# Quick Guide for VietSpider

Step-by-step Guide to VietSpider Running and Maintaining

XML Vietspider running and maintaining for Windows version allows administrators to crawl web data, index data and publish result for your own business or organizations. This quick guide explains how to use Create a crawled channel to take data from an exist URL.

How to get web data from a website.

First look of Vietspider
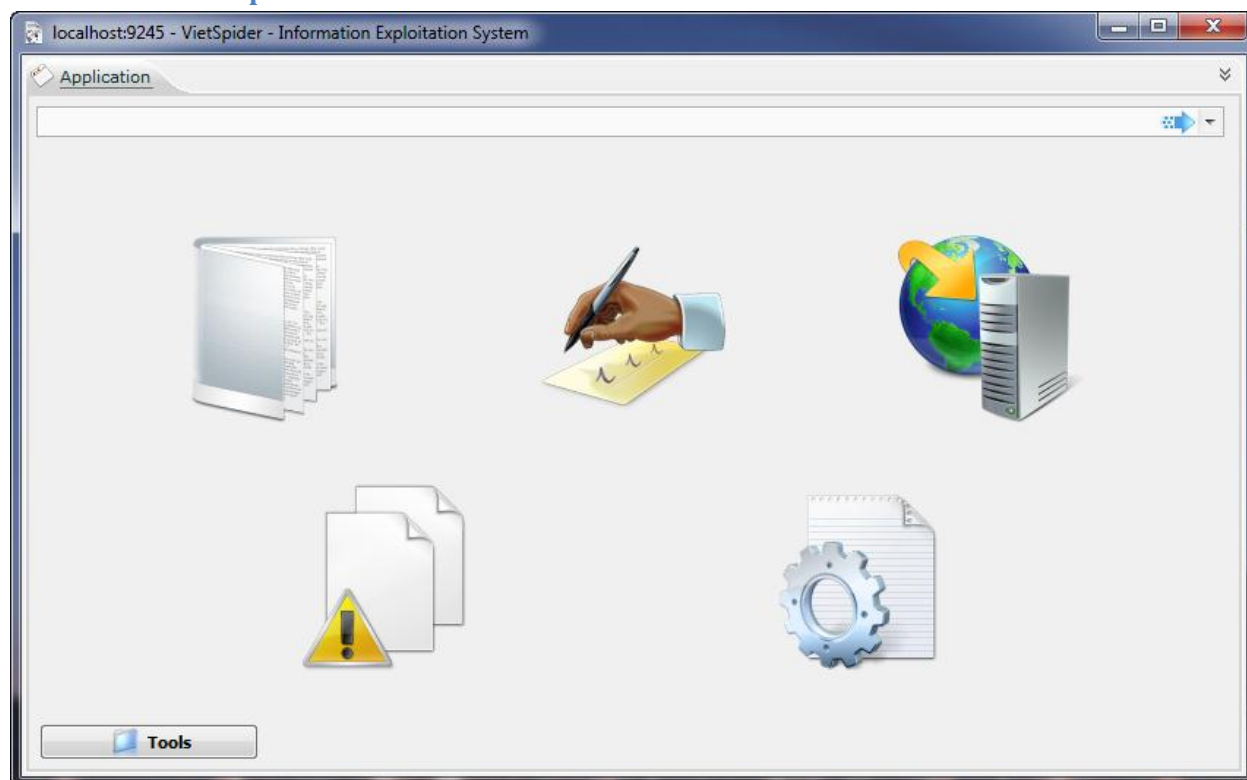
Administrators follow these steps to get web data step-by-step:

