**Specs for the amazon scraping project**

**Table of contents**

x01. The input for the scrapper program

x022. The values to be scraped from the listing for an ASIN in the list of ASIN

x03. Data scrapped from Amazon to be put in an excel spreadsheet as shown in the example spreadsheet with results:

x04. Spreadsheet column formatting:

x05. Typical Amazon.com page with ASIN in search window:

x06. Code documentation and rights

xx900 end of initial scraper specs document

**X01. The input for the full scrapper program**

Take values of UPC codes, ISBN numbers, or Amazon ASIN’s from a list of ASIN’s in a column of a Microsoft Excel spreadsheet. The input will either be a combination pf UPC codes and ISBN numbers or a combination of ASINs codes and ISBN numbers, but no UPC codes. Note that the search window in an Amazon.com webpage will find a product regardless of whether you put the ASIN, UPC code or ISBN number.

**UPC Codes** are normally 12-digit numbers but are sometimes 10-digit numbers. All the characters of the UPC code must remain numeric characters. Microsoft Excel has a tendency to change the UPC codes into scientific notation. This must not be allowed to happen. The cure for this is to format the column with the UPC coded as a number with zero decimal place.

**ISBN numbers** have either 13 numeric characters or 10 numeric characters. Like the UPC codes, all the digits of the ISBN numbers must be preserved and they cannot be allowed to change into scientific notation. For books, ISBN numbers are used instead of UPC codes.

**Amazon ASIN’s** are a string of 10 alpha-numeric characters. Most commonly they the begin with the letter “B” and are a mix of both letters and numeric characters. They can also have all 10 characters be numeric characters.

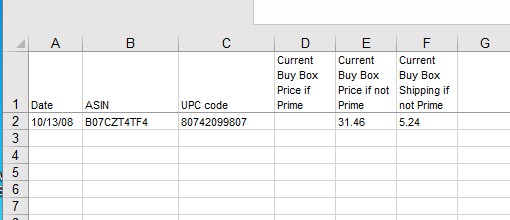
**X02. The values to be scraped from the listing for an ASIN in the list of ASIN**

For each ASIN, return the Current Buy Box Price if the buy box seller is an FBA seller (the sale will be Amazon Prime Sale); or the Current Buy Box Price and Current Buy Box Shipping, if the seller is an FBM seller (the sale is not an Amazon Prime sale). The pictures and the values in the example spreadsheet are for ASIN = B07CZT4TF4 at one instant in time.  
Similarly, the Amazon scrapper program is finding the values for one instant in time.

**X03. Data scrapped from Amazon to be put in an excel spreadsheet as shown in the example spreadsheet with results:**

**Example spread sheet with results**

ScreenHunter 2846.jpg



Use the values shown in the example spreadsheet above combined with the pictures below (for each value to be scraped there is a picture with an arrow showing where it is on an Amazon listing page).

**X04. Spreadsheet column formatting:**

**Font**

all fonts to be Arial size 8 - no colored fonts

**Highlighting and shading**

No highlighting or shading

**alignment**

Horizontal - left  
vertical - bottom

wrap text whenever there is a possibility of the text not fitting.

no combined cells

**money**

all columns with money to be formatted as a number with 2 decimal places, no commas or dollar signs, negative values indicated only by a minus sign - no parentheses or red color

**date**

The best way to describe the proper date format is with this example:

Christmas 2018 should look like this: 12/25/18

**row height**

must always be big enough so that no data is hidden

**Number type for each column**

date- see “date” above

ASIN

Current Buy Box Price if Prime - see ”money” above

Current Buy Box Price if not Prime see -”money” above

Current Buy Box Shipping if not Prime -”money” above

the number of sellers - number with zero decimal places

the Best Seller’s Rank (BSR) Category 1 - text (without hyper link)

BSR for Category 1 - number with zero decimal places

BSR category 2 - text (without hyper link)

BSR for Category 2 - number with zero decimal places

shipping weight - number with zero decimal places

the seller who has the buy box - text (without hyper link)

