



Statistics Netherlands

Division R&D

*P.O.Box 24500
2490 HA The Hague
The Netherlands*

**User guide RobotTool
version 4.0.0**

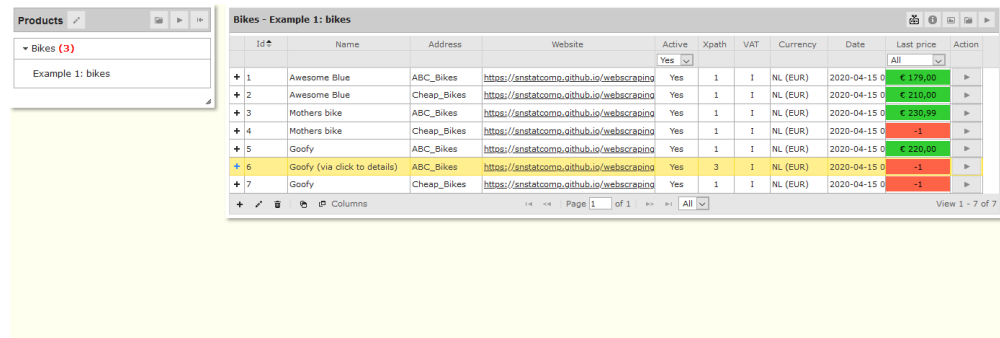
<https://github.com/SNStatComp/RobotTool>

15-4-2020

1.	Main screen.....	2
1.1	Editing product groups.....	2
1.2	Editing respondents	2
1.3	Editing the XPath expressions of a respondent.....	2
1.4	Editing Price and Price context.....	3
2.	Product group.....	3
2.1	New	3
2.2	Editing.....	3
2.3	Deleting.....	4
2.4	Import and export of product groups	4
3.	Respondents	5
3.1	New	5
3.2	Editing.....	6
3.3	Deleting.....	7
3.4	Copying.....	7
3.5	Selecting columns	8
4.	Respondent – XPath	9
4.1	New	9
4.2	Editing.....	10
4.3	Deleting.....	11
4.4	Showing price context	11
5.	Price and Price context	12
5.1	Collecting new price info for a product group.....	12
5.2	Examining Price en Price context.....	13
5.3	Taking the price from the previous response.....	13
5.4	Collecting/changing price info respondent	13
5.5	Exporting prices.....	15
	Annex A: record description export file	16
	Annex B: Configuring the RobotTool	17
	Annex C: Learning XPath	18

1. Main screen

You start the RobotTool by executing the file ‘robottool.bat’ (on windows) or typing ‘npm start’ in a command window in the RobotTool folder. After the start of the RobotTool you see a screen like below. This is the start screen after loading the example on collecting data from fake bike shops which is provided in the release of the tool:



The screenshot shows the RobotTool interface. On the left is a sidebar with a 'Products' section containing a dropdown menu labeled 'Bikes (3)' and a link to 'Example 1: bikes'. The main area displays a table titled 'Bikes - Example 1: bikes' with the following data:

	Id	Name	Address	Website	Active	Xpath	VAT	Currency	Date	Last price	Action
+	1	Awesome Blue	ABC_Bikes	https://anstatatcomio.github.io/web scraping	Yes	1	1	NL (EUR)	2020-04-15 0	€ 179,00	▶
+	2	Awesome Blue	Cheap_Bikes	https://anstatatcomio.github.io/web scraping	Yes	1	1	NL (EUR)	2020-04-15 0	€ 210,00	▶
+	3	Mothers bike	ABC_Bikes	https://anstatatcomio.github.io/web scraping	Yes	1	1	NL (EUR)	2020-04-15 0	€ 120,99	▶
+	4	Mothers bike	Cheap_Bikes	https://anstatatcomio.github.io/web scraping	Yes	1	1	NL (EUR)	2020-04-15 0	-1	▶
+	5	Goofy	ABC_Bikes	https://anstatatcomio.github.io/web scraping	Yes	1	1	NL (EUR)	2020-04-15 0	€ 220,00	▶
+	6	Goofy (via click to details)	ABC_Bikes	https://anstatatcomio.github.io/web scraping	Yes	3	1	NL (EUR)	2020-04-15 0	-1	▶
+	7	Goofy	Cheap_Bikes	https://anstatatcomio.github.io/web scraping	Yes	1	1	NL (EUR)	2020-04-15 0	-1	▶

