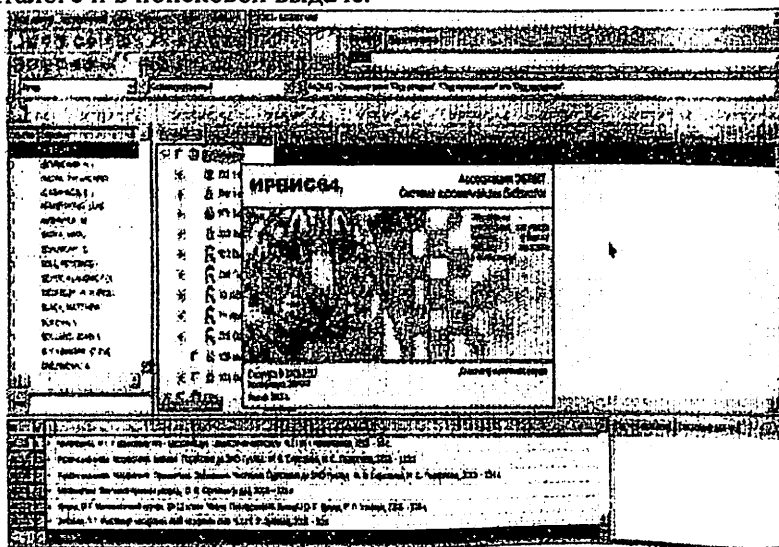
Практическая работа № 4.

Модуль поиска информации в автоматизированных библиотечных системах.

Цель задания: Изучить технологию поиска информации в автоматизированной библиотечной системе.

Содержание задания: Библиотечная система ИРБИС64 в базовой поставке позволяет организовать электронный каталог с возможностями поиска по метаданным, индексируемым этой АБИС. За коммуникацию с пользователями отвечает модуль

АРМ (автоматизированное рабочее место — прим.) «Читатель». Доступ к поисковому окну возможен через веб-браузер. Производители АБИС называют веб-интерфейс своего каталога дружественным, заявляя, что он рассчитан на пользователей, не обладающих специальными знаниями. Поиск по электронному каталогу индексирует не только записи физических фондов, но и информацию из внешних баз данных, например, из репозиториев и ЭБС. Для индексирования электронных ресурсов библиотеке необходимо импортировать нужные метаданные или добавить их вручную. После этого записи электронных книг будут доступны в каталоге и в поисковой выдаче.



Технология выполнения задания: Показать технологию поиска информации в автоматизированной библиотечной системе (система на выбор студента).

Практическая работа № 5. Алгоритмы поиска информации в массивах.

Цель задания: Изучить порядок поиска информации в массивах данных.

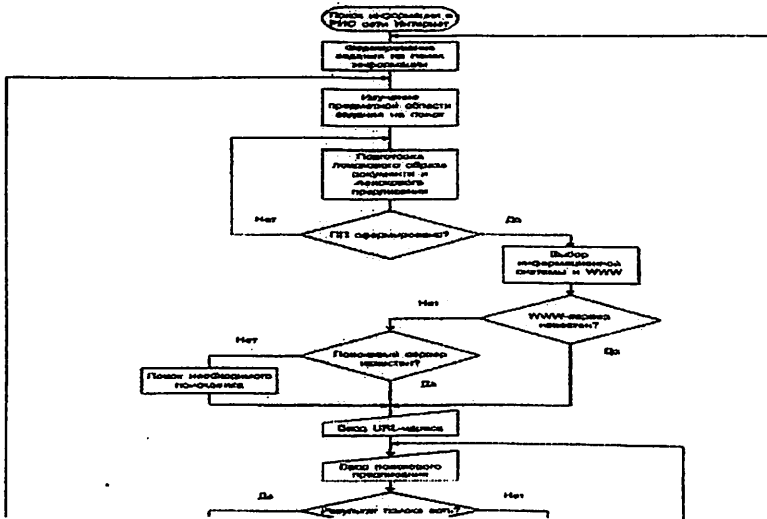
Содержание задания: Последовательный поиск (Sequential Search), называемый также линейным поиском, является самым простым из всех алгоритмов поиска. Это метод поиска одного значения в коллекции “в лоб”. Он находит значение, начиная с первого элемента коллекции и исследуя каждый последующий элемент до тех пор, пока не просмотрит всю коллекцию или пока соответствующий элемент не будет найден. Для работы этого алгоритма должен иметься способ получения каждого элемента из коллекции, в которой выполняется поиск; порядок извлечения значения не имеет. Часто элементы коллекции могут быть доступны только через итератору который предназначен только для чтения и получает каждый элемент из коллекции.

Входные данные представляют собой непустую коллекцию и целевое значение, которое мы ищем. Поиск возвращает true, если коллекция содержит целевое значение и false в противном случае. Зачастую требуется найти элемент в коллекции, которая может быть не упорядочена. Без дополнительных знаний об информации, хранящейся в коллекции, последовательный поиск выполняет работу методом “грубой силы”. Это единственный алгоритм поиска, который можно использовать, если коллекция доступна только через итератор.

Если коллекция неупорядоченная и хранится в виде связанного списка, то вставка элемента является операцией с константным временем выполнения (он просто добавляется в конец списка). Частые вставки в коллекцию на основе массива требуют управления динамической памятью, которое либо предоставляется базовым языком программирования, либо требует определенного внимания со стороны программиста.

Последовательный поиск накладывает наименьшее количество ограничений на типы искомых элементов.

Единственным требованием является наличие функции проверки совпадения элементов, необходимой для выяснения, соответствует ли искомый элемент элементу коллекции; часто эта функция возлагается на сами элементы.



Технология выполнения задания: Изучить теоретическую часть и показать алгоритм поиска в массивах данных.

Практическая работа № 6. Принципы веб-поиска.

Цель задания: Изучить особенности и принципы веб-поиска.

Содержание задания: Общий принцип работы любой поисковой системы условно можно разделить на следующие этапы:

1. **Сбор информации.** Специальная программа сканирует веб-пространство, открывает каждый доступный для неё сайт и анализирует его по заданным алгоритмам.
2. Все документы **закачиваются** на сервер поисковой системы и создается база данных, которая содержит информацию о сайте.
3. На основе полученных по сайту данных проводится построения индекса. То есть определяется, какие данные на нём содержатся, к какой группе запросов относятся данный контент их можно отнести и так далее.
4. Программа определяет релевантность страницы, в момент когда она получает пользовательский поисковый запрос, на его

основе предоставляет перечень сайтов, которые по результатам индексирования содержат запрашиваемую информацию.

