

О. А. Соседко

ОБ ОСОБЕННОСТЯХ РАЗРАБОТКИ И ПРОВЕДЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ ПО ПОИСКУ ФАЙЛОВ В ОПЕРАЦИОННОЙ СИСТЕМЕ ALT LINUX

Управление файлами и каталогами одна из важнейших компетенций учителя информатики. Несмотря на разнообразие операционных систем есть важные особенности при поиске файлов. В настоящий момент образовательные учреждения переходят на операционную систему Linux. В подавляющем числе случаев необходимо найти файл по конкретному имени или маске. Но чаще пользователь запоминает «тему» файла, т.е. какая информация содержалась в файле. Для хранения и обработки информации чаще используются текстовые редакторы. Каждый из которых сохраняет как в своем собственном формате, так и в наиболее распространённых, таких как RTF, HTML, PDF. Итак, чтобы ускорить поиск, системе надо указать три условия:

1. имя или маска имени
2. формат файла
3. часть текста, содержащегося в искомом файле.

Есть ещё фактор времени, т.е. когда и в какое время создан или изменен файл. Он используется при поиске редко, т. к. 1) пользователь помнит лишь период создания, а чем больше период, тем больше файлов найдет система и придётся вручную отфильтровывать лишнее 2) время файлу задаётся компьютером, может не совпадать с реальным, может меняться вирусом или другим программным способом.

Содержимое может храниться лишь в одной из кодировок, среди которых шесть кириллических: четыре старых MSDOS, Windows, KOI8-R, ISO, и две современных UNICODE и UTF8. Вот тут и начинаются первые проблемы, т.к. операционная система не предлагает выбор кодировки при поиске, а выбирается та, которая заложена разработчиками и, как правило, не описывается в документации или контекстной справке. Чтобы протестировать систему на поддержку поиска по кодировкам, необходимо подготовить заранее несколько файлов в разных форматах и кодировках. Для этого удобнее воспользоваться популярным редактором MS Word версии 2003.

Используя настройку Word для сохранения в HTML формате с разными кодировками (Меню «Сервис», вкладка «Общие», кнопка «параметры веб-документа», вкладка «Кодировка») сохранить фразу *«Кодировки DOS, KOI8, ISO, Windows содержат только 256 байт информации, что слишком недостаточно для одновременного хранения десятков разнообразных языков.»* из Word в кодировке по умолчанию в разных форматах: doc, rtf, htm (dos), htm (windows), htm (iso), htm (koi8-r), htm (utf-8).

Еще можно сохранить в простом текстовом формате. Меню «Файл - Сохранить как..», формат «Обычный текст», далее редактор выдаст запрос на кодировку, выбрать необходимую и для каждой из кириллической кодировки задать свое имя файла, иначе при совпадении имени запишется только новый файл, а нам для тестирования нужны все шесть.

Таким образом получили уже 12 файлов. Осталось сохранить в форматах DOC и RTF. В этих форматах кодировку поменять нельзя, т.к. заложена только разработчиками. Как правило используется Unicode, но возможен альтернативный вариант. Ms Word позволяет получить 14 файлов разных форматов и кодировок. Скопируем эти 14 файлов в один каталог и воспользуемся поисковыми возможностями Alt Linux 5.0. Заметим, что в полученных файлах есть содержимое на латинском и русском языках, а также есть число.

В среде Linux есть текстовый редактор Writer, который также имеет свой формат хранения файлов - ODT. Таким образом можно получить пятнадцатый файл для нашего эксперимента.

Для поиска файлов необходимо нажать Ctrl+F или меню «Сервис» проводника Dolphin. В результате система Linux найдет только пять из пятнадцати файлов, т.е. эффективность поиска кириллической информации $5/15 \approx 33\%$. Значит, поддерживаются кодировки Unicode и UTF8 в форматах TXT и HTM, а также собственный формат редактора Writer.

Вывод. С одной стороны, алгоритм поиска современен, т. к. поддерживаются две основные разновидности юникода, с другой стороны «старые» кодировки до сих пор часто используются пользователями не столько по незнанию, а сколько из-за настроек большинства программных продуктов при сохранении файлов.

