

# Web Scraper Instructions



# Table of Contents

Purpose.....	3
Supported Sites:.....	3
Before you begin:.....	3
Important Limitations:.....	3
Basic Steps.....	3
The Chain User Database: .....	4
Add Top 100 Hits Query: .....	4
Starting the Initial Scrape: .....	5
After the Initial Scrape: .....	5
The Reference Number Field:.....	6
The Second Scrape: .....	6
Unit of Measure Differences .....	6
Factor Edit Form: .....	7
More Info On Factoring: .....	8
Why two Factor Fields?.....	9
Next Steps.....	9
Full Update:.....	9
Partial Update: .....	9
Partial Update – Queued Items Only:.....	9
Import and Match:.....	9

## Purpose

This document covers the operation and use of the DB2 Web Scraper. There are different versions of the Web Scraper depending on the Website you want to scrape. Make sure you are authorized for the particular site before you attempt to scrape a site.

### ***Supported Sites:***

At the time of writing, there are two sites currently supported. Each supported site is assigned a code by OPSoftware. This code is very important as you begin to use Web Scraper. The current list of codes:

D = Office Depot USA.

S = Staples USA

Codes reserved for future use:

C = Office Depot Canada

T = Staples Canada

M = Office Max.

### ***Before you begin:***

You are strongly encouraged to contact Carol Brown to help with the implementation of Web Scraper. The process of gathering, manipulating, and then analyzing this data can be quite a task. Carol has experience in this area and could provide you with a very important head start.

### ***Important Limitations:***

DB2 Web Scraper works best with Windows 2000, XP, or NT operating systems. Both the server and workstation versions are acceptable. Windows 95, 98, and ME can sometimes lock because they never expected the kinds of processes that Web Scraper carries out to ever occur. These other systems may work fine, but they must be up-to-date with the latest patches and with as few as possible 'other' applications.

DB2 Web Scraper requires Microsoft Internet Explorer to operate. You must be on the absolute latest release and have all the patches loaded before you use Web Scraper.

You should, as a general practice, clear all Internet Explorer cached files and re-boot your computer before you start a Web Scrape.

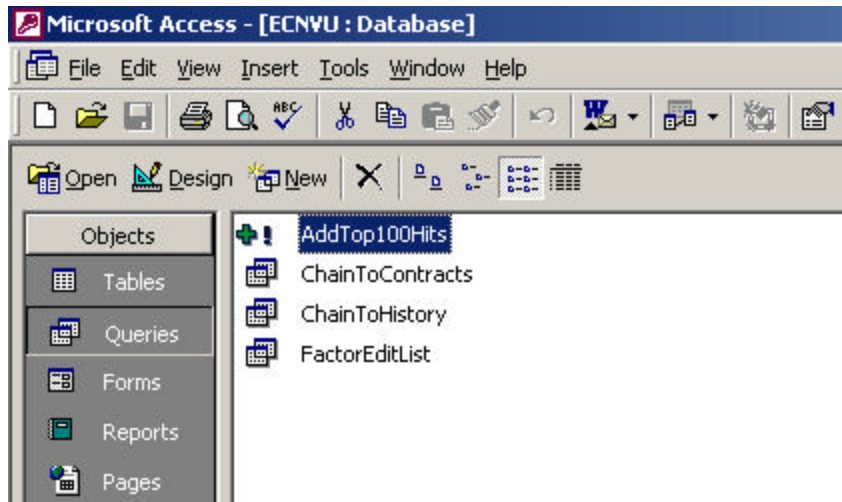
## Basic Steps

There are several basic steps that must be performed before you can download any data. The first step is instructing the Web Scraper what items you want to retrieve information for. You perform this first function by making entries into a 'queue' table included in the Web Scraper database.

**Important! You must always use your primary item number and company when adding records to the queue table!**

## ***The Chain User Database:***

Web Scraper includes a user or work database with several samples and tools you can use to work with Web Scraper data. Open the ECVU.MDB database in Access and choose the Queries Object to display a list of included queries:



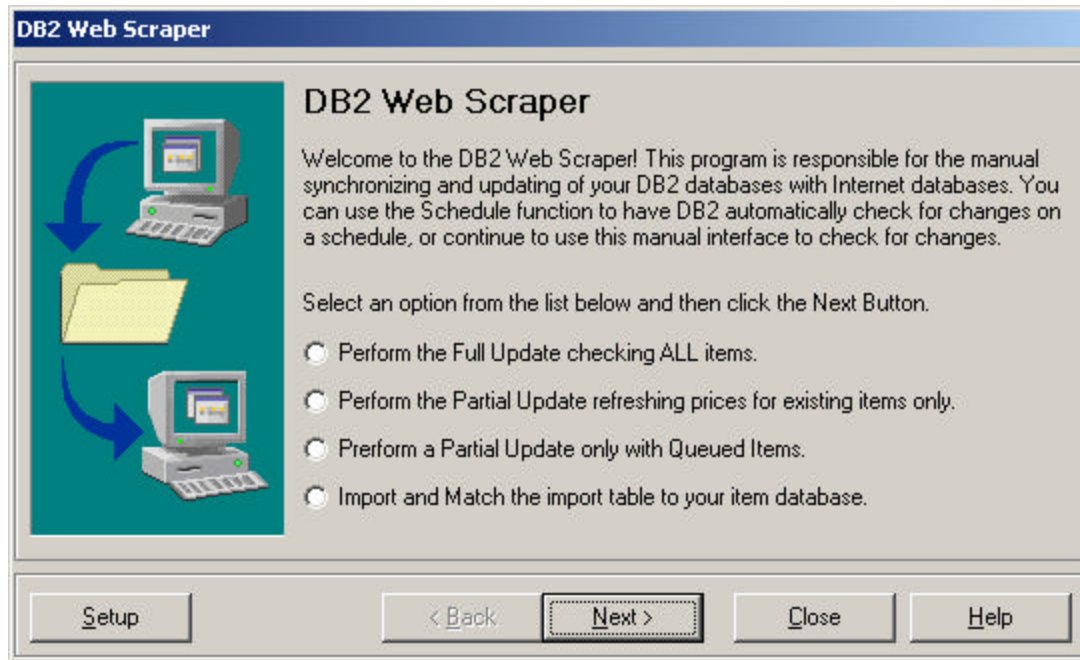
## ***Add Top 100 Hits Query:***

This sample query adds your top 100 items, based on number of year to date hits, to the queue. When you run this query, it will ask you for the chain code. You must use one of the currently active chain codes and you must be authorized for that chain code or the Web Scraper will not work.

You can also create your own queries to add records to the queue. Your queries can be based on whatever criteria you wish and can come from any DB2 table. You only need to have the item number and company code to append to the queue.

## ***Starting the Initial Scrape:***

Once you have added records to the queue table, you are ready to begin the initial scrape. Open the DB2 Web Scraper program and the following screen will appear:



Make sure you choose the Full Update option and then click the Next Button.

Internet Explorer should appear and you can watch as each item is sent to the target Website.

**Do not interfere with Web Scraper while it operates!** If you want to stop the Web Scraper, use the cancel button.

## ***After the Initial Scrape:***

Once the initial scrape completes, your chain database master table (ECNVMASTER) will contain the results. The important column at this point will be the status field. A status of 0 (zero) means that the item was found and the data was successfully recorded. A status of X means that the item was not found on the target Website.

It is the X status items you will need to address next.

## ***The Reference Number Field:***

The chain database master table contains a field called reference number. This field can be used to supply the Web Scraper with an alternate or lookup item number in those cases where the chain stores are using a different item number. If you load OPUS, then you know that sometimes they do not use the same prefixes as the wholesalers. These non-standard prefixes will not be found on the chain Websites.

Currently, and for your reference, Staples uses USSCO item numbers and Office Depot uses SPR item numbers.

What you need to do next is create a query to modify the “X” status items and change the Reference Item Number field (REFITEM) to the appropriate wholesaler item number and prefix. This can be complex and is a major reason we suggest you obtain the services of Carol Brown.

Once you have finished modifying as many reference numbers as you can, you then need to re-add all of these modified records back to the queue. You may want to just create a query that re-adds all “X” status items back to the queue. Remember, you must add items to the queue using your primary item number and company code. Never use the reference item number in the queue!

## ***The Second Scrape:***

You now need to re-run the Web Scraper. Open the Web Scraper and this time choose the Perform A Partial Update Only With Queued Items option. This will only process the queued items and change the status to 0 (zero) for any new items found.

Please note that you may be required to repeat the process of modifying the reference item number and re-scraping several times until you get all the matching items you wish.

## **Unit of Measure Differences**

There is a good chance that one or several items will be sold by the chain stores in different units of measure than you do. It is not uncommon for the chain stores to sell common items in three and five packs whereas you may sell them by the each or dozen.

There is a special form included in the chain store user database to address this problem.