5. **Сервис проводит ранжирование результатов выдачи.** То есть выстраивает порядок ссылок, которые будут показаны пользователю, отправившему запрос.

Поисковые системы бывают нескольких подвидов и существуют и другие вариации таких сервисов:

управляемые человеком (то есть каталог сайтов, каждый пункт которого и общая их база данных сформированы вручную пользователем, яркий пример каталог Rambler);

гибридные поисковые системы (где часть работы выполняет человек, часть — программа, принцип работы поисковой системы Google как раз таковой);

мета-системы (которые не составляют базу данных, а дают результат сразу из нескольких поисковых сервисов пример Vivisimo).

Как составить поисковый запрос ?

1. Грамотное написание слов обеспечивает максимальное количество совпадений с искомым информационным объектом (Хоть современные поисковые системы уже научились исправлять орфографические ошибки, но данным советом пренебрегать не стоит).

2. Благодаря использованию синонимов в запросе, можно охватить более широкий поисковой диапазон.

3. Иногда изменение слова в тексте запроса может принести больший результат осуществляйте реформирование запроса.

4. Привнесите в запрос конкретность, используйте точные вхождения фраз, которые должны определять главную суть поиска.

5. Экспериментируйте с ключевыми словами. Использование ключевых слов и словосочетаний может помочь определить главную суть, и поисковая машина выдаст более релевантный результат.

Технология выполнения задания: Составить несколько поисковых запросов и показать результат.

Практическая работа № 7. Структура и принципы работы национальной поисковой системы WWW.UZ.

Цель задания: Изучение структуры и принципы работы www.uz. Изучение основных характеристик функционирования поисковой системы.

Содержание задания:

Национальная информационно-поисковая система WWW.UZ — это механизм быстрого доступа к информации национального сегмента сети Интернет. Основными особенностями WWW.UZ системы являются многоязыковой поиск информации (узбекский, русский) и тесная интеграция с другими национальными информационными системами и базами данных.

Информационно-поисковая система WWW.UZ осуществляет поиск веб-сайта:

- По адресу
- По содержимому

В сервисы WWW.UZ входят разделы:

- Каталог
- Топ-рейтинг
- Сводная статистика,
- Новости

Раздел «Каталог» WWW.UZ — это описания зарегистрированных и систематизированных по тематическим каталогам сайтов Республики Узбекистан, находящихся в открытом доступе в сети Интернет.

Каталог облегчает поиск ресурса, предлагая рубрику поиска (Наука, Культура, Спорт, Общество, Бизнес, Новости, и т. д.). В дополнение к рубрикации по темам, пользователю предлагается классификация сайтов по типу содержащейся в них информации (Конференции и семинары Библиотека, Музеи, События, Форумы и чаты, Организации и т. д.). Каталог ежедневно пополняется новыми ресурсами благодаря активным пользователям поисковой системы и модераторам WWW.UZ. Добавить адрес нового сайта в каталог, может любой пользователь, зарегистрированный в системе ID.UZ

WWW.UZ это открытый и доступный для просмотра топ-рейтинг и сводная статистика сайтов. Система WWW.UZ ведет открытую статистику сайтов Узбекистана. Этим определяется и оценивается популярность ресурса. На всех сайтах, участвующих в рейтинге, установлены счетчики системы WWW.UZ. Работу по расширению и модернизации WWW.UZ осуществляет Центр развития и внедрения компьютерных и информационных технологий UZINFOCOM.

Технология выполнения задания: Изучить особенности функционирования поисковой системы WWW.UZ.

Практическая работа № 8. Индексирование. Статистический анализ текста (закон Ципфа).

Цель задания: Изучить понятие статистического анализа текста.

Содержание задания: В конце 40-х годов прошлого столетия Дж. Ципф, собрав огромный статистический материал, попытался показать, что распределение слов естественного языка подчиняется одному простому закону, который можно сформулировать следующим образом. Если к какому-либо достаточно большому тексту составить список всех встретившихся в нем слов, затем расположить эти слова в порядке убывания частоты их встречаемости в данном тексте и пронумеровать в порядке от 1 (порядковый номер наиболее часто встречающегося слова) до R , то для любого слова произведение его порядкового номера (ранга) в таком списке и частоты его встречаемости в тексте будет величиной постоянной, имеющей примерно одинаковое значение для любого слова из этого списка. Аналитически закон Ципфа может быть выражен в виде:

$$f_r = c,$$

где f - частота встречаемости слова в тексте;

r - ранг (порядковый номер) слова в списке;

c - эмпирическая постоянная величина.

Полученная зависимость графически выражается гиперболой. Исследовав таким образом самые разнообразные тексты и языки, в том числе языки тысячелетней давности, Дж. Ципф для каждого из них построил указанные зависимости, при этом все кривые

имели одинаковую форму - форму «гиперболической лестницы», т.е. при замене одного текста другим общий характер распределения не изменялся. Закон Ципфа был открыт экспериментально. Позднее Б. Мандельброт предложил его теоретическое обоснование. Он полагал, что можно сравнивать письменный язык с кодированием, причем все знаки должны иметь определенную «стоимость». Исходя из требований минимальной стоимости сообщений, Б. Мандельброт математическим путем пришел к аналогичной закону Ципфа зависимости

$$f \cdot r^c = c,$$

где r - величина (близкая к единице), которая может изменяться в зависимости от свойств текста.

Дж. Ципфом и другими исследователями было установлено, что такому распределению подчиняются не только все естественные языки мира, но и другие явления социального и биологического характера: распределения ученых по числу опубликованных ими статей (А. Лотка, 1926 г.), городов США по численности населения (Дж. Ципф, 1949 г.), населения по размерам дохода в капиталистических странах (В. Парето, 1897 г.), биологических родов по численности видов (Дж. Уиллис, 1922 г.) и др. Самым важным для рассматриваемой проблемы является тот факт, что и документы внутри какой-либо отрасли знаний могут распределяться согласно этому закону. Частным случаем его является закон Бредфорда, непосредственно связанный уже не с распределением слов в тексте, а с распределением документов внутри какой-либо тематической области.

