

Contents

If WebParse is not compatible with your E-mail system, or if you encounter problems, please contact Informatik Inc. at webparse@informatik.com. In most cases we are able to make the necessary changes within 24 hours, and make the new version available via our web page. We can also, for a limited time, prepare the parsing specification file for you at no charge.

[About](#)

[Introduction to HTML](#)

[Specifications](#)

[Parse a File](#)

[View and Print Extracted Data](#)

Copyright, Registration, Technical Support

This system was developed by Informatik Inc, P.O. Box 868, Devon, Pennsylvania 19333. For information on program modifications, systems integrations and custom applications, please contact Informatik Inc.

This is a copyright-protected shareware product. If, after review and evaluation, you like the software, please register and pay the registration fee. Upon registration and payment of the registration fee, Informatik Inc will send you the latest master diskette, a printed User Guide and a registration certificate. Registered users will receive lifetime technical support and are entitled to free software upgrades during first year. Shareware is based on trust, and we hope that all users will register.

To register

From the About menu, choose Registration, Copyright and follow the instructions.

To obtain technical support

From the Help menu, choose Technical Support and follow the instructions.

© 1995 Informatik Inc
All rights reserved

Introduction to WebParse

There has been an explosive growth in the number of Internet web pages. And, as the web technology advances and the public acceptance expands, more web pages will become interactive. A basic form of interactivity is the web Survey Form, where the data is entered on a web form and transmitted to the owner of the web page. The easiest way to transmit the data from such a web form back to the owner of the web page is by using the Internet's built-in e-mail capabilities. Thus, the web survey form is filled out, the Submit button is pressed and virtually instantly, the data is available in the e-mail in-box file.

Many web survey forms consist of questionnaires, where the respondent chooses suitable replies from selection lists, check boxes, buttons or types the answer in text boxes. Back at the recipient's office the incoming mail needs to be analyzed and keyed into a database.

WebParse simplifies this process. Essentially, WebParse scans thru all the incoming e-mail message file, isolates the messages containing data submitted from web reply forms, parses the text, extracts the relevant data and creates a comma-delimited ASCII text file that can be imported into any database. The system allows you to specify the beginning point of the scan, in case the e-mail file contains messages that have already been scanned earlier.

WebParse can also display the extracted data on the computer screen in raw or tabulated format, or it can print it out.

Some surveys allow multiple selections of answers. WebPen will combine multiple selections into one comma-delimited ASCII field.

When parsing the data, WebParse relies on the uniqueness of the first extract field identifier tag. Therefore, it is of utmost importance that the first field identifier tag be unique and distinctive. Please read the Specification section.

Parse a File

Before you can parse a file, you must create a specification file. Please read the [Specifications](#) topic.

1. From the File menu, choose Parse. The Parse window is displayed.
2. From the Specification File list, select the file. The first specification file is shown as a default selection.
3. In the 'First Record String Match' enter a unique text string that marks the beginning of the scan. This field is left blank if you scan the entire e-mail file. If you do enter a 'First Record String Match', make sure that the spelling is absolutely correct, with correct upper and lower case letters. It is suggested that you paste the string from the E-mail file. The match string must be in the header of the e-mail message, NOT in the data section of the message.
4. In the Destination File box, specify the file name that will receive the parsed data. The new data is ALWAYS APPENDED TO AN EXISTING FILE. You may want to delete the destination file first. The name is limited to eight characters. No spaces, no extension. You may type in a file number or select a generic file from the list.
5. Choose Extract.
6. The system parses the file, extracts the relevant data and creates a comma-delimited ASCII text file. Comma-delimited ASCII is a standard format of text in which each field is encapsulated in quotes and separated by a comma. WebParse treats all fields, including numeric values, as alpha fields. Multiple answers (if permitted) are combined into one field.
7. If the destination file exists, the parsed data will be appended, but the system will inform you first. To delete the current destination file, choose Delete Extraction file from the File menu, or press Shift+Del.