1. Select icon   to Create a new channel. The interface of Create New Channel as shown below:
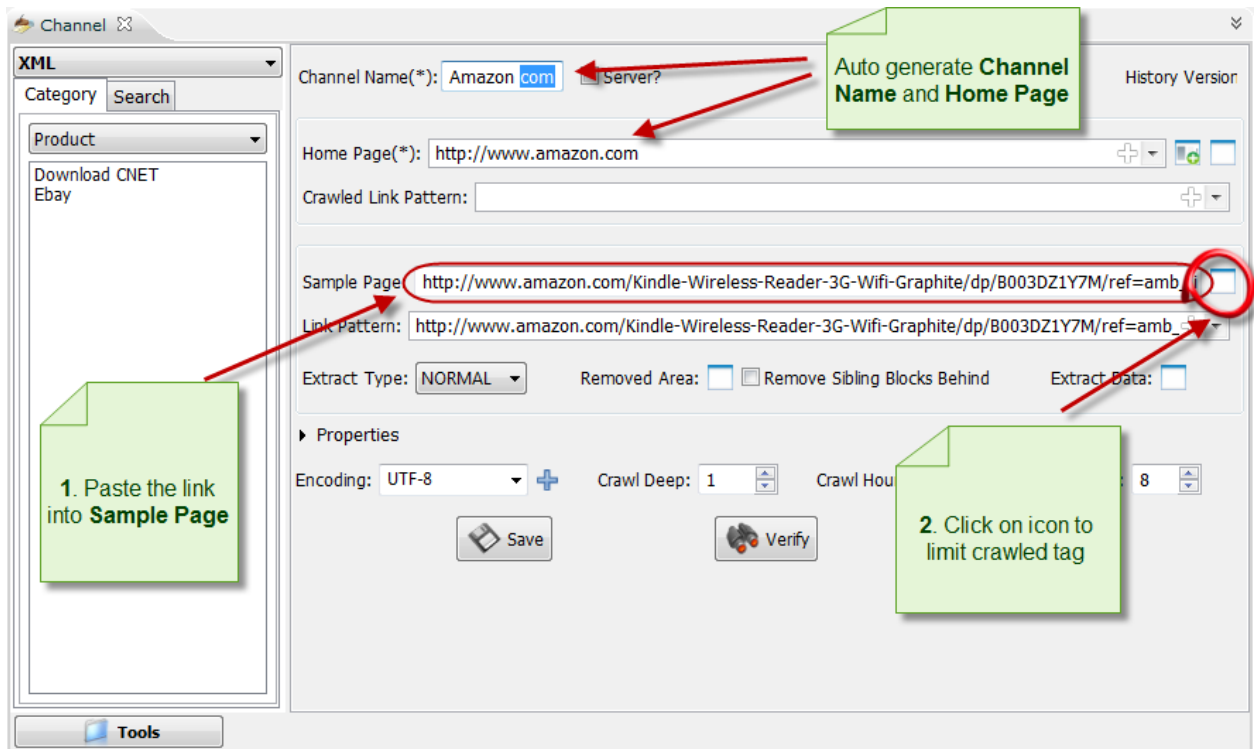