Технология выполнения задания: Изучить общие сведения по закону Ципфа.

Практическая работа № 9.

Изучение основ классифицированных информационно поисковых систем.

Цель задания: Изучить теоретические и практические сведения классификации ИПС.

Содержание задания: Информационные системы обеспечивают подготовку, ввод, хранение, обработку, контроль и передачу данных. Они отличаются иерархической структурой. Степень их автоматизации обычно относительно высока. Информационные системы бывают реализованы как сеть взаимосвязанных вычислительных машин разной величины и абонентских пунктов (терминалов). Их подсистемы выполняют функции на различных уровнях управления, как правило, используя общий банк данных. Классифицировать информационные системы достаточно сложно из-за их разнообразия и постоянного развития структур и функций.

В качестве признаков классификации используются: область применения, охватываемая территория, организация информационных процессов, направление деятельности, структура и др.

По территориальному признаку информационная поисковая система (ИПС) классифицируются на геоинформационные, международные, общегосударственные, областные, республиканские, окружные, городские, районные и т.д.

В полнотекстовых ИПС собираются и систематизируются тексты документов или их библиографическое описание. Поскольку документированная информация - это зафиксированная на материальном носителе путем документирования информация с реквизитами, позволяющими определить такую информацию. Опыт практического применения ИПС показал, что наиболее точной, соответствующей самому назначению ИПС следует считать классификацию по степени сложности технической, вычислительной, аналитической и логической обработки используемой информации. При таком подходе к классификации можно наиболее тесно связать ИПС и соответствующие информационные технологии.

Соответственно, можно выделить следующие виды ИПС:

- - информационно-справочные системы (ИСС);
- - информационно-логические системы (ИЛС);
- - экспертные системы (ЭС) и системы поддержки принятия решений;
- - автоматизированные рабочие места (АРМ);

- - автоматизированные системы управления (АСУ).

Технология выполнения задания: Дать полную характеристику практического применения классификации ИПС.

Практическая работа № 10.

История и принципы работы поисковой системы Google.

Цель задания: Изучить принципы работы поисковой системы Google.

Содержание задания: Своим появлением поисковые системы напрочь изменили привычный для нас способ сбора информации. Интересно ли вас обновление данных фондового рынка или вы хотите найти лучший ресторан в районе, либо пишете академический отчет об Эрнесте Хемингуэе - поисковик даст ответ на все запросы. В 80 годы ответы на вопросы потребовали бы посещения местной библиотеки. Теперь же все решается в течении миллисекунды с использованием алгоритмических полномочий поисковика. В этом отношении главная цель поисковой системы заключается в том, чтобы максимально быстро найти уместную и актуальную информацию, как ответ на введенные поисковые термины, также называемые ключевыми словами. Поэтому центральным аспектом для любой поисковой системы, желающей выдать действительно полезный результат, является понятие цели поиска, того, как именно люди ищут. Результат работы Google можно сравнить с интернет-каталогом, отобранным с помощью рейтинговой системы на основе алгоритмов. Более конкретно алгоритм поиска можно описать как «нахождение элемента с заданными свойствами среди списка элементов».

Сканирование может быть описано, как автоматизированный процесс систематического изучения общедоступных страниц в Интернете. Проще говоря, во время этого процесса Google обнаруживает новые или обновленные страницы и добавляет их в свою базу. Для облегчения работы он использует специальную программу. «Googlebots» (можно встретить альтернативные названия: «боты» или «роботы») посещают список URL-адресов, полученных в процессе прошлого сканирования и дополненных

данными карты сайта, которую предоставляют веб-мастера и анализируют их содержание. При обнаружении ссылок на другие страницы во время посещения сайта, боты также добавляют их в свой список и устанавливают систематические связи. Процесс сканирования происходит на регулярной основе в целях выявления изменений, изъятия «мертвых» ссылок и установления новых взаимосвязей.

Индексация — процесс сохранения полученной информации в базе данных в соответствии с различными факторами для последующего извлечения информации. Ключевые слова на странице, их расположение, мета-теги и ссылки представляют особый интерес для индексации Google. Для того чтобы эффективно хранить информацию о миллиардах страниц в базе данных поисковой системы, Google использует крупные центры обработки данных в Европе, Азии, Северной и Южной Америке. В этих центрах, как было подсчитано, на основе энергопотребления Google в 2010 году, работает около 900,000 серверов.

Технология выполнения задания: На основе собранных сведений подготовить презентацию о принципах работы поисковой системы Google.

Практическая работа № 11.

История и принципы работы поисковой системы Яндекс.

Цель задания: Изучить исторические сведения и работы поисковой системы Яндекс.

Содержание задания:

- **Алгоритм прямого поиска.**

Что это такое – вы помните, что читали замечательную историю в одной из книг. И вы начинаете по очереди искать. Взяли одну книгу – полистали – не нашли, взяли другую... Принцип понятен, но этот способ чрезвычайно долгий. Это тоже понятно.

- **Алгоритм обратного поиска.**

Для этого алгоритма создается из каждой страницы твоего блога – создается текстовый файл. В этом файле перечисляются в

алфавитном порядке ВСЕ слова, которые ты использовал. Даже позиция этого слова в тексте указывается (координаты в тексте). Это достаточно быстрый способ, но уже поиск происходит с какой-то погрешностью. Алгоритм этот ищет не в интернете, не поиском по блогу. А в отдельно взятом текстовом файле, который создан был когда-то давно. Когда робот заходил к тебе. И эти файлы (обратные индексы) хранятся на серверах Яндекса. Так, это были базовые алгоритмы поиска. Т.е. как Яндекс просто находит нужные документы. С этим вроде бы проблем не должно быть.

Но ведь документов Яндекс знает не один и даже не 100, а по последним данным из моих источников – Яндекс знает порядка 11 млрд. документов (10 727 736 489 страниц).

И среди всего этого количества нужно выбрать документы, подходящие под запрос. И что еще важнее – нужно как-то ранжировать их. Т.е. выстроить по степени важности, а точнее по степени полезности для читателя.

Математические модели поиска

Для решения этого вопроса на помощь приходят математические модели. Вот о простейших моделях мы сейчас и поговорим.