At the bottom of the table, there are controls for columns, page navigation (Page 1 of 1), and a view count (View 1 - 7 of 7).

1.1 Editing product groups


The RobotTool is designed to monitor the prices of products from web pages of respondents. The left-hand panel of the dashboard shows the products organised in a number of product groups. You can edit the product groups via the “Edit” button on that panel, see also chapter 2. The red figure between parentheses behind a product group indicates the number of respondents in that product group for whom no price is known yet (price=-1). By clicking on a product group the respondents belonging to that product group are retrieved and shown in the right hand section of the screen.

1.2 Editing respondents

Each product group can have several respondents. For editing the respondents of a product group: see chapter 3.

1.3 Editing the XPath expressions of a respondent

The RobotTool uses the XPath language to configure the scraping path through websites and to define the exact position on the webpage of the product (the price context). To retrieve the price context for a respondent from the internet you have to

link one or several XPath expressions with that respondent. By clicking on  for a respondent (see screen 1) you can edit the instructions and XPath expressions for that respondent. More details can be found in chapter 4.

1.4 Editing Price and Price context


You can use the XPath expressions belonging to a respondent to retrieve the price context for that respondent from the internet. For editing the price and price context: see chapter 5.

2. Product group


The screen below is the basic screen for editing product groups (see chapter 1.1).

Productgroups			
Id	Description	Comment	Action
11111	Electronics	Prices of electronic products in 5 countries	
22222	Cinema tickets		
33333	Driving lessons		

Page 1 of 1 10 View 1 - 3 of 3

NB: In these kinds of tables the number of lines shown can be changed by clicking the downward arrow: 

2.1 New

You can add a new product group by clicking on icon  which gives:

Add Record

Product

Description

Comment

Submit

Cancel

2.2 Editing

You can edit an existing product group by clicking  getting this screen:

Edit Record

Product


Description

Comment

Submit


Cancel

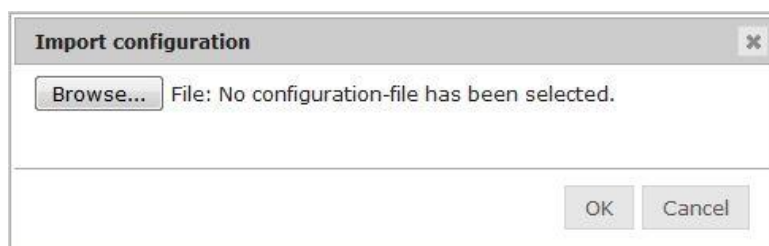
2.3 Deleting

You can delete an existing product group by clicking on icon 




2.4 Import and export of product groups

You can **import** a configuration of product groups (i.e. a product group with its respondents and XPath expressions) by clicking on icon  which give you the following windows:



You can select a previously saved configuration file (extension ".csv") by using "Browse".

Note: the Browse button can be used to navigate to a folder on your computer and select a file (extension '.csv'), however the path from which the browser tries to read that file from the folder as specified in the configuration file of the RobotTool (see Annex B) which is by default the folder 'ImportExport' in the RobotTool folder.

You can **export** a configuration of product groups (i.e. a product group with its respondents and XPath expressions) by clicking on icon .

The configuration that is loaded at that particular moment will be saved in a configuration file (.csv) in the Export folder of the RobotTool (see Annex B).