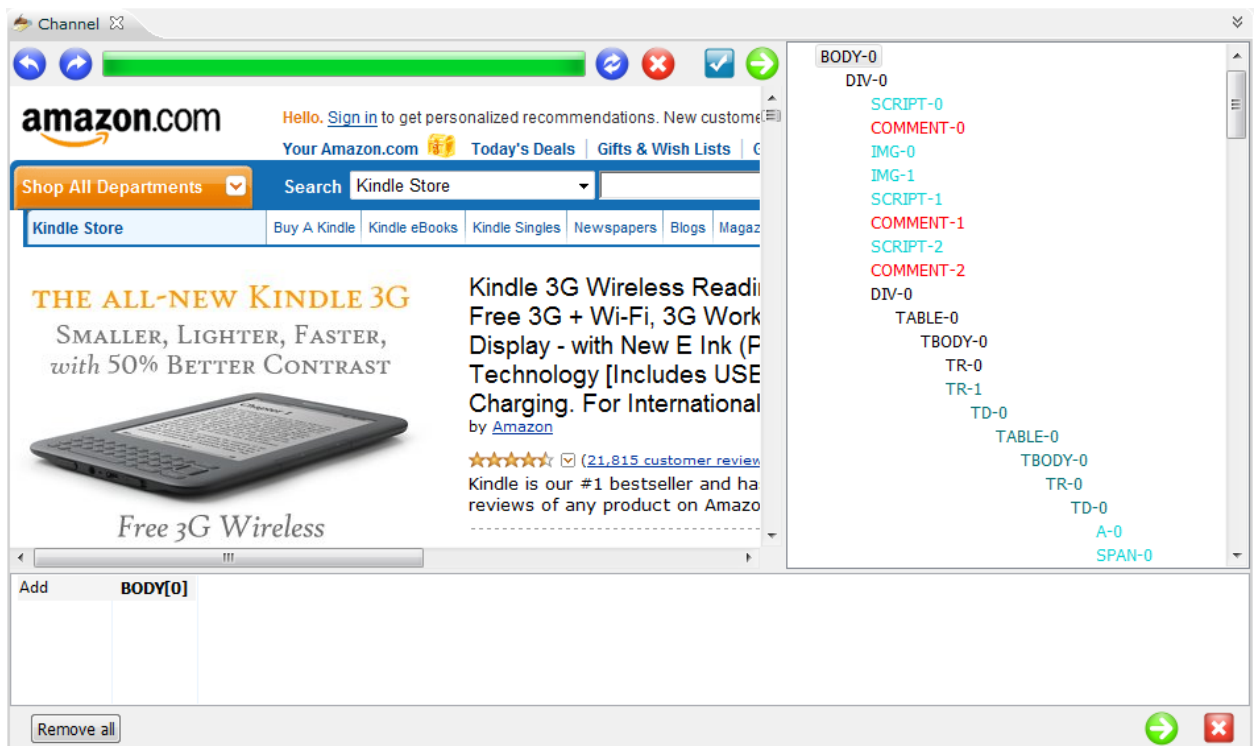2. Go to a website, take example, this case we use Amazon Kindle

Copy the link of Kindle on Address bar, it looks like this:

http://www.amazon.com/Kindle-Wireless-Reader-Wifi-Graphite/dp/B003DZ1Y8Q/ref=amb_link_355368562_2?pf_rd_m=ATVPDKIKX0DER&pf_rd_s=center-1&pf_rd_r=07YFZNAJCGRFF6CXDSNQ&pf_rd_t=101&pf_rd_p=1289229502&pf_rd_i=507846

3. And then, paste into **Sample Page** on Channel Tab of VietSpider:

Channel ✕

XML ▾

Category | Search

Product ▾

Download CNET
Ebay

Channel Name(*): Amazon **com** □ Server?

Auto generate **Channel Name** *and* **Home Page**

History Version

Home Page(*): http://www.amazon.com

Crawled Link Pattern:

Sample Page: http://www.amazon.com/Kindle-Wireless-Reader-3G-Wifi-Graphite/dp/B003DZ1Y7M/ref=amb_

Link Pattern: http://www.amazon.com/Kindle-Wireless-Reader-3G-Wifi-Graphite/dp/B003DZ1Y7M/ref=amb_

Extract Type: NORMAL ▾   Removed Area: □ □ Remove Sibling Blocks Behind   Extract Data: □

1. Paste the link into **Sample Page**

2. Click on icon to limit crawled tag

▸ Properties

Encoding: UTF-8 ▾ ✚   Crawl Deep: 1 ⇕   Crawl Hou   8 ⇕

Save   Verify

Tools

4. When you click on icon ☐ The following interface will appear.

Channel ✕

BODY-0
  DIV-0
    SCRIPT-0
    COMMENT-0
    IMG-0
    IMG-1
    SCRIPT-1
    COMMENT-1
    SCRIPT-2
    COMMENT-2
    DIV-0
      TABLE-0
        TBODY-0
          TR-0
          TR-1
            TD-0
              TABLE-0
                TBODY-0
                  TR-0
                    TD-0
                      A-0
                      SPAN-0

Add   BODY[0]