Булевская мат.модель – Если слово встречается в документе – документ считается найденным. Просто на совпадение и ничего сложного.

Но тут есть проблемы. Например, если ты как пользователь введешь какое-то популярное слово, а еще лучше предлог «в», который является самым распространенным словом в русском языке и встречается в КАЖДОМ документе – то тебе выдаст такое количество результатов, что ты даже не осознаешь такую цифру, сколько тебе документов нашлось. Поэтому появилась следующая мат.модель.

Векторная мат.модель – эта модель определяет «вес» документа. Уже не только совпадение встречается, но и это слово должно встречаться несколько раз. Причем чем больше слово встречается – тем выше релевантность (соответствие).

Вероятностная модель – более сложная. Принцип такой: поисковик нашел сам эталон страницы. Например, вы ищите информацию об истории Яндекса. У Яндекса хранится какой-то

эталон, допустим это будет моя предыдущая статья о Яндексе. И все остальные документы он будет сравнивать с этой статьёй. И логика здесь такая: чем более страница твоего блога похожа на мою статью – тем вероятнее тот факт, что твоя страница блога тоже будет полезна читателю и тоже рассказывает об истории Яндекса.

Чтобы сократить количество документов, которые нужно показывать пользователю – было введено понятие релевантности, т.е. соответствия.

Технология выполнения задания: Изучить историю и провести анализ принципов работы поисковой системы Яндекс.

Практическая работа № 12. Содержание HTML - документов.

Цель задания: Изучить содержание HTML - документов.

Содержание задания: HTML - документ — это обычный текстовый документ, может быть создан как в обычном текстовом редакторе (Блокнот), так и в специализированном, с подсветкой кода (Notepad++, Visual Studio Code и т.п.). HTML - документ имеет расширение `html`.

HTML-документ состоит из дерева HTML - элементов и текста. Каждый элемент обозначается в исходном документе начальным (открывающим) и конечным (закрывающим) тегом (за редким исключением).

Начальный тег показывает, где начинается элемент, конечный - где заканчивается. Закрывающий тег образуется путем добавления слэша / перед именем элемента: `<имя элемента>... </имя элемента>`. Между начальным и закрывающим тегами находится содержимое элемента — контент. Элементы, представленные одиночными тегами, не могут хранить в себе содержимого напрямую, оно прописывается как значение атрибута, например, элемент `<input type="button" value="Кнопка">` создаст кнопку с текстом Кнопка внутри.

Элементы могут вкладываться друг в друга, например, `<p><i>Текст</i></p>`. При вложении следует соблюдать порядок их закрытия (принцип «матрёшки»),

например, следующая запись будет неверной: `<p><i>Текст</i></p></i>`.

HTML - элементы могут иметь атрибуты (глобальные, применяемые для всех HTML - элементов, и собственные). Атрибуты прописываются в открывающем теге элемента и содержат имя и значение, указываемые в формате `имя атрибута="значение"`. Атрибуты позволяют изменять свойства и поведение элемента, для которого они заданы.

Каждому элементу можно присвоить несколько значений `class` и только одно значение `id`. Множественные значения `class` записываются через пробел, `<div class="nav top">`.

Значения `class` и `id` должны состоять только из букв, цифр, дефисов и нижних подчеркиваний и должны начинаться только с букв или цифр. Браузер просматривает (интерпретирует) HTML-документ, выстраивая его структуру (DOM) и отображая ее в соответствии с инструкциями, включенными в этот файл (таблицы стилей, скрипты). Если разметка правильная, то в окне браузера будет отображена HTML-страница, содержащая HTML-элементы — заголовки, таблицы, изображения и т.д.

Процесс интерпретации (парсинг) начинается прежде, чем веб-страница полностью загружена в браузер. Браузеры обрабатывают HTML-документы последовательно, с самого начала, при этом обрабатывая CSS и соотнося таблицы стилей с элементами страницы.

HTML-документ состоит из двух разделов — заголовка — содержимое элемента `<head>` и содержательной части — содержимое `<body>`.

Технология выполнения задания: Показать пример содержания HTML - документов.

Практическая работа № 13. Алгоритмы поиска XML. Программирование поиска XML.

Цель задания: Изучить алгоритмы и программирование поиска XML.

Содержание задания: XML расшифровывается как eXtensible

Markup Language — «расширяемый язык разметки». Один из языков разметки тебе, возможно, уже знаком: ты слышал об HTML, с помощью которого создаются веб-страницы :) XML — это язык для описания данных.

Его главная составная часть — теги: вот такие штуки в угловых скобках:

```
<book>
</book>
```

Теги бывают открывающими и закрывающими. У закрывающего есть дополнительный символ — “/”, это видно на примере выше. Каждому открывающему тегу должен соответствовать закрывающий. Они показывают, где начинается и где заканчивается описание каждого элемента в файле. Теги могут быть вложенными! В нашем примере с книгой у тега <book> есть 3 вложенных тега — <title>, <author> и <year>. Это не ограничивается одним уровнем: у вложенных тегов могут быть свои вложенные теги, и т. д. Такая конструкция называется деревом тегов. Давай рассмотрим дерево на примере XML-файла с описанием автосалона:

```
<?xml version="1.0" encoding="UTF-8"?>
<carstore>
  <car category="truck">
    <model lang="en">Scania R 770</model>
    <year>2005</year>
    <price currency="US dollar">200000.00</price>
  </car>
  <car category="sedan">
    <title lang="en">Ford Focus</title>
    <year>2012</year>
    <price currency="US dollar">20000.00</price>
  </car>
  <car category="sport">
    <title lang="en">Ferrari 360 Spider</title>
    <year>2018</year>
    <price currency="US dollar">150000.00</price>
  </car>
</carstore>
```

Здесь у нас есть тег верхнего уровня — <carstore>. Его еще называют «root» — корневой тег. У <carstore> есть один дочерний тег — <car>. У <car>, в свою очередь, тоже есть 3 своих дочерних тега — <model>, <year> и <price>. Каждый тег может иметь атрибуты — дополнительную важную информацию. В нашем примере у тега <model> есть атрибут «lang» — язык, на котором написано название модели:

```
<model lang="en">Scania R 770</model>
```

Так мы можем указать, что название написано на английском языке. У нашего тега <price> (цена) есть атрибут «currency» — «валюта».

```
<price currency="US dollar">150000.00</price>
```

Так мы можем указать, что цена за машину указана в американских долларах.