3. Respondents

The main screen below is the basic screen for editing respondents. This example shows the panel after running the example configuration on fake bike shops produced with the release as an example.

Bikes - Example 1: bikes										
Id	Name	Address	Website	Active	Xpath	VAT	Currency	Date	Last price	Action
				Yes					All	
+ 1	Awesome Blue	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 179,00	▶
+ 2	Awesome Blue	Cheap_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 210,00	▶
+ 3	Mothers bike	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 230,99	▶
+ 4	Mothers bike	Cheap_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	-1	▶
+ 5	Goofy	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 220,00	▶
+ 6	Goofy (via click to details)	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	3	I	NL (EUR)	2020-04-15 0	-1	▶
+ 7	Goofy	Cheap_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	-1	▶

3.1 New

You can add a new respondent by clicking on icon  getting this window:

Add Record

Id

Name

Address

Website

Active

Yes

Comment

Note 1

Note 2

Note 3

Note 4

Note 5

VAT

Included

Currency

NL (EUR)

Product

Bikes


Submit

Cancel

Most fields are easy to understand. The Website field should always be filled with a URL. The RobotTool uses this to start scraping.

The Active field can be used to deactivate respondents from the following scraping run.. Sometimes collecting prices for a respondent becomes impossible after a while (e.g. the respondent has stopped the activities or the prices cannot be retrieved for technical reasons). In those cases it is handy to switch the respondent to inactive (Active = No), still maintaining all configuration and data for this respondent. When inactive, no more prices will be retrieved for this respondent.

3.2 Editing

You can edit a respondent by clicking on icon  giving something like:

Edit Record

Id

6

Name

Goofy (via click to details)

Address

ABC_Bikes

Website

<https://snstatcomp.github.io/web scrapingtests/R>

Active

Yes

Comment

Getting the Goofy bike via click on button, wait for loading detail page and getting the price from the detail page

Note 1

Note 2

Note 3

Note 4

Note 5

VAT

Included

Currency

NL (EUR)


Product

Bikes

Submit

Cancel

3.3 Deleting

You can delete a respondent by clicking on icon  giving you a warning:


Delete ✕

Warning: all observations from this source will be deleted also!

Delete

Cancel

3.4 Copying

You can copy a respondent by clicking on icon  raising:

Copy respondent ✕

Id

Name

Goofy (via click to details)

Address

ABC_Bikes

Website

<https://snstatcomp.github.io/web scraping tests/R>

Active

Yes

▼

Comment

Getting the Goofy bike via click on button, wait for loading detail page and getting the price from the detail page

⋮

Note 1

⋮

Note 2

⋮

Note 3

⋮

Note 4

⋮

Note 5

⋮

VAT

Included

▼

Currency

NL (EUR)

Product


Bikes

▼

Submit

Cancel

3.5 Selecting columns

You can indicate which columns you want to see in screen 3. By clicking on icon  Columns you will get the following screen which allows you to select and deselect columns for display:

Select columns

8 items selected

Remove all

Add all

Id	-	Comment	+
Name	-	Note 1	+
Address	-	Note 2	+
Active	-	Note 3	+
Xpath	-	Note 4	+
VAT	-	Note 5	+
Currency	-		
Date	-		

Ok


Cancel

4. Respondent – XPath

The following screen is the basic screen for editing the XPath expressions of a respondent. You reach this screen by pressing the ‘+’ icon before an entry in the right panel. A subpanel will open containing the steps (also called Xpaths for short) that belong to that product to be found on the website specified. Below is an example showing one step, the retrieval of the part of a webpage specified in the website filed specified by the XPath expression `//*[@id='price1']`. See Annex C for some links to resources for learning how XPath works.