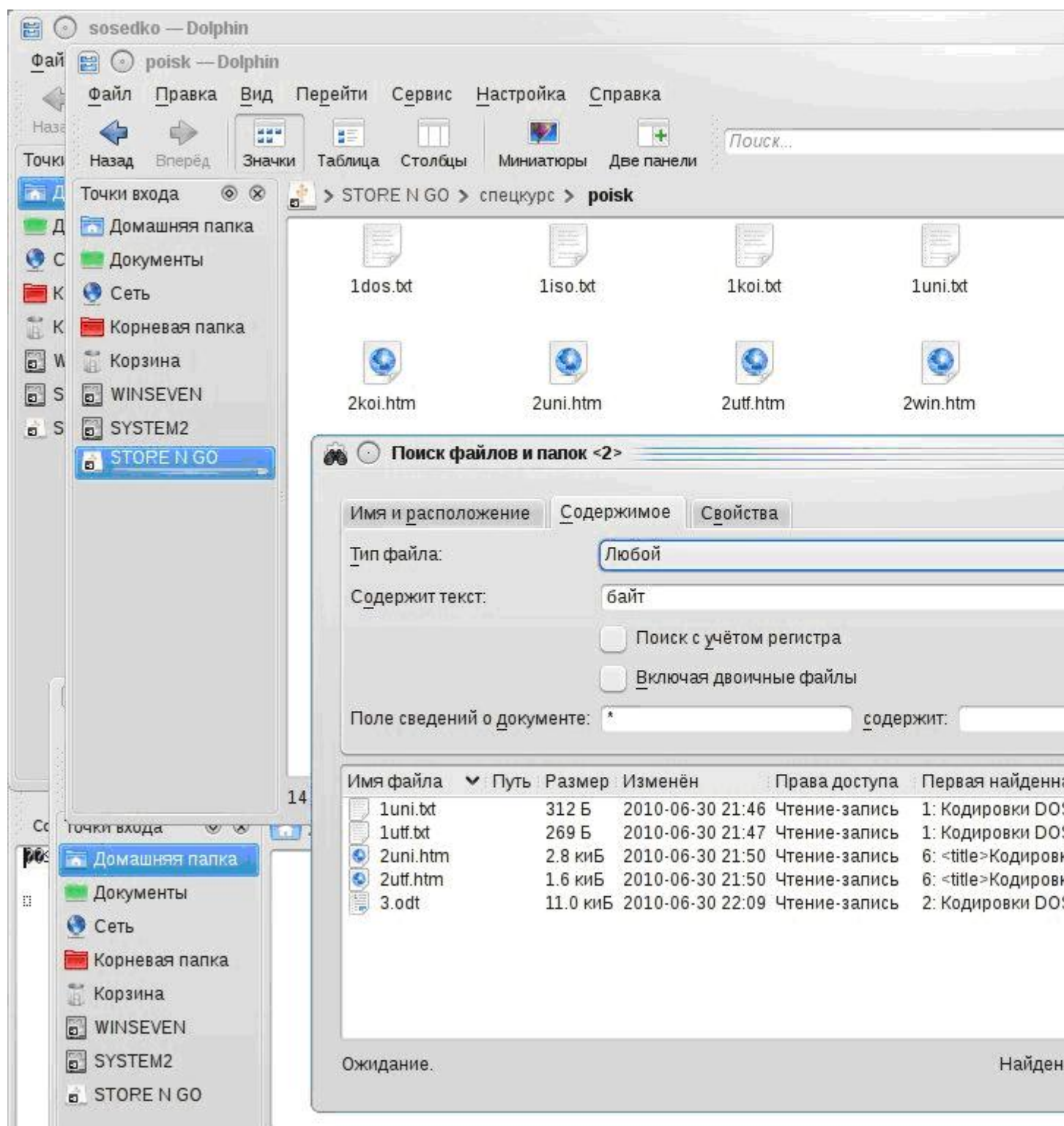


Рис. 1. Окно результата поиска

По аналогии в качестве самостоятельных заданий студентам можно предложить сделать анализ поиска файлов по содержимому в системах семейства Windows. Для отчёта можно использовать следующую таблицу:

файл	<u>Windows</u> : поиск по англ. слову	Windows: поиск по русск. слову
PRIMER_дос.TXT		
PRIMER_вин.TXT		
PRIMER_исо.TXT		
PRIMER_кои.TXT		
PRIMER_юни.TXT		
PRIMER_втф.TXT		
PRIMER_дос.HTM		
PRIMER_вин.HTM		
PRIMER_исо.HTM		
PRIMER_кои.HTM		
PRIMER_юни.HTM		
PRIMER_втф.HTM		
PRIMER.RTF		
PRIMER.DOC		
PRIMER.ODT		
% поиска		

Сделать выводы. Сравнить с результатами в Linux.