Технология выполнения задания: Привести примеры алгоритма поиска XML.

Практическая работа № 14.

Булева алгебра в поиске. Принципы поиска векторной модели.

Цель задания:

Содержание задания: Булевский поиск в контексте поисковой системы - это тип поиска, в котором вы можете использовать специальные слова или символы для ограничения, расширения или определения поиска.

Это возможно с помощью логических операторов, таких как AND, OR, NOT и NEAR, а также с помощью символов + (сложение) и — (вычитание).

Когда вы включаете оператор в булевский поиск, вы либо вводите гибкость для получения более широкого диапазона результатов, либо определяете ограничения для уменьшения числа несвязанных результатов.

Булевы операторы поиска

Для целей булева веб-поиска вам необходимо знать следующие термины и символы:

Логический оператор	Символ	объяснение	пример
И	+	Все слова должны присутствовать в результатах	компьютер и интернет
ИЛИ ЖЕ		Результаты могут включать любое из слов	палео или первобытный
НЕ	—	Результаты включают все, кроме термина, который следует за оператором	диета НЕ веганская
ОКОЛО		Поисковые термины должны появляться в определенном количестве слов друг друга	казахский министр

Вектор - это направленный отрезок, то есть отрезок, имеющий длину и определенное направление.

Векторная модель (англ. **vector space model**) — в информационном поиске представление коллекции документов векторами из одного общего для всей коллекции векторного пространства.

Технология выполнения задания: Изучив теоретические сведения булевого поиска провести поиск информации по различным

параметрам, с помощью выбранной студентом поисковой системы, показать результат.

Практическая работа № 15.

Ранжирование результатов поиска и важность терминов.

Цель задания: Определить основные понятия и термины, ранжирования результатов поиска в существующих поисковых системах.

Содержание задания: Ранжирование - это сортировка сайтов поисковыми системами в результатах выдачи по соответствующему поисковому запросу.

Ранжирование - это порядок выдачи найденных сайтов.

Алгоритмы ранжирования - это методы фильтрации результатов пользовательских запросов в поисковых системах, которые определяют порядок отображения сайтов в результатах поисковой выдачи Google, «Яндекс» и других поисковых систем.

Поисковые системы учитывают следующие факторы:

Возраст домена;	Поведенческие факторы;
Ссылочная база;	Регулярность обновления контента;
Время загрузки;	Внутренние факторы – структура.

Процесс ранжирования выглядит следующим образом:

- Отправка поискового запроса.
- «Понимание» этого запроса системой.
- Вывод релевантных результатов поиска.
- Фильтрация результатов, исключение дублей и нерелевантных страниц.
- Сортировка согласно внутренним алгоритмам поисковика и вывод конечного списка сайтов на страницу результатов поиска.

Технология выполнения задания: Изучить основные термины и определения, основные характеристики процесса ранжирования, показать алгоритм ранжирования на примере выбранной поисковой системы, представить заключение.

4. ТЕМЫ ДЛЯ ЛАБОРАТОРНЫХ ЗАНЯТИЙ

№	темы лабораторных занятий	КОЛ-ВО ЧАСОВ
1.	Работа с каталогами, рубриками	2
2.	Роль электронных каталогов в автоматизированных библиотечных системах	4
3.	Структура и принципы работы национальной поисковой системы WWW.Uz.	2
4.	Содержимое HTML-документов: формы, фреймы, апплеты.	4
5.	Программирование поиска информации в массивах	4
6.	История и принципы работы поисковой системы Google	2
7.	Wiki - энциклопедия и поисковая система	4
8.	Принципы поиска векторной модели	4
9.	Ранжирование результатов поиска	4
Всего		30

4.1. ТЕМЫ, РЕКОМЕНДОВАННЫЕ ДЛЯ ПРОВЕДЕНИЯ ЛАБОРАТОРНЫХ ЗАНЯТИЙ ПО ПРЕДМЕТУ “ИНФОРМАЦИОННО – ПОИСКОВЫЕ СИСТЕМЫ”.

Лабораторная работа № 1. Работа с каталогами, рубриками.

Цель задания: Укрепление теоретических знаний, выработка умений и навыков при работе с библиотечными каталогами, рубриками.

Содержание задания: Библиотечный каталог - совокупность расположенных по определенным правилам библиографических записей на документы, раскрывающая состав и содержание фонда библиотеки или информационно- библиотечного

учреждения. Библиотечный каталог может функционировать в карточной или машиночитаемой форме, на микроносителях, а также в форме книжного издания.

Каталог - в широком смысле - список элементов данных, файлов, серверов, принтеров, магнитных накопителей и других объектов, составленный в порядке, облегчающем их нахождение. Каталоги упорядочиваются по алфавиту, датам, размеру содержащихся в них объектов и другим признакам.

Система библиотечных каталогов и картотек - часть справочно-библиографического аппарата библиотеки; совокупность планомерно организованных, взаимосвязанных и дополняющих друг друга библиографических каталогов и библиографических картотек.

Информационно-поисковая система (ИПС) - это упорядоченная совокупность документов (массивов документов) и информационных технологий, предназначенных для хранения и поиска информации - текстов (документов) или данных (фактов). Информационно - поисковыми системами являются любые определенным образом организованные хранилища информации. Причем информационно-поисковые системы могут быть и неавтоматизированными. Главное - это целевая функция: хранение и поиск информации. Важнейшим компонентом информационно-поисковых систем является информационно-поисковый язык. Информационно-поисковый язык - знаковая система, предназначенная для описания (путем индексирования) основного смыслового содержания текстов (документов) или их частей, а также для выражения смыслового содержания информационных запросов с целью реализации информационного поиска.

Технология выполнения задания: Дать характеристику библиотечным каталогам, указать по каким параметрам необходимо производить поиск информации в библиотечном каталоге, а также по предметным рубрикам каталога.

Лабораторная работа № 2.

Роль электронных каталогов в автоматизированных библиотечных системах.

Цель задания: Изучить теоретические сведения об электронных каталогах, определить роль каталогов в деятельности современных библиотечных учреждениях, дать характеристику электронному каталогу и возможности работы АБИС.