Remove all

5. On browse web interface, you select text (**1**) from title of Kindle, Vietspider will automatically detect (**2**) which tag belong to on **Tree tag** on the right. Then, you right click and select Add block (**3**). Finally, you got position of title (**4**).
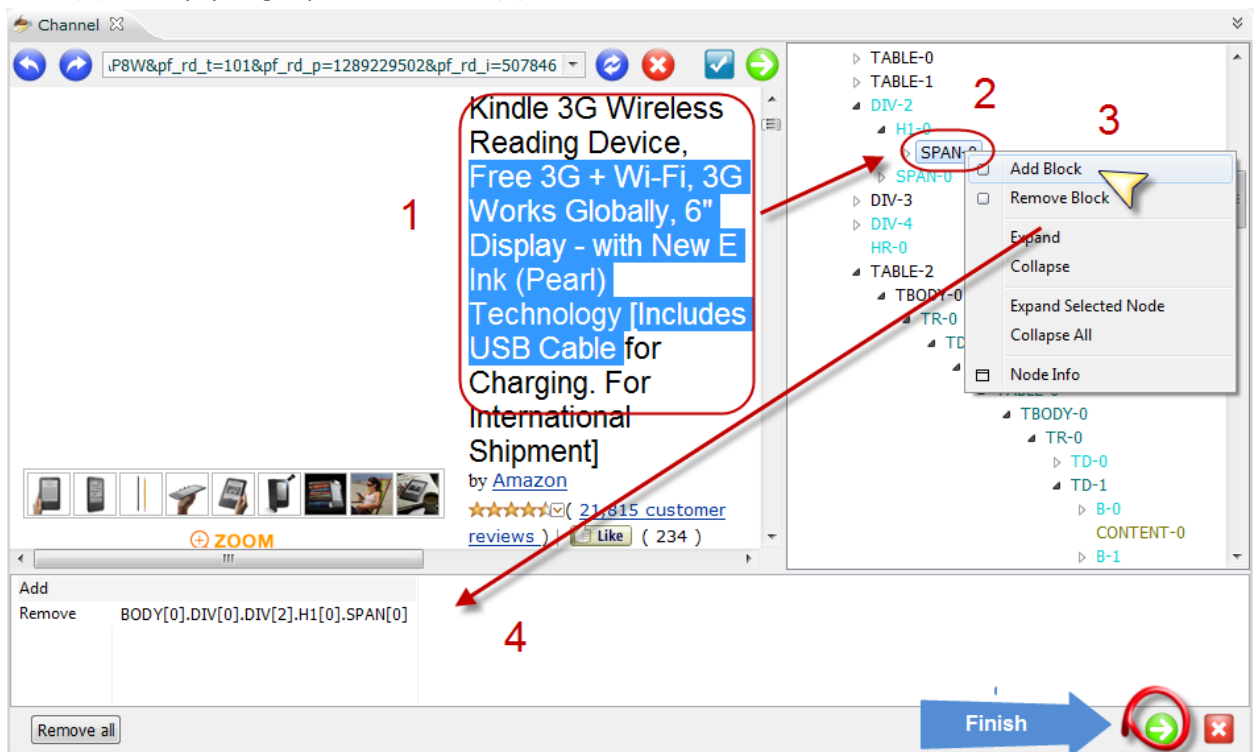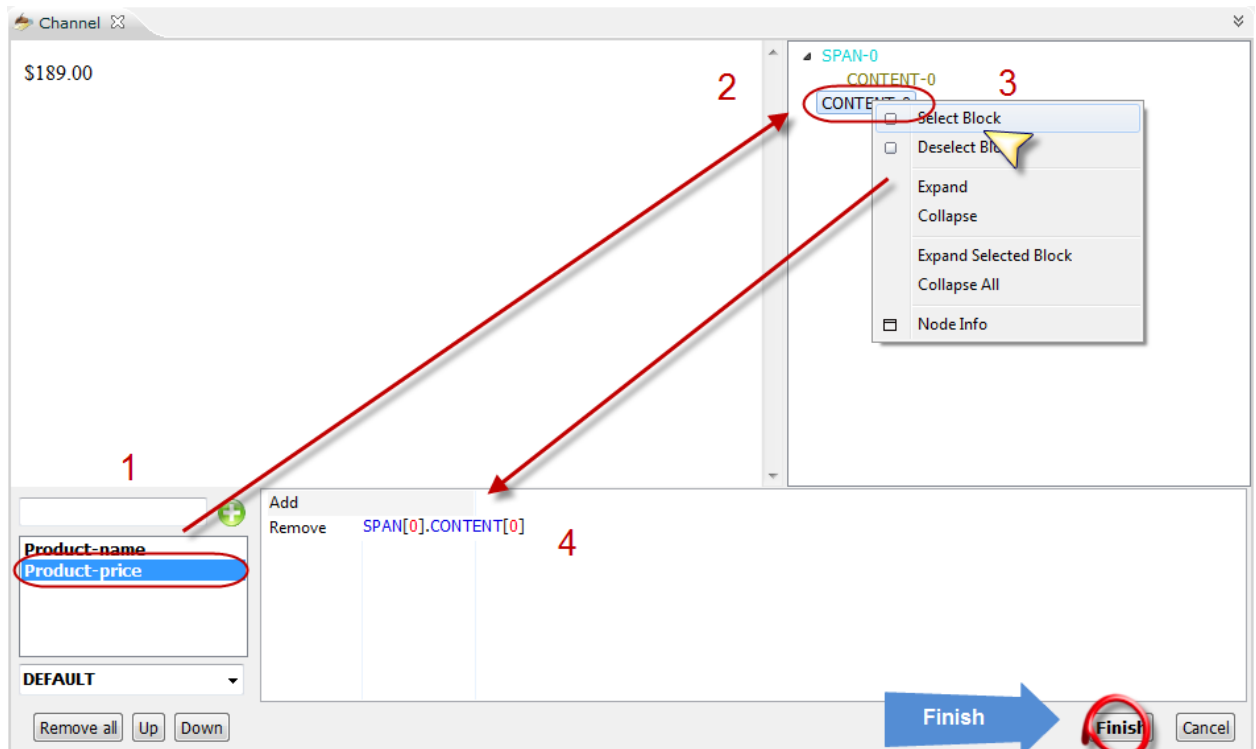