**Column widths**

Date = 7

ASIN 13

Current Buy Box Price if Prime = 8

Current Buy Box Price if not Prime = 8

Current Buy Box Shipping if not Prime = 8

the number of sellers = 7

the Best Seller’s Rank (BSR) Category 1 = 24

BSR for Category 1 = 8

BSR category 2 = 24

BSR for Category = 24

shipping weight = 7

the seller who has the buy box = 24

**Borders**

No borders

**Bold and underlining**

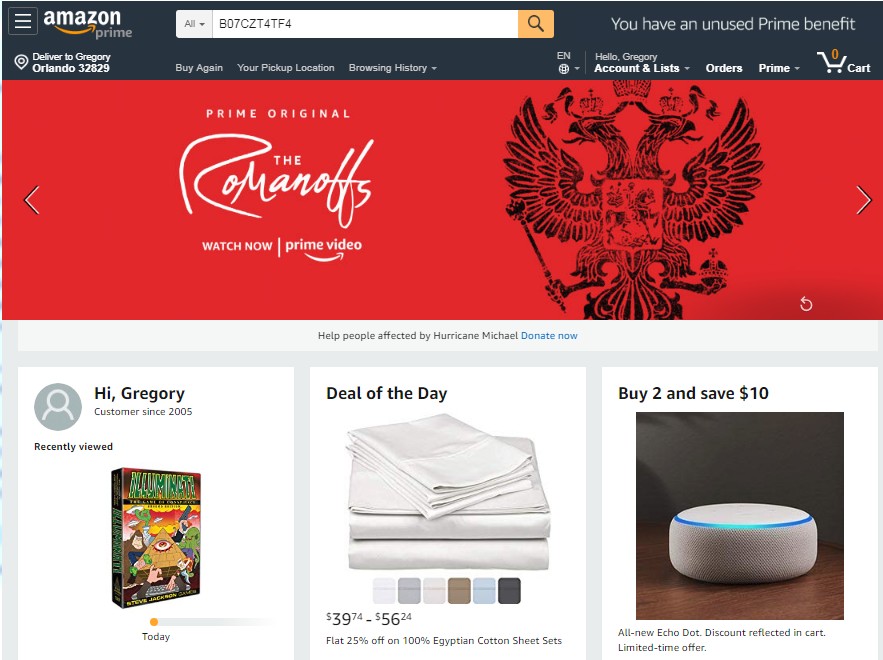
No Bold or underlining

**Hyperlinks**

No hyperlinks

**X05. Typical Amazon.com page with ASIN in search window:**

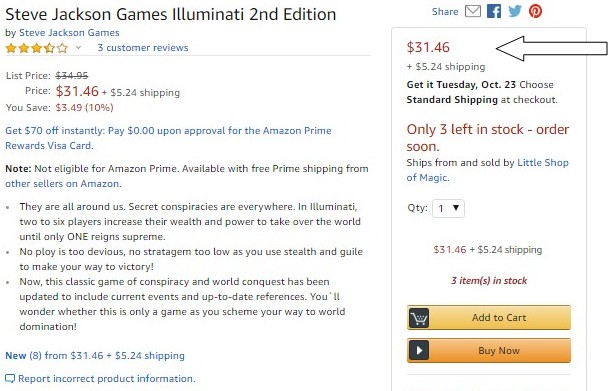
ScreenHunter 2827.jpg



**Pictures of the values to be scrapped from an Amazon listing for a given ASIN for one instant in time.**

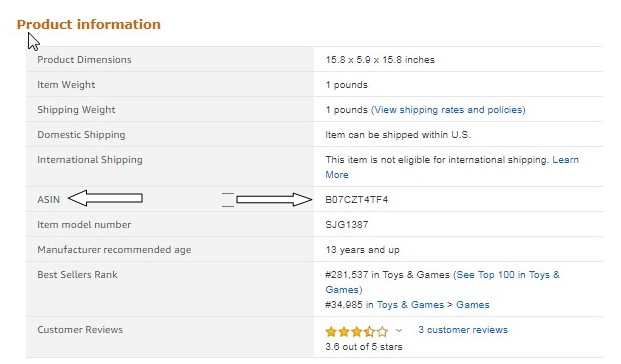
**Buy Box Price**

ScreenHunter 2819.jpg



**ASIN**

ScreenHunter 2825.jpg



**X07. Code documentation and rights**

The code of the computer programs shall be sufficiently documented so that third party developers will be able to maintain and improve the code and all rights to the code and its use, including the right to use it in future upgrades and together with other software shall belong exclusively to Markham Lake Products, LLC and its assigns.