Содержание задания: Автоматизированные процессы обработки книг, книговыдачи, хранения и доступа к библиографическим базам расширили представления о работе с информацией. Стали очевидны все особенности и принципиальные отличия в работе между традиционными и автоматизированными схемами. При этом особое внимание уделялось электронному каталогу как основному элементу информационной среды библиотеки. В нем, как в зеркале, отражались все ошибки и просчеты в ее работе, он стал системообразующим центром практически для всех видов ее деятельности. Основным преимуществом автоматизации для этой сферы является предоставление читателям более полных, более достоверных информационных данных о состоянии библиотечного фонда, о возможностях эффективной и быстрой ориентации в весьма больших объемах информации. Функциональное назначение автоматизированных библиотечных систем заключается в автоматизации основных библиотечных процессов: комплектование (включая подписку на периодические издания), обработка литературных источников, печать стандартных выходных форм, таких, как каталожные карточки, книжные формуляры, инвентарные книги, бюллетени новых поступлений и пр. Кроме того, автоматизированные библиотечные системы являются идеальным средством автоматизации для малых и средних библиотек, позволяющим при самых минимальных затратах перейти к использованию новых информационных технологий.

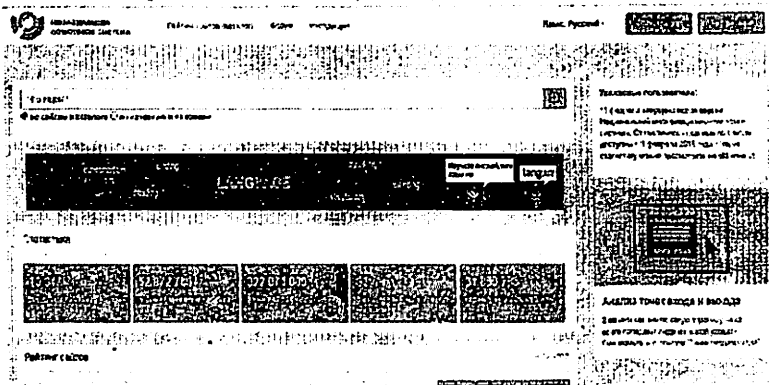
Технология выполнения задания: На примере действующей АБИС, дать характеристику и особенности электронных каталогов автоматизированных информационно – библиотечных систем.

Лабораторная работа № 3.

Структура и принципы работы национальной поисковой системы www.uz.

Цель задания: Изучение структуры и принципы работы **www.uz**.
Изучение основных характеристик функционирования поисковой системы.

Содержание задания:



Национальная информационно - поисковая система **WWW.UZ** — это механизм быстрого доступа к информации национального сегмента сети Интернет. Основными особенностями **WWW.UZ** системы являются многоязыковой поиск информации (узбекский, русский) и тесная интеграция с другими национальными информационными системами и базами данных.

Информационно-поисковая система **WWW.UZ** осуществляет поиск веб-сайта:

- По адресу
 - По содержимому
- В сервисы **WWW.UZ** входят разделы:
- Каталог
 - Топ-рейтинг
 - Сводная статистика,
 - Новости

Раздел «Каталог» **WWW.UZ** — это описания зарегистрированных и систематизированных по тематическим каталогам сайтов Республики Узбекистан, находящихся в открытом доступе в сети Интернет.

Каталог облегчает поиск ресурса, предлагая рубрику поиска

(Наука, Культура, Спорт, Общество, Бизнес, Новости, и т. д.). В дополнение к рубрикации по темам, пользователю предлагается классификация сайтов по типу содержащейся в них информации (Конференции и семинары Библиотека, Музеи, События, Форумы и чаты, Организации и т. д.). Каталог ежедневно пополняется новыми ресурсами благодаря активным пользователям поисковой системы и модераторам WWW.UZ. Добавить адрес нового сайта в каталог, может любой пользователь, зарегистрированный в системе ID.UZ WWW.UZ это открытый и доступный для просмотра топ-рейтинг и сводная статистика сайтов. Система WWW.UZ ведет открытую статистику сайтов Узбекистана. Этим определяется и оценивается популярность ресурса. На всех сайтах, участвующих в рейтинге, установлены счетчики системы WWW.UZ. Работу по расширению и модернизации WWW.UZ осуществляет Центр развития и внедрения компьютерных и информационных технологий UZINFOCOM.

Технология выполнения задания: Изучить рейтинг и статистику ресурсов WWW.UZ.

Лабораторная работа № 4.

Содержимое HTML-документов: формы, фреймы, апплеты.

Цель задания: Изучить содержимое HTML-документов.

Содержание задания: HTML - это теговый язык разметки документов, то есть любой документ на языке HTML представляет собой набор элементов, причем начало и конец каждого элемента обозначается специальными пометками, называемыми тегами. Регистр, в котором набрано имя тега, в HTML значения не имеет. Элементы могут быть *пустыми*, то есть не содержащими никакого текста и других данных (например, тег перевода строки
). В этом случае обычно не указывается закрывающий тег. Кроме того, элементы могут иметь *атрибуты*, определяющие какие-либо их свойства (например, размер шрифта для тега). Атрибуты указываются в открывающем теге. Вот пример части разметки HTML-документа:

<p>Текст между двумя тегами - открывающим и закрывающим.</p>

Здесь элемент содержит атрибут href.