6. After finishing, click on to finish. We will get back to **Create New Channel** interface. Please hit as following to select exact data to get.

7. If you click, you will see the interface as bellow:

8. Please write field that you want to get.
   In this field we type: **Product-name, Product-price** to put Title and price of Kindle.

9. Select **Product-price** and choose **Content [0]** on **Tree Tag** , then hit **Select Block**, finally we got position for **Product-price**. As same with **Product-name**. Click **Finish** to save configuration. Alright, almost done.

10. Click on [Verify] icon to check correct configuration of this channel.

And, now we got data look like this:

```
Channel ⊠                                                        ≫
  <?xml version="1.0" encoding="utf-8" ?>
 - <document>
     <Product-price>$189.00</Product-price>
     <Product-name>Kindle 3G Wireless Reading Device, Free 3G + Wi-Fi, 3G
       Works Globally, 6" Display - with New E Ink (Pearl) Technology [Includes
       USB Cable for Charging. For International Shipment]</Product-name>
 - <src>
     <![CDATA[ http://www.amazon.com/Kindle-Wireless-Reader-3G-Wifi-
       Graphite/dp/B003DZ1Y7M/ref=amb_link_355368562_4/176-5847544-3545461?pf_r
     </src>
 </document>
```

http://www.amazon.com/Kindle-Wireless-Reader-3G-Wifi-Graphite/dp/B003DZ1Y7M/ref=amb_link_35 **Back**
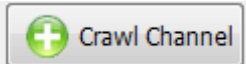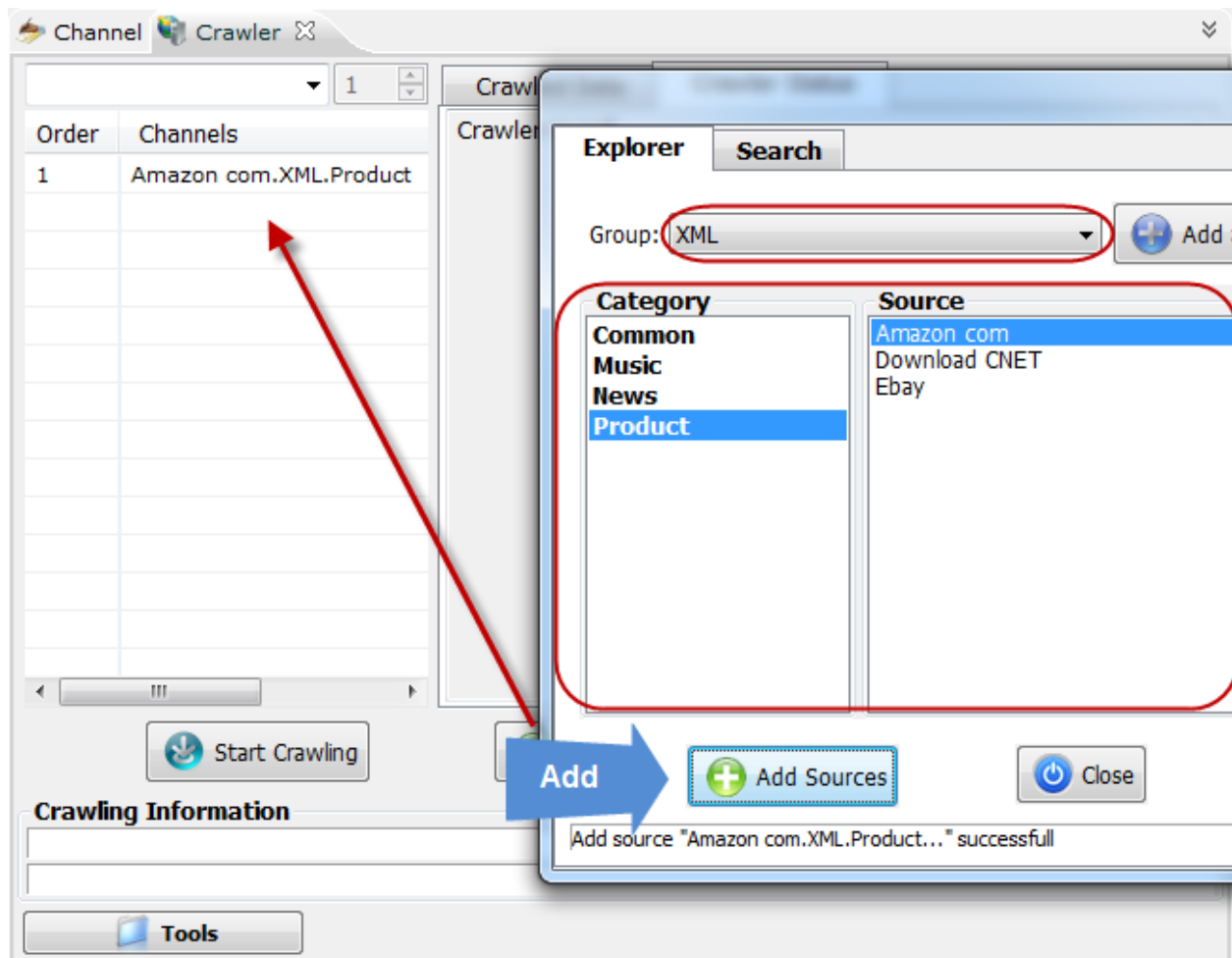
11. If nothing wrong, please hit **Back**, and click on [Save] icon in the previous window to Save all information of this channel.