2	Awesome Blue		Cheap_Bikes	https://snstatcomp.github.io/webscraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 210,00	
#	Type	XPath			Parameter				Action		
1	Context	//*[@id='price1']									
<div><div><div></div><div></div><div></div></div><div><div></div><div></div><div></div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></</div></div>											

4.1 New

You can add a new XPath expression by clicking on icon . A step number is automatically generated that is one up from the largest number of steps existing for this respondent. The step number can be changed but cannot be the same as another step number.

Add Record

XPath

Parameter

Type
Context

#
3

Submit
Cancel

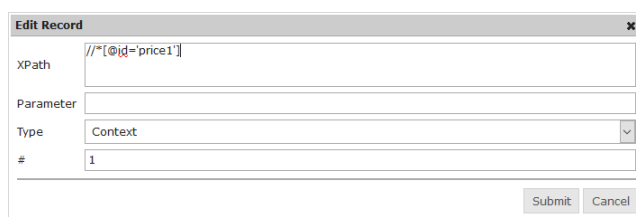
For the **Type** field you can choose from several types of instructions summarised in the table below. All instructions except the Wait command need an **XPath** expression to determine on which part of the web page they work. Some instructions take an additional parameter from the **Parameter** field.

Type	Description	Example
Context	Get the text part of the element specified in the XPath field	XPath: <code>//*[@id='price1']</code>
Link	Navigate to the page referred to in the <i>href</i> attribute of the selected element. If the tag of this element is not equal to <code><a></code> a click action is performed.	Xpath: <code>Id('menu-item')/a</code>
Input	Insert text into the selected element. This text is taken from the <i>Parameter</i> field.	Xpath: <code>//input[@class='search']</code>

		Parameter: lasagna
Submit	Submit the webform of the selected webform element	
Screenshot	Take a screenshot of the current webpage and save as HTML and .png. The screenshot will be save in the folder 'screenshot'	
Click	Click on the selected element.	XPath: /html/body/div/a
Wait	<p>If the field <i>Xpath</i> is empty the robot waits a number of milliseconds as specified in the <i>Parameter</i> field.</p> <p>If the field <i>Xpath</i> is not empty the robot waits until the selected element is visible by the robot or a timeout occurs. The timeout value (in milliseconds) is taken from the <i>Parameter</i> field.</p> <p>Default: 10000 (10 seconds)</p>	Parameter: 1000
mouseMove	<p>Move the mouse pointer to the selected element.</p> <p>Use this option when an element you are interested in is a part of a popup menu.</p>	
Download	<p>Download the file referred to in the selected element.</p> <p>If the tag is <code><a></code> the <i>href</i> tag is used, is the tag <code></code> then the <i>src</i> attribute is used.</p> <p>For this file a hash is computed and saved in the context.</p>	

4.2 Editing

You can edit an XPath expression by clicking on icon  giving:



Edit Record

XPath

//*[@id='price1']

Parameter

Type

Context


#

1

Submit

Cancel



4.3 Deleting

You can delete an XPath expression by clicking on icon  giving you a warning:



4.4 Showing price context

There are two options for testing the XPath expressions:

1. Click on icon  in the "Action" column
You will see the internet page of the respondent to which the XPath expression is applied.
2. Click on icon  in the "Action" column
You will see that part of the respondent's web page to which the XPath expression is applied. If you have clicked on the icon in the last line then that is the price context. If the price context could not be retrieved that the screen shows an error message, see also paragraph 5.1. Below are two examples of what you might get, depending on product, webpage etc.:




5. Price and Price context

The main screen below is the basic screen for editing respondents. This example shows the panel after running the example configuration on fake bike shops produced with the release as an example.

