

image not found or type unknown



Огромную важность приобретают вопросы поиска и систематизации информации.

1. Файл и папка как единицы хранения информации на компьютере

Информацию на компьютере организуют так, чтобы на поиск документа затратить не более минуты или еще меньше.

Представьте, что компьютерный жесткий диск (винчестер) – это огромное книгохранилище на миллион книг. Книга – это не более, чем компьютерный файл. Следовательно, наименьшая организационная единица хранения информации на компьютере – файл, или некая совокупность информации, записанная на каком-то носителе и имеющая название.

Файл с английского переводится, как подшивка или скоросшиватель. Он характеризуется тремя параметрами:

1. именем; имя в зависимости от операционной системы, можно давать или только латинскими буквами (в старых ОС), или любыми буквами вперемешку с цифрами и некоторыми знаками (дефис, например);
2. размером; в размере файла ограничений нет никаких, лишь бы места на жестком диске хватило;
3. расширением; расширение показывает, какого вида информация хранится в данном файле (текстовый документ, графическое изображение, медиа и так далее), причем, расширение файлу присваивает программа, в которой создается файл.

Например, файл такого вида: “инструкция.doc” – информация, созданная текстовым редактором Word. В этом примере “инструкция” – имя файла, а “.doc” – расширение.

На компьютере хранится огромное количество файлов и если их не сгруппировать (как книги по шкафам по жанрам), то немислимо, как что-то можно будет найти!

Вот для такого случая существуют папки (folder), в которые размещают и группируют файлы.

У каждой папки имеется имя, которое может быть любым, лишь бы можно было понять, что в ней находится. Расширений у папок нет, так как они предназначены только для хранения файлов. Причем, в одну папку можно вложить еще несколько папок (подпапок), что очень удобно при сортировке информации. Например, у нас на компьютере уже есть стандартная папка “Мои документы”, где находятся подпапки, например, “Документы кота Васьки”.

Такая структура вложенных папок называется древовидной структурой и помогает нам легко найти нужную информацию.

Самая большая единица в организации хранения информации на компьютере – это части жесткого диска, на которые который мы можем его разбить и эти части называются буквами латинского алфавита. начиная от “С” и до последней буквы “Z” (буквами “А” и “В” называют дисководы). Части жесткого диска можно назвать только одной буквой из алфавита, а не цифрами и не символами.

2. Файловая структура

Файловая система (ФС) - функциональная часть ОС, т.е. это порядок хранения и - организации файлов на диске

1. Виды файловой структуры

Виды файловой структуры: 1) Одноуровневая ФС - линейная последовательность имен файлов, используется для дисков с небольшим количеством файлов; 2) Многоуровневая иерархическая ФС - представляет собой древовидную структуру, служит для хранения сотни и тысячи файлов. Каталог (Папка) верхнего уровня содержит вложенные папки 1 уровня, которые могут содержать папки 2 уровня и тд.

Для хранения информации каждый диск разбивается на 2 области: 1) каталог (directory) или папка - содержит названия файлов и указание на начало их размещения на диске; 2) область хранения файлов, содержит текст.

Чтобы найти файл надо знать: 1) имя файла; 2) где храниться файл

например:

C:\GAMES\CHESS\zena.exe

Имя диска – С

папка 1уровня –GAMES

папка 2уровня –CHESS; имя файла - zena.exe

В процессе работы наиболее часто над файлами производят следующие операции: копирование, перемещение, удаление, переименование.

3. Создаем свою файловую систему в ОС Windows на .Net

Существует великое множество файловых систем. Это и файловые системы для носителей информации (FAT*, NTFS, ext* и т.д.), и сетевые файловые системы (NFS, CIFS и т.д.), и виртуальные файловые системы, и великое множество других. А появлялась ли у тебя, %habrauser%, потребность в своей, еще несуществующей файловой системе? О том, как ее сделать для ОС Windows на managed-коде (.net), и пойдет речь.