## Factor Edit Form:

Open the chain store user database and click on the Forms objects. Open the Factor Edit form and the following form is shown:

The screenshot shows the Microsoft Access interface for the FactorEditList form. The form contains the following data:

Code:	ItemNumber:	30301	Company:	PAP	RefItem:	181529	Status:	0
SKU:	181529	Manufacturer:	Papermate/Parker/Waterman					
Description:	PENCIL_MECH_DISPOSABLE_YW Sharpwriter's cushion point lead adjusts to writing pressure, reduces breakage and never needs sharpening. Retractable with a convenient pocket clip. Non-refillable.							
	Sharpwriter Pencil, #2 Lead, Yellow							
	#2; Yellow							
Chain Description:	Paper Mate® Sharpwriter Pencils, 0.7 mm, Yellow Barrel, Pack Of 12							
Unit Measure:	EA	12 EA/BX	Your Factor:	1	Pack 2:	36 BX/CT		
Chain Unit:	dozen		Chain Factor:	1				
Chain Price:	\$4.99		List Price:	\$0.52				
Difference Factor:	.10		Factored Price:	\$4.99				

In the example shown, your unit of measure is EA and the chain store unit of measure is dozen. To correct this problem, and to compare the prices using the same unit of measure, you will need to make entries into the “Factor” boxes.

In this case, you would want to enter 12 as the chain store factor. Your unit is EA (1) and their unit is dozen (12).

Once you enter the factors, click the Next Record button (>) at the bottom of the form to display the next record.

Note the data pulled, and the criteria for selecting the items to check, comes from a query of the same name in the chain user database. You can modify this query if need be.

## More Info On Factoring:

You probably already know that one method used by chain stores to prevent consumers from comparing prices is packaging in unusual quantities.

Although not contained in the most recent catalogs, I have seen cases where 'house' brand transparent tape is 900 inches where most dealers carry 1296-inch rolls.

A more common unit is five packs (not evenly divisible by 12, which is how most independent dealers may carry an item).

This has been addressed in the Web Scraper by allowing you to input two conversion fields. The two fields were used to provide the most straightforward conversion and to provide for all possibilities.

Suppose for example, an item is carried by a chain in a five pack, you carry the item by the dozen. You would enter in the two conversion fields, Chain Factor and Your Factor the following:

Chain Factor = 5

Your Factor = 12

The Web Scraper will then automatically generate a "Factored Price" using these factors you enter. Here is how the calculation works:

Chain Factor / Your Factor = Factor

Chain Price / Factor = Factored Price

Using the numbers in the above example with a chain store 5 pack price of \$6.99:

$$5 / 12 = .41666$$

$$6.99 / .41666 = \$16.776$$

So the chain store factored price is \$16.78 per dozen, which matches your unit selling price.

Note that this can work either way. For example suppose you sold the item by the each:

Chain Factor = 5

Your Factor = 1

$$5 / 1 = 5$$

$$6.99 / 5 = \$1.398$$

Again the factored chain store price is reflected in your unit, each in this case.

This method also addresses those more unusual situations such as the 900-inch VS 1296-inch roll.

Chain Factor = 900

Your Factor = 1296

And again, you will be comparing apples to apples as this example arrives at the factored price by breaking it down to the price per inch. See for yourself. Their price \$.99 for a 900 inch roll, your price \$1.29 for a 1296 inch roll. Who has the lower price?

$$900 / 1296 = .69444$$

$$.99 / .69444 = \$1.425$$

## Why two Factor Fields?

Why not just one field containing the factor?

It is much easier for you to spot check. For example, looking at a report or display, what would you rather see?

They sell it by 5, I sell it by 12.

or,

Factor = .41666

It is also much easier to enter and you don't even need a calculator.

## Next Steps

Once you get your data loaded, it then becomes the task of keeping it updated. This last task is much simpler and only requires a periodic running of the Web Scraper on your existing data.

Open the DB2 Web Scraper and choose one of the following options:

### ***Full Update:***

This option will re-add all items from your existing Master table to the queue, and then scrape those items, one at a time, removing them from the queue as they are processed. The Full Update option adds ALL records to the queue, even those with a status of "X".

### ***Partial Update:***

This options will re-add all items from your existing Master table only if the current status is 0 (zero). This prevents re-scraping items you know are not carried by the chain stores anyway. As with the Full Update, the items are first added to the queue and then scraped one at a time, and removed from the queue.

### ***Partial Update – Queued Items Only:***

This option will not add any items to the queue. It will only process items already in the queue. You want to use this option initially as your Master table will not contain any items when first delivered.

### ***Import and Match:***

This special option allows you to import and match someone else's Web Scraper data. This option will read the special import table and then match the data to your item files and your item number, and then add the items to your Chain Master table.