Contents

If WebParse is not compatible with your E-mail system, or if you encounter problems or have questions, please contact Informatik Inc. at webparse@informatik.com or   (610) 640-0339.   In most cases we are able to make the necessary changes within 24 hours, and make the new version available via our web page or be E-mail.   WebParse can be adapted to virtually all E-mail formats.   To adapt your particular E-mail system, please contact Informatik.

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 print file of the 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.

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 'mailto' 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 response forms, parses the text, extracts the relevant data, formats the text 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.

Most users will use the 'single message' extraction method.   With this method, you highlight the response section of the incoming message, activate WebParse, click on the 'Parse' button, and the text will be parsed and displayed.   The extracted data can then be converted to an comma-delimited ASCII text and exported to a text file for later import to a database.

If your incoming email responses are held in individual files, you can combine the files into a temporary 'working' file quick global parsing.

WebParse can also display the extracted data on the computer screen in raw or tabulated format, and the data can be printed.

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 <u>Specifications</u> 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.   <u>This field is left blank if your scan the entire e-mail file</u>.   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 ech 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. 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.   If your E-mail system uses the hexadecimal value for a space (%20), leave the Filler Character field blank; WebParse converts all hexadecimal characters during the course of the conversion.

6. The end tag specifies the symbol that is used to separate the reply fields.   Eudora, for example, uses the ampersand character (&).

7. Click on Automatic Spec Generation.   The system now analyses the survey data and creates the specifications for parsing.   It is important that the fields contain the hexadecimal values, exactly as shown in the E-mail file.

8. 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.   This new field must NOT be the first field.

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 mmddyy.

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

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.

Special Situations

Your web form replies may already be formatted and tabulated.   With certain exceptions and limitations, WebParse can parse formatted text.   The specifications for formatted text parsing must be entered manually (you cannot use the Automatic Spec Generation).   Because of the diversity of formatters, the WebParser may not provide the results you expect.   If you encounter problems or have any questions, please contact Informatik Inc.

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.

   The column headings of the spreadsheet can be customized.   See section below.

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.

Customizing the Column Headings of the Spreadsheet

With an text editor (such as the Windows Notepad), create a text file and in it list the headings for the columns, one heading per line.   The base name of the text file must be the same as the file name of the data file, but with an extension of HDR.   For example, if you want to customize the headings for the extracted data in the TEMP1.TXT file, you must create a 'companion' file TEMP1.HDR file.   Both files must be in the same directory.

Parse a Single Response

Sometimes, you want to extract, parse, view and print individual responses.   The fields of the extracted text must have end tags and the field labels should end with an equal sign (=).

To parse a single response

1.   In your e-mail program, open the e-mail message.   Highlight the survey response section. From the Edit menu, choose Copy.   The highlighted text is copied to the Windows Clipboard.

2.   Activate WebParse.   If WebParse is already running, simply press Alt+Tab once or repeatedly.

3.   If necessary,   specify the end tag and filler settings.   The default settings are: Ampersand (&) for the end tag, equal sign (=) for label end mark and plus sign (+) for filler character.   To change the defaults, choose the Settings menu.

4.   From the WebParse File menu, choose Single Parse, or press the Single button in the tool bar.

5.   The system parses the selected response section and displays the result.

6.   To file the parsed data (formatted), choose Save As from the File menu.   In the Save As dialog box, specify the file.   If the file already exists, the data will be appended.

7.   To append the parsed data (formatted) to a file you have opened before, choose Append from the File menu.

8   To export and append the data to a comma-delimited ASCII file, choose Export as ASCII from the File menu.   The data will be appended to the EXPORT.TXT file in the application's directory.   (The Export.txt file can be deleted from the File menu).

   Some email processors (for reasons unknown) reverse the field order of the responses.   To export such a 'reversed' response, choose Export as ASCII (in Reversed Order) from the File menu.

9.   To print the parse data, choose Print from the File menu.

10.   To display the saved data, choose Open from the File menu.   Specify the Export.txt file.   The file can be viewed 'raw' or in spreadsheet format.

Your web form replies may already be formatted and tabulated.   With certain exceptions and limitations, WebParse can parse formatted text.   Because of the diversity of formatters, the WebParser may not provide the results you expect.   If you encounter problems or have any questions, please contact Informatik Inc.

