



LINK EXTRACTOR

Manual

TABLE OF CONTENTS

1 APP documentation.....	3
1.1 HOW IT WORKS.....	3
1.2 Input data	4
1.3 Output data	4
1.4 Basic workflow example.....	5
2 API documentation.....	7

1 APP DOCUMENTATION

1.1 HOW IT WORKS

Link Extractor APP is an acquisition task that enables to find urls from a web.
User can specify the deep level of search into the web site pages tree.

1.2 INPUT DATA

This app allows You to specify the url from which start the link search and the depth of search (number of depth levels).

Example: If You specify depth=2 then the APP search all url contained in the specified page and all urls contained in pages referring this first level of url.

You can see an example of input data in the following figure:



Link Extractor Task
App Name
url
depth

Figure 1 - example of input data.

1.3 OUTPUT DATA

This app gives a set of two elements back:

- a. depth_level: specify the depth level of the found url;
- b. link: url found.

This set of elements is given back in an excel spreadsheet or a data cube whose metadata depending on the data extracted and presented in the output.

You can see an example of output data in the following figure:

[-] (2 Items)	
depth_level	1
link	http://www.techopedia.com/topic/226/personal-tec
[-] (2 Items)	
depth_level	1
link	http://www.techopedia.com/it-terms/v
[-] (2 Items)	
depth_level	1
link	http://www.techopedia.com/it-terms/k
[-] (2 Items)	
depth_level	1
link	http://www.techopedia.com/topic/206/windows-8

Figure 2 - example of output data

1.4 BASIC WORKFLOW EXAMPLE

Link Extractor APP doesn't need any APP before in the workflow, so you can create a simple flow as shown in the following Figure 3.

You can set input parameter and watch the results by the watcher button as shown in the following Figure 4.

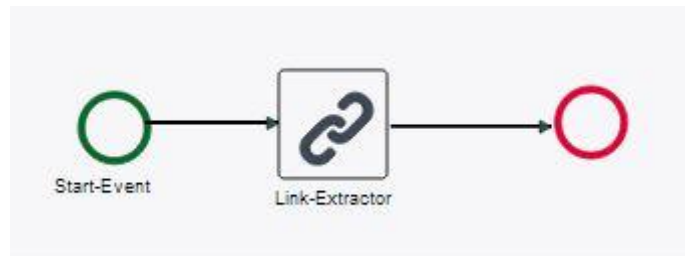


Figure 3 - Link Extractor workflow example

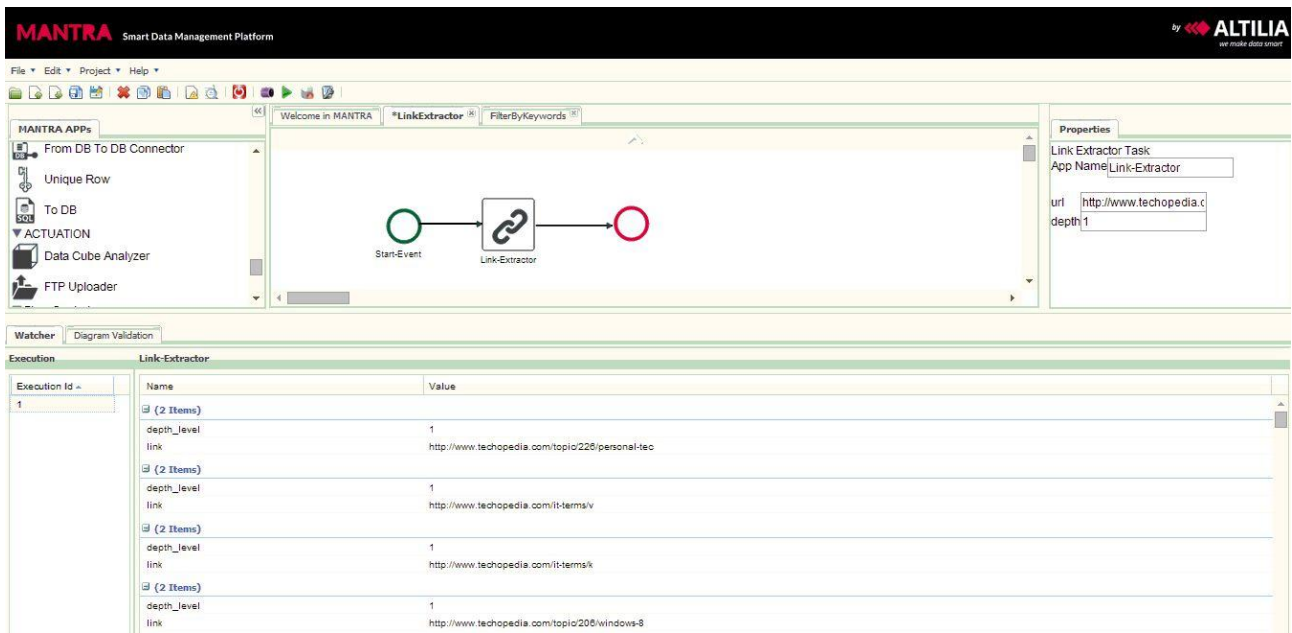


Figure 4 - the watcher panel.

2 API DOCUMENTATION

For information about how to use Link Extractor API in your application, send us a message to info@altiliagroup.com.