

## Website Input - Feature #1168

### Output raw data

01/19/2016 05:29 PM - Luke Murphey

<b>Status:</b>	Closed	<b>Start date:</b>	05/18/2016
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Luke Murphey	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	2.1		
<b>Description</b> Add the ability to just output the raw content of the page that is matched. The user can then parse the content themselves in SPL (such as rex). This way, the user can import things such as: <ul style="list-style-type: none"><li>• JSON files</li><li>• HTML that is not valid or is rendered by Javascript (especially document.write() calls).</li><li>• Raw content that needs no parsing (like the Internet Storm Center status)</li></ul>			
<b>Subtasks:</b> Task # 1343: Allow empty selector on input page if raw content is included			
			<b>Closed</b>
<b>Related issues:</b> Related to Website Input - Feature #1748: Add parsing of JSON fields			
			<b>New</b>
			<b>02/16/2017</b>

### History

#### #1 - 01/19/2016 05:29 PM - Luke Murphey

- Description updated

#### #2 - 01/19/2016 06:21 PM - Luke Murphey

<https://answers.splunk.com/answers/341483/website-input-another-device-to-scrape-information.html>

#### #3 - 05/04/2016 04:26 PM - Luke Murphey

Here are some issues that would need to be addressed to get this to work:

1. how to handle source-typing.
  1. For XML, it would need to output raw XML so that xpath could be used (<http://docs.splunk.com/Documentation/Splunk/6.0.7/SearchReference/Xpath>) and KV\_MODE=xml
  2. For JSON, the sourcetype would need to be such that Splunk would treat the content as JSON (using INDEXED\_EXTRactions = json)
2. How to handle the extra fields that are usually included as key-value pairs

#### #4 - 05/04/2016 04:27 PM - Luke Murphey

spath: <http://docs.splunk.com/Documentation/Splunk/6.2.3/SearchReference/Spath>

#### #5 - 05/14/2016 07:48 PM - Luke Murphey

Not sure how to output raw data and define sourcetype, index

#### #6 - 05/18/2016 06:30 AM - Luke Murphey

I currently have to strip endlines. To include endlines, I would need to configure a better line-break (<http://docs.splunk.com/Documentation/Splunk/6.2.2/Data/Indexmulti-lineevents>).

#### #7 - 05/19/2016 10:00 PM - Luke Murphey

- Status changed from New to Closed

**#8 - 03/06/2017 09:05 PM - Luke Murphey**

- Related to Feature #1748: Add parsing of JSON fields added