А вот пример пустого элемента:

Каждый HTML-документ, отвечающий спецификации HTML какой-либо версии, обязан начинаться со строки декларации версии HTML <!DOCTYPE>, которая обычно выглядит примерно так:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
```

Если эта строка не указана, то добиться корректного отображения документа в браузере становится труднее.

Далее обозначается начало и конец документа тегами <html> и </html> соответственно. Внутри этих тегов должны находиться теги заголовка (<head></head>) и тела (<body></body>) документа.

Теги и их параметры нечувствительны к регистру. То есть и означают одно и то же. В последних версиях HTML практически у каждого тега огромное число необязательных параметров — обычно не меньше 15.

Технология выполнения задания: Изучить HTML - документы: формы, фреймы, апплеты и представить в работе.

Лабораторная работа № 5.

Программирование поиска информации в массивах.

Цель задания: Изучить процесс программирования поиска информации в массивах.

Содержание задания: Поиск - обработка некоторого множества данных с целью выявления подмножества данных, соответствующего критериям поиска.

Все алгоритмы поиска делятся на

- поиск в неупорядоченном множестве данных;
- поиск в упорядоченном множестве данных.

Упорядоченность - наличие отсортированного ключевого поля.
Сортировка - упорядочение (перестановка) элементов в подмножестве данных по какому-либо критерию. Чаще всего в качестве критерия используется некоторое числовое поле, называемое ключевым. Упорядочение элементов по ключевому полю предполагает, что ключевое поле каждого следующего элемента не больше предыдущего (*сортировка по убыванию*). Если ключевое поле каждого последующего элемента не меньше предыдущего, то говорят о *сортировке по возрастанию*.
Цель сортировки - облегчить последующий поиск элементов в отсортированном множестве при обработке данных. Все алгоритмы сортировки делятся на

- алгоритмы внутренней сортировки (сортировка массивов);
- алгоритмы внешней сортировки (сортировка файлов).

Сортировка массивов

Массивы обычно располагаются в оперативной памяти, для которой характерен быстрый произвольный доступ. Основным критерием, предъявляемым к алгоритмам сортировки массивов, является минимизация используемой оперативной памяти. Перестановки элементов нужно выполнять на том же месте оперативной памяти, где они находятся, и методы, которые пересылают элементы из массива А в массив В, не представляют интереса.

Методы сортировки массивов можно разбить на три класса:

- сортировка включениями;
- сортировка выбором;
- сортировка обменом.

Сортировка файлов

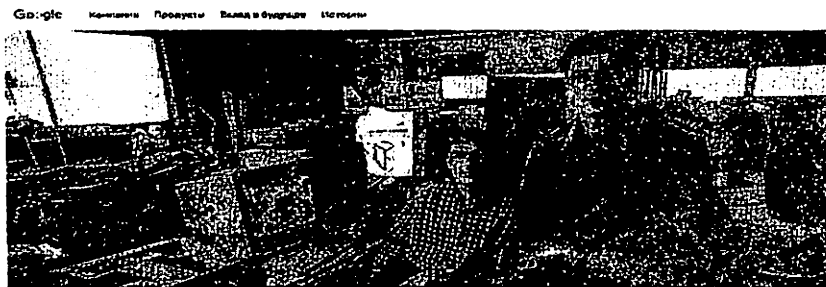
Файлы хранятся в более медленной, но более вместительной внешней памяти, на дисках. Однако алгоритмы сортировки массивов чаще всего неприменимы, если сортируемые данные расположены в структуре с последовательным доступом, которая характеризуется тем, что в каждый момент имеется непосредственный доступ к одному и только одному компоненту.

Технология выполнения задания: Провести анализ и показать алгоритм программирования поиска в информационных массивах.

Лабораторная работа № 6. История и принципы работы поисковой системы Google.

Цель задания: Изучить исторические сведения и принципы работы поисковой системы Google.

Содержание задания:



В 1996 году студенты Стэнфордского университета Ларри Пейдж и Сергей Брин начали работу над проектом BackRub, посвященным изучению ценности веб-страниц с точки зрения количества ссылающихся на них сайтов. В итоге разработали алгоритм PageRank. Анализ поисковой выдачи на основе этого алгоритма показал более высокую релевантность результатов, чем в случае с существующими алгоритмами (на тот момент приоритет в выдаче получали страницы, на которых чаще встречалась ключевая фраза). В 1998 году вышла статья «Анатомия системы крупномасштабного гипертекстового веб-поиска», в которой подробно описывался механизм расчета PageRank и работы новой поисковой системы Google. Поисковик, которым сегодня пользуется более 70 % населения планеты. Google получает информацию из разных источников. Вот некоторые из них:

- веб-страницы;
- пользовательский контент, например профиль вашей компании или информация, которую кто-нибудь добавил на Google Карты;
- отсканированные книги;
- открытые базы данных в интернете;

- множество иных источников.

Технология выполнения задания: Изучив сведения и принципы работы поисковой системы Google, показать отличительные характеристики и особенности системы.

Лабораторная работа № 7.

Wiki - энциклопедия и поисковая система.

Содержание задания: Изучить поисковую систему Wikipedia.

Ожидаемый результат: Wikipedia – это всемирная сетевая энциклопедия, статьи для которой пишутся и правятся самими пользователями: по своей инициативе, добровольно, то есть даром. Такой демократический подход в идеале должен исключить недобросовестные манипуляции со стороны «режима» и обеспечить всем гражданам свободный доступ к достоверной информации. Технологическая платформа для реализации интернет-ресурса, на котором пользователи сами могут создавать, публиковать и редактировать контент, классифицируется как «вики-движок» – специфическая разновидность CMS (система управления сайтом). В 1995 году программист из США Говард Каннингем (Howard Cunningham), будучи проездом в аэропорту солнечного Гонолулу, прокатился на местной достопримечательности – вики-вики шаттле. Это такой маленький автобус-челнок, перевозящий путешественников от терминала к терминалу.

Мистер Каннингем разработал первую вики-CMS и назвал ее WikiWikiWeb, что как бы намекало на простоту и скорость публикации и редактирования контента пользователями вики-проекта.