В общем случае, для создания своей файловой системы (далее ФС), необходимо написать драйвер этой ФС и зарегистрировать его в ОС. Т.к. драйвер взаимодействует с ядром ОС, то его создание на managed-коде является делом нетривиальным, а производительность такого драйвера будет на невысоком уровне. В связи с этим, целесообразнее иметь драйвер ФС, написанный на native-коде и некий промежуточный слой между этим драйвером и managed-кодом. Такой драйвер существует в проекте Dokan, равно как и набор классов DokanNet для взаимодействия с этим драйвером в managed-коде. Оба проекта распространяются под лицензией GPLv3.

1. Попробуем разобраться, как это работает

Dokan — это драйвер, работающий на уровне ядра Windows, с которым могут взаимодействовать любые приложения из userspace-а. Драйвер существует как для 32-ух разрядных систем, так и для 64-ух разрядных. Драйвер не несет в себе реализацию какой-либо ФС, а лишь выступает в роли прокси, пропуская через себя все операции ввода-вывода (IO) из ядра ОС в userspace. Реализация функционала ФС лежит на наших плечах, мы создаем базовые функции ФС, такие как открытие/закрытие файла, чтение из файла, запись в файл и т.д., которые будут вызываться драйвером Dokan при возникновении в системе соответствующих событий. В момент регистрации нашей ФС в драйвере Dokan, мы указываем некоторые параметры работы нашей ФС (кол-во потоков-обработчиков операций ввода/вывода, точка монтирования нашей ФС (Dokan поддерживает монтирование в качестве съемного или сетевого диска и только в корень ФС), автоматическое

размонтирование нашей ФС, в случае ошибок в работе, и некоторые другие параметры). После этого в ОС появляется новый диск в корне ФС, с которым любые приложения и сама ОС взаимодействует, как и с обычным диском. Какие операции разрешено производить в данной ФС, зависит только от разработчика этой ФС, то есть от нас.

DokanNet – это обертка над драйвером Dokan. Позволяет разрабатывать свою ФС на managed-коде. DokanNet описывает интерфейс DokanOperations, который необходимо реализовать для регистрации нашей ФС в драйвере Dokan. Автор DokanNet вместе с исходниками самого проекта положил 2 примера работы с данной оберткой:

- DokanNetMirror — ФС-зеркало уже существующего каталога
- RegistryFS — ФС, представляющая структуру системного реестра Windows в директории и файлы.

2. Коды

Коды примеров легко читаемы и позволяют ознакомиться с возможностями, предоставляемые драйвером Dokan.

Реализация интерфейса DokanOperations сводится к реализации следующих функций:

- CreateFile — создание файлов/директорий
- OpenDirectory — открытие директории
- CreateDirectory — создание директории
- Cleanup — удаление файла/пустой директории
- CloseFile — закрытие файлового дескриптора
- ReadFile — чтение куска файла указанной длины с указанным смещением
- WriteFile — запись данных в файл с указанным смещением
- FlushFileBuffers — очистка буферов (кэша) файла
- GetFileInformation — получение информации о размере, атрибутах, времени создания/последнего доступа/модификации файла/директории

- FindFiles — получение списка файлов/директорий в указанной директории
- SetFileAttributes — установка атрибутов файла/директории
- SetFileTime — установка времени создания/последнего доступа/модификации файла/директории • DeleteFile — пометка файла на удаление (удаление проводится в Cleanup)
- DeleteDirectory — пометка директории на удаление (удаление проводится в Cleanup)
- MoveFile — перемещение/переименование файла/директории
- SetEndOfFile — установка размера файла (используется при создании пустого файла определенной длины)
- SetAllocationSize — автор не указал для чего необходима данная функция, на практике передача управления в нее не замечена
- LockFile — блокировка файла в single-доступе
- UnlockFile — снятие блокировки
- GetDiskFreeSpace — получение информации о кол-ве доступного/общего/свободного места в ФС
- Unmount — размонтирование/отключение ФС

Как мы видим, функционал, предоставляемый драйвером Dokan и оберткой DokanNet очень богат. На базе данного проекта разработана ФС SSHFS. Каждый может разработать ФС, отвечающий своим требованиям.