Bikes - Example 1: bikes										
Id	Name	Address	Website	Active	Xpath	VAT	Currency	Date	Last price	Action
				Yes						
+ 1	Awesome Blue	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 179,00	►
+ 2	Awesome Blue	Cheap_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 210,00	►
+ 3	Mothers bike	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 230,99	►
+ 4	Mothers bike	Cheap_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	-1	►
+ 5	Goofy	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	€ 220,00	►
+ 6	Goofy (via click to details)	ABC_Bikes	https://snstatcomp.github.io/web scraping	Yes	3	I	NL (EUR)	2020-04-15 0	-1	►
+ 7	Goofy	Cheap_Bikes	https://snstatcomp.github.io/web scraping	Yes	1	I	NL (EUR)	2020-04-15 0	-1	►

5.1 Collecting new price info for a product group

By clicking on icon  in the top right-hand corner you can start a data collection run for all products in that panel and retrieve the price (and price context) for all respondents of the product group and save it as new observations. It is saved under the date on which you saved the observation. If you succeeded in collecting the price context for a respondent and the price context is the same as in the previous observation, then the price of the previous observation is taken and shown against a green background. If the price context changed or if the collection of the price context failed for a respondent, the price = “-1” is shown against a red background. If the collection of the price info failed, an error message with the reason is saved in the price context:

```
Error: Wait timed out after 5124ms
```

At the end you receive the following message:

Get new priceinfo Electronics: completed

OK

The new observations are recorded in the database. The application shows the new observations in the ‘Last price’ column.

5.2 Examining Price en Price context

By clicking on the price (or “-1”) in the ‘Last price column’ of a respondent you can examine the price and its price context of the observations (up to five of the latest) of the respondent. An example is shown below.

Price and Pricecontext Electronics					
Source: 15552 Afuture - Samsung Galaxy Tab 3 10.1 16GB		20141030	20141029	20141028	20141023
Date	20141030	20141029			Action
Price	€ 272,60	€ 272,60			
Quantity	-1	-1			
Comment					
Context	€ 272,60	€ 272,60			

The default panel shows the two most recent observations. By clicking on the tabs with dates on the top right-hand you can show and hide the other observations so that it is easy to inspect the collection history.

Price and Pricecontext Electronics						x
Source: 11786 De - Apple iPad Air 64GB Wifi		20141030	20141029	20141028	20141023	20141016
Date	20141030	20141029				Action
Price	463,17 €	463,17 €				
Quantity	-1	-1				
Comment						
Context	Apple iPad Air Wi-Fi, iOS 7, A7 Chip und M7 Motion Coprocessor, 64 GB, 24.63 cm (9.7) Retina Display Multi-Touch Multi-Touch (2048 x 1536), FaceTime HD/iSight... 463,17 € Preis gilt nur bei Verwendung des Gutschein-Codes: 7SPAREN (7% Rabatt ab 40 EUR - 10.000 EUR Einkaufswert - Keine Einlösung des Gutscheins auf Bücher, Edelmetalle und Versandkosten. Keine Auszahlung. Nur solange der Vorrat reicht.) Vorkasse, Kreditkarte, Lastschrift, Sofortüberweisung, Paypal, Giropay, Sofortüberweisung: Giropay, 4,90 € inkl. Versand (Gesamtpreis: 468,07 €) *29.10.2014 09:16 / * 3 Tage LocalPCdirect.de 49 Meinungen 48,97% Positive Bewertungen					Apple iPad Air Wi-Fi, iOS 7, A7 Chip und M7 Motion Coprocessor, 64 GB, 24.63 cm (9.7) Retina Display Multi-Touch (2048 x 1536), FaceTime HD/iSight... 463,17 € Preis gilt nur bei Verwendung des Gutschein-Codes: 7SPAREN (7% Rabatt ab 40 EUR - 10.000 EUR Einkaufswert - Keine Einlösung des Gutscheins auf Bücher, Edelmetalle und Versandkosten. Keine Auszahlung. Nur solange der Vorrat reicht.) Vorkasse, Kreditkarte, Lastschrift, Paypal, Giropay, Sofortüberweisung: 4,90 € inkl. Versand (Gesamtpreis: 468,07 €) *29.10.2014 09:16 / * 3 Tage LocalPCdirect.de 49 Meinungen 48,97% Positive Bewertungen

If collecting the price context failed or if the price context has changed, you can adjust the price by editing the field “Price” of the last observation date. You can also edit the Quantity in case of products that sell per quantity. The Comment field allows you to add a remark that should be stored with the observation. All editing can only be done with the most recent observation.


