

Web Scraping Pricing Analytics



Contents

HOW DID WE GET HERE



Then



Now



Q & A

Appendix

Gathering Web Scrape Data

1

Web Scrape Data Quality

2

What to do with all that Data

3

Index Calculation

4

When is a Match not a Match

5

Time Calculation

6

Overview Scrape Process

7

Early Days



Manual



Excel



Small Scale



Slow
Process

Mid Term



Third Party
scrape partner



More
Data



Use of
Access



Excel used for
stakeholders



Slow
Process

Now



Integration into
Data warehouse



Analysis for
excel



New Third Party
(expansive)



Increased
data



Consistency
Data



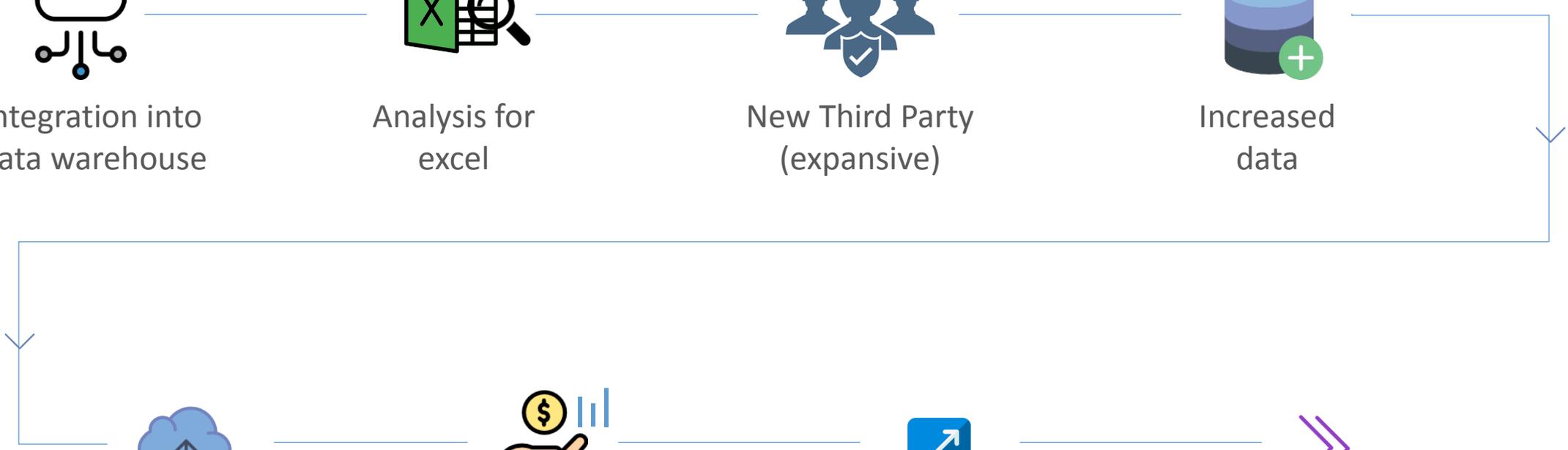
Not only Price
Indexing



Scale



Further
Options





eClerxDigital

Competitive Intelligence
Use Cases



Move up the growth curve

Uncover key growth opportunities faster and discover new & unknown potential threats



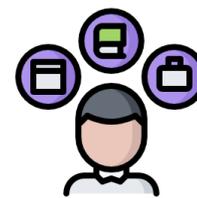
Rapid and improved business decisions

Make the right decisions even with fragmented or incomplete data to stay ahead of market changes



Gain advantage in complex environment

Access high-volume data in high activity situations to monitor the changing market landscape



Consultative and proactive

Proactive approach to actionable insights and recommendations for a positive business impact



Price | Promotion

- Competitor, Marketplace/Metasearch, Google shopping/PLA, other channels | Site, Mobile App
- Contract Violations, Price Alerts
- Market and FX trends vs. commodity prices vs demand vs sales trends
- Associated products impact on pricing decision
- Anchoring, Corridors, Optimization



Product

- Product gaps/overlaps, New entrants
- Full catalogue listing/delisting movements, Matched product movements
- Product availability, product mix behaviors/trends
- Inventory trends, Fast moving products/brands
- Cross-referencing, Alternative products, Functional equivalents



Content

- Manufacturer/3rd party content sourcing, preparation (text, images, video, files)
- Consistency (owned & referral digital channels)
- Enhance/ enrich assets on product pages
- Compare content strategy with direct/indirect competition, Taxonomy benchmarking



Placement

- Top x% products visibility across channels
- Market share of search
- Compare listing of top movers
- Rating & Review, Customer feedback/ sentiment, Changes over time, Key themes for action

Thank you for your attention.



