Extract Contents

The entire help file consists of the topics below.   It is suggested you read all the topics.

Descriptions
<u>What is Extract</u> and what about distributing it?
<u>Input Datafile</u>
<u>Comma Delimited Text File</u>
Using Extract
5 easy steps to extract data
    1 <u>Loading Datafiles into Extract</u>
    2 <u>Selecting Search Text (Labels)</u>
    3 <u>Selecting Relative Data</u>
    4 <u>Adding a Field</u>
    5 <u>Running the Extraction</u>
Automated Extractions from the Command Line
<u>Command Line Options</u>

What is Extract

Extract is a Windows 3.1 program designed to extract data from irregularly formatted <u>data files</u> into <u>comma delimited ASCII files</u>, easily readable by Dbase, Excel and other programs.

Extract uses an extract-by-example interface.   You define to the program how to locate all the fields within just one record and Extract will automatically read all records, based on your one-record-only definition.

Distribution

Extract is a freely distributable and may not be sold or distributed with any other package. Although I've attempted to make Extract as bug-free as possible it is provided AS IS.

Datafile Format
Extract will read any normal ASCII text file.   The data in the text file must have at least one piece of text constant to each record for Extract to work properly.

The sample below is a typical text file that, because of its format, is unreadable by programs that require <u>delimited text files</u>.   Notice that there is a label associated with each piece of data (name, last name, address, city, postal code and age) and that the format is the same for each record.   So to perform an extraction on this file select the label as the <u>search text</u> and the actual data as the <u>relative data</u>.   Remember that Extract uses an extract-by-example interface so you only need to select the search text and relative data for each field in the FIRST RECORD.

Sample Datafile 1:
```
record 1
first name: Bob
last name: Smith
address: 123 East Street          city: Williams Lake
postal code: V2G 1R8
age: 32

record 2
first name: Jane
last name: Jones
address: 321 South Ave            city: Williams Lake
postal code: V2G 1R8
age: 26
```

Sample Datafile 1 above would be ideal input for Extract because each field has a label but with a little creativity you can usually find one piece of data that is constant for each record.   The datafile below doesn't have labels for each field (or even each record) but the 'Rain' field does have a decimal point in the same place in each record that could be used for the search text.   It is easy to use the decimal point for the search text and use a different relative data offset for each field.

Sample Datafile 2:

				Wind	Wind		
Date	Rain	Temp	Humidity	Speed	Dir	FFMC	ISI
14	0.8	5	98	6	3	4	0
15	0.5	6	98	2	4	5	0
16	0.3	7	99	8	6	3	0
17	0.0	6	98	4	8	2	0
18	0.3	4	99	3	3	3	0

Comma Delimited Text (ASCII) Files

Extract converts irregularly formatted <u>datafiles</u> into delimited ASCII files. Delimited text files are importable into many programs (Dbase, Excel, Access, etc) and are the output from Extract.

Sample Comma Delimited File:

```
"Bob", "Smith", "123 East Street", "Williams Lake", "V2G 1R8", 32
"Jane", "Jones", "321 South Street", "Williams Lake", "V2G 1R8", 26
```

Loading Example Text

To start using the program select the *File|Load Example Text* menu.   This will allow you to load in your <u>datafile</u> and select the data you wish to extract from it.

If the datafile is larger than 20,000 characters you will be asked whether it is OK to truncate the file.   Answer YES.   This simply means that you will not be able to view the entire <u>datafile</u> while selecting the fields to extract; when you run the actual extraction the entire file WILL be processed no matter what size.

Next quick learn step: <u>Selecting Search Text (Labels)</u>

Set Search Text (Labels)
What is search text?   This is the text Extract will search for to tell it where the data you wish to extract is.

Method 1:
Use the mouse (or keyboard) to select the text that is associated with a field and select the *Set|Search Text* menu.
Method 2:
Use the mouse (or keyboard) to select the text that is associated with a field and select the *View|Field Info* menu.   This will open the View Field Info dialog box.   Click on the Set button in the Search Text area of the View Field Info dialog box.

Next quick learn step: Selecting Relative Data

Selecting Relative Data
What is relative data?   This is the data you wish to extract, specified as an offset from the search text you have already selected (you have selected your <u>search text</u> haven't you?).

Method 1:
Use the mouse (or keyboard) to select the actual data and select the *Set|Relative Data* menu.

Method 2:
Use the mouse (or keyboard) to select the actual data and select the *View|Field Info* menu.   This will open the View Field Info dialog box.   Click on the Set button in the Relative Data area of the View Field Info dialog box.

Next quick learn step: <u>Adding a Field</u>

Adding a Field
To add a new field skip to one past the last field.

Method 1:
Select the *View|Next Field* until a message box appears with the question "Create new field?".
Click on Yes.
Method 2:
Select the *View|Field Info* menu.   This will open the View Field Info dialog box.   Click on the
Next button until a message box appears with the question "Create new field?". Click on Yes.

Next quick learn step:

Running the Extraction

Once you have entered and created the fields necessary for your extraction click on the *Run* menu.   This will bring up a dialog box requesting the <u>datafile</u> to run the extraction on.   Enter the datafile name and you are presented with another dialog box, this one requests the name of the <u>delimited text file</u> to save to.   After selecting the <u>delimited text file</u> all the Extract windows disappear and are replaced by a counter window.   The counter window reports how far the extraction has completed and disappears when Extract is finished processing.

Running Extract From the Command Line

The format for starting EXTRACT from the command line is:

```
EXTRACT.EXE FIELD.FLD INPUT.TXT OUTPUT.TXT
```

Where:
    FIELD.FLD   = A pre-saved field definition file.
    INPUT.TXT   = The <u>datafile</u> to run the extraction on.
    OUTPUT.TXT = The <u>delimited text file</u> to append to.

Note: If OUTPUT.TXT already exists it will be appended to and can not be overwritten from the command line.