Specifications

Before you can run the first extracts, you must create a specification file. The specification file holds the parsing instructions, such as fields names, match strings, etc. You may need to create separate specification files for various applications. The specification files are kept in the application's main directory and they have a .PAR extension name. It is important that the specification files be created with great care; the data must be accurate and the text relating to match strings is case sensitive, i.e. upper-case characters and lower-case characters are treated as different characters.

When parsing the data, WebParse relies on the uniqueness of the first extract field identifier tag. Therefore, it is of utmost importance that the first field identifier tag be unique and distinctive. A field identifier of 'name=' will not always work because it may not be unique within an e-mail message. It is suggested that the first name have some unique characters, such as ##name=.

When reviewing the e-mail text, you will notice some strange characters and symbols. There is an simple explanation: E-mail converts the characters outside the normal alpha-numeric range (A-Z and 0-9) to hexadecimal values. However, WebParse will convert the hexadecimal values back to readable characters.

To create a specification file

1. Using Windows Write (or Notepad), open the incoming E-mail file. For Eudrora, the file is called IN.MBX and is located in the EMAIL subdirectory, possible in the C:\Windows\Internet directory.
2. In the E-mail file, locate a typical web survey reply and highlight the reply section. The reply section is located immediately below the header information of each message and it is a tightly packed string of (mainly obscure) data. The section is typically 2-3 lines long but may be longer or shorter. **Copy the complete section into your clipboard.**
3. Return to the WebParse system and select Specification/Setup from the File menu, or click on the hammer icon in the toolbar. The Specification window appears.
4. Click on the Browse button and select the directory and file of the E-mail in-box (the same file you viewed earlier).
5. Click on Automatic Spec Generation. The system now analyses the survey data and creates the specification for parsing. It is important that the fields contain the hexadecimal values, exactly as shown in the E-mail file.
6. Optionally, you may re-arrange the sequence of the extract fields for the comma-delimited ASCII file and/or add constant text or blank fields.

To specify a constant value, enter three equal signs followed by the constant text (No spaces between the equal signs and the text!). For example ===Survey will insert the word 'Survey' in each record.

To specify the system date, enter ===Date (three equal signs followed by the word 'Date'. The date can be formatted, using the standard formatting codes. For example, to enter the current system date formatted as 12/31/95, you would enter ===Date mm/dd/yy. To specify the same date without slashes, you would enter ===Date mmdyy.

Finally, to specify a running counter, starting with 1, you enter ==Counter.

7. In the Filler Character cell, enter the character that your E-mail system uses as 'filler'. Eudora, for example, uses plus signs (+) to fill blank spaces in the answer section.
8. The end tag specifies the symbol that is used to separate the reply fields. Eudora, for example, uses the ampersand character (&).
9. From the File menu, choose Save As. Specify the file name. The extension name must be .PAR. If you save the file without extension name, the system will assign the correct extension name.

There are up to 24 fields for field identifiers. If you need more identifiers, you should add the remaining fields with a text editor.

To amend a specification file

1. From the File menu, choose Open and select the file.
2. Make the necessary corrections, as explained above.
3. From the File menu, choose Save.

For improved accuracy of input, it is recommended that you copy and paste the entries from the E-mail in-box. This method will prevent all typing errors. Remember, the text is case sensitive.

View and Print the Extracted Data

The extracted data can be view online or printed.

To view the extracted data

1. From the File menu, choose View. You can view the 'raw' file or a tabulated format. The raw format is displayed with the Windows Notepad.
2. If you are in the main menu, select the file that you wish to view, otherwise the current file will be displayed.
3. Press OK.
4. If necessary, adjust the width of the columns, by dragging the column separator in the first row. If a field text is multi-line, you may see the complete text by dragging downward the row separator in the first column.
5. Close the screen, by clicking the Control in the upper-left corner of the screen.

To print the extracted data

1. From the File menu, choose Print.
2. If you are in the main menu, select the file that you wish to print; otherwise the current file will be printed. Press OK.
3. In the Print dialog box, specify the number of copies (and the printer). Press OK.

The file is printed in a vertical layout (one field per line). If you wish to print the 'raw' extract file, you should view the file, as explained above, and print it from the Notepad print menu.