4. Поисковые системы

Поисковая система — программно-аппаратный комплекс с веб-интерфейсом, предоставляющий возможность поиска информации в интернете. Под поисковой системой обычно подразумевается сайт, на котором размещён интерфейс (фронт-энд) системы. Программной частью поисковой системы является поисковая машина (поисковый движок) — комплекс программ, обеспечивающий функциональность поисковой системы и обычно являющийся коммерческой тайной компании-разработчика поисковой системы.

Большинство поисковых систем ищут информацию на сайтах Всемирной паутины, но существуют также системы, способные искать файлы на FTP-серверах, товары в интернет-магазинах, а также информацию в группах новостей Usenet.

Улучшение поиска — это одна из приоритетных задач современного интернета (см. про основные проблемы в работе поисковых систем в статье Глубокая паутина).

1. Статистика использования поисковых систем По данным компании Net Applications, в ноябре 2011 года использование поисковых систем распределялось следующим образом:

- Google — 83,87 %;
- Yahoo! — 6,20 %;
- Baidu — 4,22 %;
- Bing — 3,69 %;
- Yandex — 1,7 %;
- Ask — 0,57 %;
- AOL — 0,36 %.

2. Популярные поисковые системы

Согласно данным LiveInternet в 2012 году об охвате русскоязычных поисковых запросов:

- Всеязычные:

- o Google (24,9 %)

- o Bing (0,8 %)

- o Yahoo! (0,2 %) и принадлежащие этой компании поисковые машины:

- Inktomi

- AltaVista

- Alltheweb

- Англоязычные и международные:

- o AskJeeves (механизм Teoma)

- Русскоязычные — большинство «русскоязычных» поисковых систем индексируют и ищут тексты на многих языках — украинском, белорусском, английском, татарском и др. Отличаются же они от «всеязычных» систем, индексирующих все документы подряд, тем, что в основном индексируют ресурсы, расположенные в доменных зонах, где доминирует русский язык или другими способами ограничивают своих роботов русскоязычными сайтами.

- o Яндекс (61,3 %)

- o Mail.ru (8,5 %)

- o Рамблер (1,9 %)

- o Нигма (0,3 %)

Некоторые из поисковых систем используют внешние алгоритмы поиска. Так, Qip.ru использует поисковый механизм Яндекса, а Nigma сочетает в себе как свой алгоритм, так и сборную выдачу от других поисковиков.

3. Необычные поисковые системы

- DuckDuckGo — гибридная поисковая система с политикой конфиденциальности пользователя и его поисковых запросов.
- Koogler (с его помощью ортодоксальные иудеи могли найти контент, признанный раввинами удовлетворяющим религиозным требованиям).
- Yauba (индийский поиск «для параноиков» — в отличие от привычных поисковиков, за пользователями не следят, а все записи о его действиях удаляются с сервера).
- TinEye — поисковая система, специализирующаяся на поиске изображений в Интернете.
- Генон — поисковая система, собирающая и создающая контент у себя на сайте.
- ImHalal — первый исламский поисковик. Новый поисковик может отличить разрешённые для мусульман (халяльные) результаты поиска от запретных

(харамных). Был открыт в 2009 году. Ориентировочно с начала сентября 2011 прекратил функционирование в связи с недостатком финансирования.

5. Поиск информации в Интернете

Сеть Интернет растет очень быстрыми темпами, поэтому найти нужную информацию среди сотен миллиардов Web-страниц и сотен миллионов файлов становится все сложнее. Для поиска информации используются специальные поисковые системы, которые содержат постоянно обновляемую информацию о местонахождении Web-страниц и файлов на сотнях миллионов серверов Интернета.

Поисковые системы содержат тематически сгруппированную информацию об информационных ресурсах Всемирной паутины в базах данных. Специальные программы-роботы периодически "обходят" Web-серверы Интернета, читают все встречающиеся документы, выделяют в них ключевые слова и заносят в базу данных Интернет-адреса документов.

Большинство поисковых систем разрешают автору Web-сайта самому внести информацию в базу данных, заполнив регистрационную анкету. В процессе заполнения анкеты разработчик сайта вносит адрес сайта, его название, краткое описание содержания сайта, а также ключевые слова, по которым легче всего будет найти сайт.