Questions

eClerxDigital

Saurabh Sharma (Head of Practice – Competitive Intelligence)
Saurabh.Sharma@eClerx.com

Prashant Patel (EMEA Client Engagement Team)
Prashant.Patel@eClerx.com



Gary Malpas (Commercial Analytics Manager)
Gary.Malpas@rs-components.com

Backup – eClerx Digital

WHY ECLERX

- **17+ years of experience** supporting market & competitive intelligence programs
- **Consultative and customized** solution design
- **Award winning** proprietary platform processing billions of data-points per month
- **Quick adaptation** to changing requirements
- **Lower total cost** and faster time to market

Size & Scale

3.5 billion+

DATA POINTS
processed per month

1.2 billion+

MATCHES
suggested and published

10,000+

CUSTOMISED DASHBOARDS
delivered per month

8,000+

DIGITAL SOURCES
tracked for market information

25 times

FASTER
than standard matching process

99%+

ACCURACY in automated data processing

400+

DEPLOYMENT done across industries

500 million+

PRODUCTS
cross referenced



INDUSTRY AWARDS received in last 3 years



Challenges



eClerx Solution



Business Outcomes

- 1 Frequent changes in source websites
- 2 Disruptions in data due to access problems
- 3 Inconsistent data structure
- 4 Variety in data types required
- 5 Precise output formats required

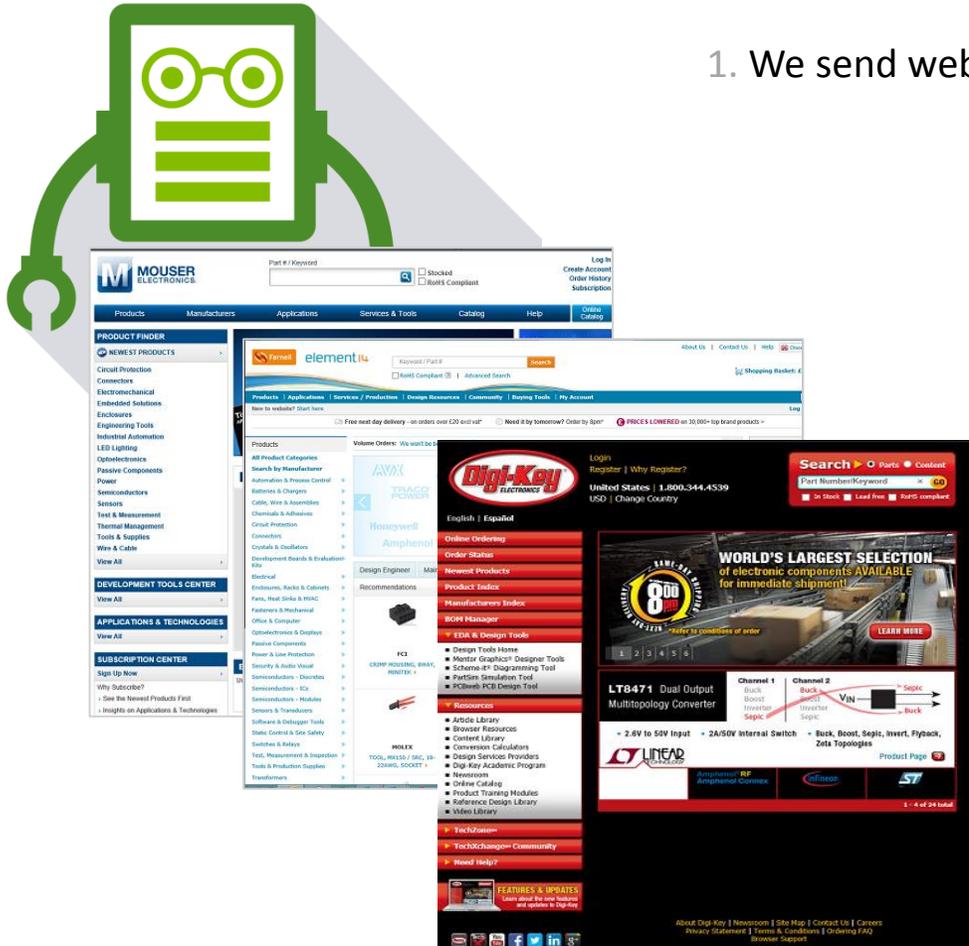
- Machine learning based detection of changes to site structure
- Reconfiguration of data extraction engine within minutes
- Multipoint data sourcing system; adjustments for website policies and advance stealth
- ML based attribute matching & feature value adjustment
- Data cleansing and normalization tools plus comprehensive data QA
- Capture any available data (text, image, video, files) from digital sources
- Data output is structured to match customer needs, whether for reporting or as an input to other systems