Note: the RobotTool configuration (See Annex B) contains a setting that allows you to switch between overwriting observation for the same day or not.

5.3 Taking the price from the previous response

Apart from adjusting the price (and optionally the quantity) of the last observation by hand as in the previous section, you can also take the price from the previous observation of the respondent. You can do this by clicking on the icon . You can only use this when no price has been entered yet (price = -1).

5.4 Collecting/changing price info respondent

If the price context of a respondent has changed (e.g. the price is in a different place on the respondent’s website) then you may have to change the XPath expressions

(see chapter 4). If collecting the price context failed, you can collect the price context of a respondent collecting by clicking on icon 

Get/Edit priceinfo 2014-10-30 09:30:17

Apple iPad Air Wi-Fi, iOS 7, A7 Chip und M7 Motion Coprozessor, 64 GB, 24.63 cm (9.7") Retina Display Multi-Touch (2048 x 1536), FaceTime HD/iSight... 463,17 € Preis gilt nur bei Verwendung des Gutschein-Codes: 7SPAREN (7% Rabatt ab 40 EUR - 10.000 EUR Einkaufswert - Keine Einlösung des Gutscheins auf Bücher, Edelmetalle und Versandkosten. Keine Auszahlung. Nur solange der Vorrat reicht.) Vorkasse, Kreditkarte, Lastschrift, Sofortüberweisung, Giropay, Paypal: 4,90 € inkl. Versand (Gesamtpreis: 468,07 €) *30.10.2014 09:17 /* */ 3 Tage LocalPCdirect.de 49 Meinungen 48,97% Positive Bewertungen


Save

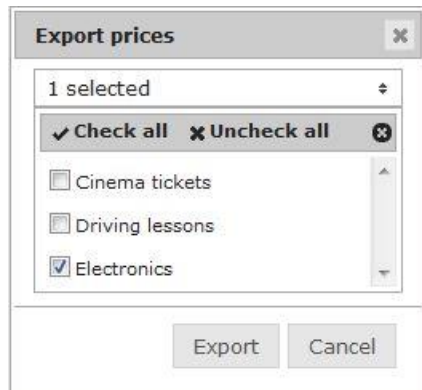
Cancel

By clicking on “Save” you save the price and price context of the last observation of the respondent in the database, overwriting the old values.

14

5.5 Exporting prices

The product group panel (the left panel in the RobotTool user interface) contains a button name 'Export prices'  that you can use to export the most recently observed price information of all active respondents of all active respondents. Depending on your choice of exporting, you can tick all product groups or just one product group. You can change your selection in the following window:



NB. You can select the folder to which prices are saved in the RobotTool configuration file (see Annex B).

Annex A: record description export file

Naam	Omschrijving
product_group	Identification of the product group
source	Identification of the respondent
observation_date	Date/time of the observation
value	Observed price
quantity	Observed quantity
comment	Any comment on the observation
context	Observed price context
user_id	The PID of the observer

Annex B: Configuring the RobotTool

The configuration file of the RobotTool can be found in the folder:

<RobotTool folder>\inst\app\config

The configuration file is called 'config.json'.

The same folder contains a file 'config.json.explanation' with comments that explain the meaning of the different parameters.

Annex C: Learning XPath

For an explanation of the XPath language used in the configuration of products on webpages in the RobotTool, we advise to use one or more of the following resources:

- https://www.w3schools.com/xml/xpath_intro.asp
- <https://devhints.io/xpath>
- <https://blog.scrapinghub.com/2016/10/27/an-introduction-to-xpath-with-examples>
- <https://www.tutorialspoint.com/xpath>