Website Input - Bug #2807

Results don't retain original sorting

07/03/2020 10:49 PM - Luke Murphey

Status:	Closed	**Start date:**	07/03/2020
Priority:	Normal	**Due date:**	
Assignee:	Luke Murphey	**% Done:**	100%
Category:	Input: Web Spider	**Estimated time:**	0.00 hour
Target version:	4.5.10		

Description
See https://community.splunk.com/t5/forums/replypage/board-id/apps-add-ons-all/message-id/62252

Associated revisions

Revision 762 - 07/04/2020 04:59 PM - luke.murphey

Fixing issue where results got reordered

Closes #2807

History

#1 - 07/03/2020 10:51 PM - Luke Murphey

The preview works correctly, as does the search command.

Reproduction:

- Input with URL of https://www.w3schools.com/html/html_tables.asp and selector of tr

This works:

```
| webscrape selector="tr" url="https://www.w3schools.com/html/html_tables.asp" depth_limit=25 empty_matches=0
```

#2 - 07/04/2020 12:08 AM - Luke Murphey

Qs:

- Where is the ordering getting messed up?
- Does this work if I output without using the stash writer?
    - No change
- Does this only happen when output_fx is used?
    - No change
- What is the relevant stack?
    - web_input.py::output_results
    - web_scraper.py::scrape_page
    - web_scraper.py::get_result_single

Obs:

- test_scrape_page_mv might be able to detect this problem
    - Ordering looks right here actually
- get_result_single uses an OrderedDict
- The result is correct in get_result_single() but not in output_results()

#3 - 07/04/2020 05:00 PM - Luke Murphey

- Status changed from New to Closed

- % Done changed from 0 to 100