- Zero data gaps
- Complete and accurate data availability
- Matched & clean data in correct format for analysis and data feeds
- Robust data collected to enable detailed analysis of market positioning
- Actionable insights and seamless integration with reporting environments

Backup – RS Components

Gathering Web Scrape Data

1. We send web-scrafer bots out onto our competitors web sites....



Vendor Matching
& Alignment



We then bring all the data into our BI infrastructure for processing and reporting

Timing to scrape 500 Milliseconds = ½ a second

1.5 - 3.5 Milliseconds to do one scrape multiplied by 3.7Million records for

Mouser equals over 3.5 hours to complete,

that's if there are no issues which is rarely the case!

We scrape 15+ Farnell sites with an Avg of 650,000 articles each

Web Scrape Data Quality

Keywords:

In stock
 Lead free
 RoHS Compliant

Search Again

Product Index > Cables, Wires - Management > Heat Shrink Fabric > TE Connectivity Raychem Cable Protection HFT5000-25/12-0-SP

Digi-Key Part Number	HFT5000-25/12-0-SP-01-ND	Price Break	Unit Price	Extended Price
Digi-Key Stock	220	1	4.00000	4.00
Quantity Available	C an ship immediately	2	3.47000	6.94
Value Added Item		5	3.06000	15.34
Manufacturer	TE Connectivity Raychem Cable Protection	12	2.66833	32.02
Manufacturer Part Number	HFT5000-25/12-0-SP	250	2.13504	533.76
Description	HEAT SHRINK FABRIC 1+1M			
Lead Free Status / RoHS Status	Lead free / RoHS Compliant			

Quantity: Item Number: Customer Reference: Add to Cart

When requested quantity exceeds displayed pricing table quantities, a lesser unit price may appear on your order. You may submit a request for quotation on quantities which are greater than those displayed in the pricing table.

Datasheets: [HFT5000 Drawing](#), [HFT5000 Catalog Page](#), [HFT5000 PCB Sheet](#), [HFT5000 Cable Heat Shk](#)

Product Photos: [HFT5000-25/12-0-SP](#)

Standard Package: 1

Category: Cables, Wires - Management
Family: Heat Shrink Fabric
Series: HFT5000-Raychem
Shrinkage Ratio: 2 to 1
Length: 3.28' (1.00m)
Diameter - Inner, Recovered: 0.47" (12.0mm)
Diameter - Inner, Supplied: 0.58" (15.0mm)
Wall Thickness - Nonrecovered: -
Color: Black
Other Names: HFT5000-25/12-0-SP-01

Report an Error

Complexity of website



LOADING
PLEASE WAIT...

Part No. / Keyword:

Stocked
 RoHS Compliant

Products Manufacturers Applications Services & Tools Catalogue Help

All Products > Optoelectronics > LED Lighting > LED Emitters > High Power LEDs - White > OSRAM Opto Semiconductors GW MAFJ1 CM-RUSS-4053

OSRAM Opto Semiconductors

Mouser Part No: T20-GWMAFJ1CMRUS40
Manufacturer Part No: GW MAFJ1 CM-RUSS-4053
Manufacturer: OSRAM Opto Semiconductors
Description: High Power LEDs - White White 4000K SOLDER P P
Lifecycle: New Product, New from this manufacturer.