Please note ....

Unlike the 'global parsing', the 'single parsing' does not use the parse specification file.   Therefore, certain functions like the re-sequencing, date field and automatic numbering are not available.

Tutorial

This tutorial demonstrates the use of WebParse in a 'single parse' mode.   WebParse can also parse collective e-mail files.   To adapt your particular e-mail system, please contact Informatik Inc. at (610) 640-0339 or email@informatik.com.

1.   In your e-mail program, open the e-mail message.   Highlight the survey response section (only the response section, not the header information).

     A typical response section is as follows:

     name=John+Miller&address=150+Main+Street&city=Anytown&state=PA&zip=12345&comments=I+like+it%

     Notice that all reply fields (except the last one) end with the ampersand sign.   We call this the 'end tag'.   The spaces are indicated by a plus (+)sign.   Also, any non-standard characters, like the exclamation sign, is expressed as a hexadecimal value.   But, you don' t need to read the message; WebParse will translate it for you.

2.   From the Edit menu in the email application, choose Copy.   The highlighted text is copied to the Windows Clipboard.

3.   Activate WebParse.   If WebParse is already running, simply press Alt+Tab once or repeatedly.

4.   From the WebParse File menu, choose Single Parse, or press the Single button in the tool bar (yellow square).

5.   The system parses the selected response section and displays the result.

6.   If the parsing did not work correctly, review the tags.   If the field end tag is not an ampersand symbol, or if the filler symbol is not a plus sign, or if the end symbol of the label is not an equal sign,   specify these tags in the Setting menu.   The default settings are: Ampersand (&) for the end tag, equal sign (=) for label end mark and plus sign (+) for filler character.   Click on the Single Parse button again.

7.   To export and append the data to a comma-delimited ASCII file, choose Export as ASCII from the File menu.   The data will be appended to the EXPORT.TXT file in the application's directory.   (To delete the Export.txt file, choose Delete Export.txt File from the File menu).

8.   Go back to your email file and select another response section.   The best way to return to the email program is by pressing Alt+Tab once or repeatedly.   When the response section has been highlighted and copied to the Windows Clipboard, return to the WebParse window (simply press Alt+Tab once or repeatedly) and press on the Single Parse button in the toolbar (yellow square).

9.   Repeat this step for a few more messages.

10. Now, we want to review the parsed data.   While in the Single Parse window, choose Open File from the File menu.   There are two modes of display.   The 'raw' display will show you the comma-delimited ASCII text.   The 'tabulated' display will show you the data in a spreadsheet format.

11. These are just the basics.   You can print the file, save individual messages, etc.   Of course,
      WebParse can parse an entire email file in one process.   Please read the online help for detail.

_____

Typically, the parsing is performed on 'raw' unformatted e-mail messages.   You can parse formatted
text if the text includes end tags.   For all practical purposes,   in the case of formatted text,   these
'end tags' are affixed at the beginning of each label (... the beginning of a text is also the end of the
previous test).   The following is an example:

&name =   John Doe
&address = 150 Main Street
&city = Anytown, NY 12345

(the end tag in front of the first label is optional)

Combining Files

The fast 'global' parsing can only be done if all the incoming email responses are collected in one file.   Some email systems create separate files for each mail message.   WebParse allows you to combine these individual files into one combined temporary working file.

To combine individual files into a 'working file'

1.   From the File menu of the main screen (all windows closed), choose Combine Files.   The Combine Files window is displayed.

2.   Specify the drive and the directory of the files you want to combine.   Double-click on the directory name.   The files in the directory are listed.

3.   Select all the files.   You can select a range of files by clicking on the first file of the range and the last file while holding down the Shift key, or by moving the mouse over the files while pressing the left mouse button.

4.   In the First Label of Response, type the label of the first field of the response section, for example 'Name:='   .   The exact spelling is crucial.   The system will save this setting for your next session.

5.   Choose OK.   The Open File dialog box is displayed.   Select the destination file.   The file is defaulted to the TEMPWORK.TXT file in the application's directory.   (Make sure that you use the same file name in the Parse Specification file).   Click on OK.   If the selected file is not empty, the system will alert you.   The TEMPWORK.TXT file can be deleted by choosing Delete TEMPWORK.TXT from the File menu.