1. Поиск по ключевым словам

Поиск документа в базе данных поисковой системы осуществляется с помощью введения запросов в поле поиска.

Запрос должен содержать одно или несколько ключевых слов, которые являются главными для этого документа. Например, для поиска самих систем поиска в Интернете можно в поле поиска ввести ключевые слова "российская система поиска информации Интернет".

Через некоторое время после отправки запроса поисковая система вернет список Интернет-адресов документов, в которых были найдены заданные ключевые слова. Для просмотра этого документа в браузере достаточно активизировать указывающую на него ссылку.

Если ключевые слова были выбраны неудачно, то список адресов документов может быть слишком большим (может содержать десятки и даже сотни тысяч

ссылок). Для того чтобы уменьшить список, можно в поле поиска ввести дополнительные ключевые слова или воспользоваться каталогом поисковой системы.

Одной из наиболее полных и мощных поисковых систем является Google (www.google.ru), в базе данных которой хранятся 8 миллиардов Web-страниц и каждый месяц программы-роботы заносят в нее 5 миллионов новых страниц. В Рунете (российской части Интернета) обширные базы данных, содержащие по 200 миллионов документов, имеют поисковые системы Yandex (www.yandex.ru) и Rambler (www.rambler.ru).

2. Поиск в иерархической системе каталогов

В базе данных поисковой системы Web-сайты группируются в иерархические тематические каталоги, которые являются аналогами тематического каталога в библиотеке.

Тематические разделы верхнего уровня, например: Интернет, Компьютеры, Наука и образование и т. д., содержат вложенные каталоги. Например, каталог Интернет может содержать подкаталоги Поиск, Почта и др.

Поиск информации в каталоге сводится к выбору определенного каталога, после чего пользователю будет представлен список ссылок на Интернет-адреса наиболее посещаемых и содержательных Web-сайтов. Каждая ссылка обычно аннотирована, т. е. содержит короткий комментарий к содержанию документа.

Наиболее полный многоуровневый иерархический тематический каталог русскоязычных Интернет-ресурсов имеет поисковая система Апорт (www.aport.ru). Каталог содержит подробную аннотацию содержания Web-сайтов и указание на их географическое положение.

3. Поиск файлов

Для поиска файлов на серверах файловых архивов существуют специализированные поисковые системы, в том числе поисковая система FileSearch (www.filesearch.ru). Для поиска файла необходимо ввести имя файла в поле поиска, и поисковая система выдаст Интернет-адреса серверов файловых архивов, на которых хранится файл с заданным именем.

Поиск информации в русскоязычной части Интернета с помощью наиболее поисковых систем: Google, Rambler, Апорт и файловой поисковой системы Research

можно производить с использованием интегрированной поисковой системы Google.ru. Для этого достаточно ввести ключевые слова в строку поиска, с помощью переключателей установить тип необходимой информации и щелкнуть по кнопке с названием поисковой системы Google.ru. Для этого достаточно ввести ключевые слова в строку поиска, с помощью переключателей установить тип необходимой информации и щелкнуть по кнопке с названием поисковой системы.

Вывод

Задачи поставленные в работе выполнены. Исследовали понятия файла и папки как единицы хранения информации на компьютере, рассмотрена организация файловой структуры как способ систематизации информации на локальном компьютере, изучено создание файловой структуры для организации личной информационной среды на персональном компьютере, раскрыты поисковые системы, и поиск информации в Интернете.

Итак, можно сделать вывод, что единица хранения информации на компьютере это файл. Файл может быть с разным именем, размером и расширением. Папки существуют для размещения и группирования файлов.

Функциональной частью операционной системы является файловая структура она служит для порядка хранения и организации файлов на диске.

Существует великое множество файловых систем. И при создании необходимо учитывать виды файловой структуры.

Поисковая система — программно-аппаратный комплекс с веб-интерфейсом, предоставляющий возможность поиска информации в интернете.

Для поиска информации используются специальные поисковые системы, которые содержат постоянно обновляемую информацию о местонахождении Web-страниц и файлов на сотнях миллионов серверов Интернета.