Learn more about OSRAM Opto Semiconductors GW MAFJ1 CM-RUSS-4053

Data Sheet

Images are for reference only
See Product Specifications

Add to Compare List

Share

Specifications	Value	Check
Manufacturer:	Osram Opto Semiconductor	<input checked="" type="checkbox"/>
Product Category:	High Power LEDs - White	<input checked="" type="checkbox"/>
RoHS:	Details	<input checked="" type="checkbox"/>
Brand:	OSRAM Opto Semiconductors	<input type="checkbox"/>
Illumination Colour:	Neutral White	<input type="checkbox"/>
Colour Temperature:	4000 K	<input type="checkbox"/>
Luminous Flux/Radiant Flux:	1640 lm	<input type="checkbox"/>
Colour Rendering Index - CRI:	95	<input type="checkbox"/>
Viewing Angle:	120 deg	<input type="checkbox"/>
Forward Current:	700 mA	<input type="checkbox"/>
Forward Voltage:	29 V	<input type="checkbox"/>
Mounting Style:	SMD SMT	<input type="checkbox"/>
Packaging:	Tray	<input type="checkbox"/>
Height:	1.1 mm	<input type="checkbox"/>
Length:	15.2 mm	<input type="checkbox"/>
Maximum Operating Temperature:	+ 85 C	<input type="checkbox"/>
Minimum Operating Temperature:	- 40 C	<input type="checkbox"/>
Width:	15.2 mm	<input type="checkbox"/>
Part # Aliases:	Q8511A5056	<input type="checkbox"/>

To add to a project, please Log In.

Log In to create a note about this product or see notes that you previously created on this product.

Changes to site structure



What To Do With All That Data



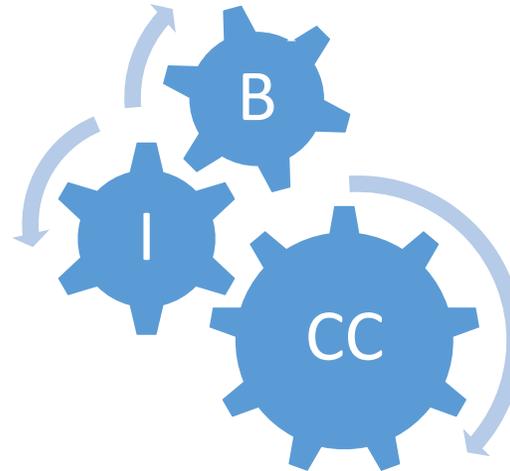
www.variouswebsitedata.com



RAW DATA



MATCHED DATA



5xRule

Index Calculation

Competitor Index measures our price position. At an article level, it refers to the RS price as a percentage of the competitor Price.

Article	Competitor Price	RS Price	Calculation	Index
A	10	9.2	=9.2/10	92%
B	40	50	=50/40	125%

Indices are weighted according to Sales Volumes. So, our key top selling products feature more prominently in the overall weighting.

Article	Competitor Price	RS Price	Calculation	Index	Sales Volume	Weighted Index
A	10	9.2	=9.2/10	92%	500	94%
B	40	50	=50/40	125%	10	

Index <100 = CHEAPER

Index >100 = EXPENSIVE

When Is A Match Not A Match

Battery Name and Type

MN2400



RS Code:- 448-8460
MPN:- 15071690

MX2400



Farnell code 2345270
MPN:- 5000394002746

Specification

1.5 V

pack of 8

AAA

Alkaline

Power Check Gauges

Also known as LR03 Batteries

Defined as Ultra Power (by Duracell)

1.5 V

pack of 8

AAA

Alkaline

Power Check Gauges

Also known as LR03 Batteries

Defined as Ultra Power (by Duracell)

Difference

In truth only a newer version and little to no difference.

[BACK](#)

Time Calculation



Double click on work sheet and enter new numbers for Milliseconds and # of records to see impact

Second	0.0035
# of Records	3,789,000 Enter
Minutes	221
Hours	3.7
Days	0.2

Overview Scrape Process