В дальнейшем сложилось три основных типа wiki-платформ.

1. Публичные вики со свободным доступом посетителей к редактированию и созданию контента.
2. Корпоративные – как понятно из названия, системы для обмена документами и совместной работы над проектами в удаленном формате в рамках одной организации.

3. Частные вики-CMS для реализации персональных дневников, путевых журналов, все то, что сегодня называется web-log или блог.

Специфика вики-движков в простоте и скорости работы с контентом. От автора публикаций не требуется знания программирования, разметки HTML и сложных веб-технологий.

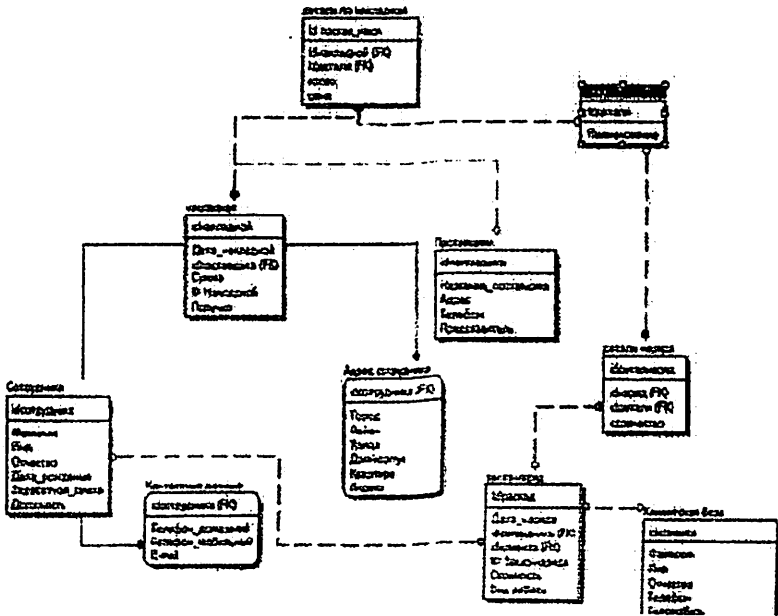
Технология выполнения задания: Изучить исторические сведения и особенности работы системы, дать полную характеристику.

Лабораторная работа № 8. Принципы поиска векторной модели.

Цель задания: Изучить принципы поиска векторной модели.

Содержание задания: Векторная модель (англ. vector space model) — в информационном поиске представление коллекции документов векторами из одного общего для всей коллекции векторного пространства. Векторная модель является основой для решения многих задач информационного поиска, как то: поиск документа по запросу, классификация документов, кластеризация документов.

Векторная модель представляла документы и поисковые запросы как векторы в N -мерном пространстве (N равнялось размерности словаря коллекции документов). Близость между документом и запросом измерялась на основе близости соответствующих векторов. Чем меньше был угол между векторами, тем выше считалось соответствие документа запросу. В качестве метрики близости использовалась косинусная мера, т.е. значение косинуса угла между соответствующими векторами.



Технология выполнения задания: Провести анализ принципов поиска векторной модели, представить в информационной модели и пояснить необходимые сведения.

**Лабораторная работа № 9.
Ранжирование результатов поиска.**

Цель задания: Дать полную характеристику определению и задачам ранжирования результатов поиска.

Содержание задания: Ранжирование – это сортировка сайтов поисковыми системами в результатах выдачи по соответствующему поисковому запросу. Алгоритм определяет, какой сайт лучше решает проблему пользователя и поднимает его выше конкурентов. Чем лучше ресурс отвечает на поставленные вопросы, тем больше у него шансов попасть в топ Яндекса или Google (именно они составляют 95% трафика с поисковых систем). Поисковики учитывают следующие факторы:

- Возраст домена;
- Ссылочная база;
- Время загрузки;
- Поведенческие факторы;
- Регулярность обновления контента;
- Внутренние факторы – структура, юзабилити.

Контент сайта, а это не только текст, но и графика, должен содержать слова и фразы из подобранного семантического ядра запросов. Страница сайта должна быть релевантна продвигаемой фразе. Контент должен быть уникальным и понятным для пользователя. Алгоритмы ранжирования понимают смысл текста, поэтому просто напечатать текст запросами уже не получится – сайт попадет под санкции поисковика за переспам, и выходить из него придется долго после устранения ошибок.

Алгоритмы ранжирования меняются и дорабатываются в ходе своего функционирования. Поэтому страницы, которые были в топе, могут терять позиции. Чтобы бороться с этим, веб-мастерам приходится регулярно дорабатывать сайты.

Технология выполнения задания: Изучить практический опыт ранжирования результатов поиска на примере поисковой системы.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ:

Основная литература	
1.	Christopher D. M., An Introduction to Information Retrieval/ Prabhakar R. Hinrich Schütze.- Cambridge University Press .- 2009., England
2.	Kowalski G. J., Information storage and retrieval ayatems: theory and implementation (second edition) / Maybury M. T.- Kluwer academic publishers.- 2009., USA
3.	Advances in Information Retrieval. 32nd European Conference on IR Research, ECIR 2010. Milton Keynes, UK, March 28-31, 2010 Proceedings.
4.	Terri Riyz, Kichik, Kayl Banerjiy. Raqamli kutubxonani barpo qilish. “Bu qanday qilinadi” yo‘riqnomasi -T.: Baqtria Pres, 2012.-272 b.;

5.	Internet – axborot qidiruvi. Xundibayev A.M. tahriri ostida. T.:FAN, 2006, 190 b.
6.	Гусев В.С. Google. Эффективный поиск. Краткое руководство.- Режим доступа: https://booksee.org/book/754618
7.	Е.Е.Gasanov.Teoriya slojnosti informatsionnogo poiska, M.: MGU, 2005, 144 s.
8.	Advances in Information Retrieval. Giambattista Amati Claudio Carpineto Giovanni Romano (Eds.) 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007
Дополнительная литература	
1.	Terri Riyz, Kichik, Kayl Banerjiy. Raqamli kutubxonani barpo qilish. “Bu qanday qilinadi” yo‘riqnomasi -T.: Baqtria Pres, 2012.-272 b.;
2.	Internet – axborot qidiruvi. /Xundibayev A.M. tahriri ostida. T.:FAN, 2006.-190 b.
Электронные ресурсы:	
1.	www.uz – национальная поисковая система
2.	www.Ziyonet.uz – образовательный портал
3.	www.rsl.ru – Государственная библиотека РФ
4.	www.rgu.ac.uk – Университет Роберта Гордона (Великобритания, Шотландия)
5.	www.press-service.uz – пресс-служба администрации Президента РУз
6.	www.natlib.uz – Национальная библиотека РУз
7.	www.cyberleninka.ru – электронный научный портал
8.	http://www.gpntb.ru –ГПНТБ (Государственная публичная научно-техническая библиотека) РФ
9.	https://yandex.uz/ - поисковая система
10.	https://www.google.com – поисковая система

Bichimi 60x84 1/16. Bosma tabog'i 2,5

Adadi 10. Buyurtma № 168

Al Xorazmiy nomidagi

Toshkent axborot texnologiyalari universiteti

«Taxririyl nashriyot» bo'limida chop etildi.

Toshkent sh. Amir Temur ko'chasi 108-uy.