12. Now go to **Tools**, open **Crawler** by clicking on **Crawler** , the interface will be displayed as bellow:
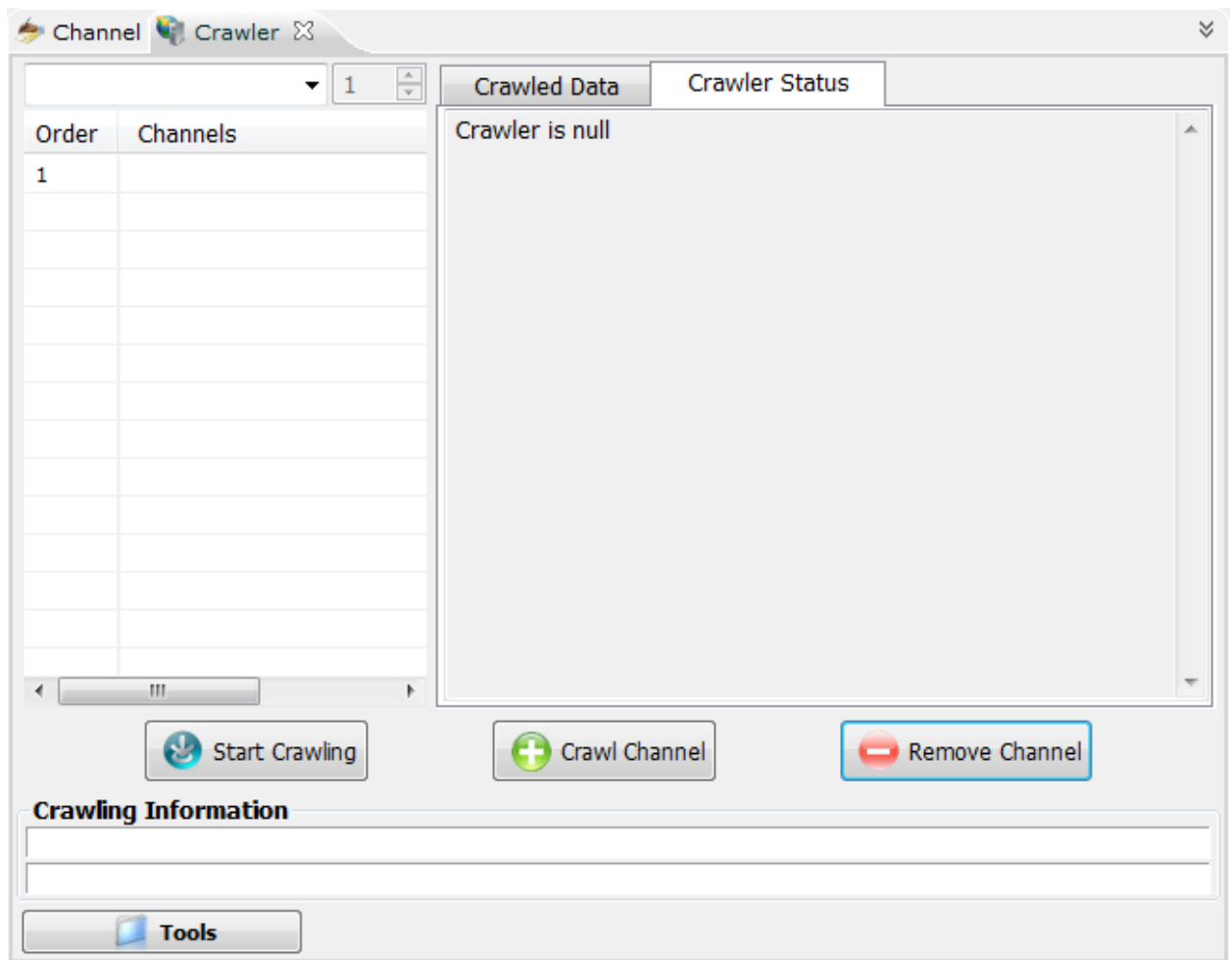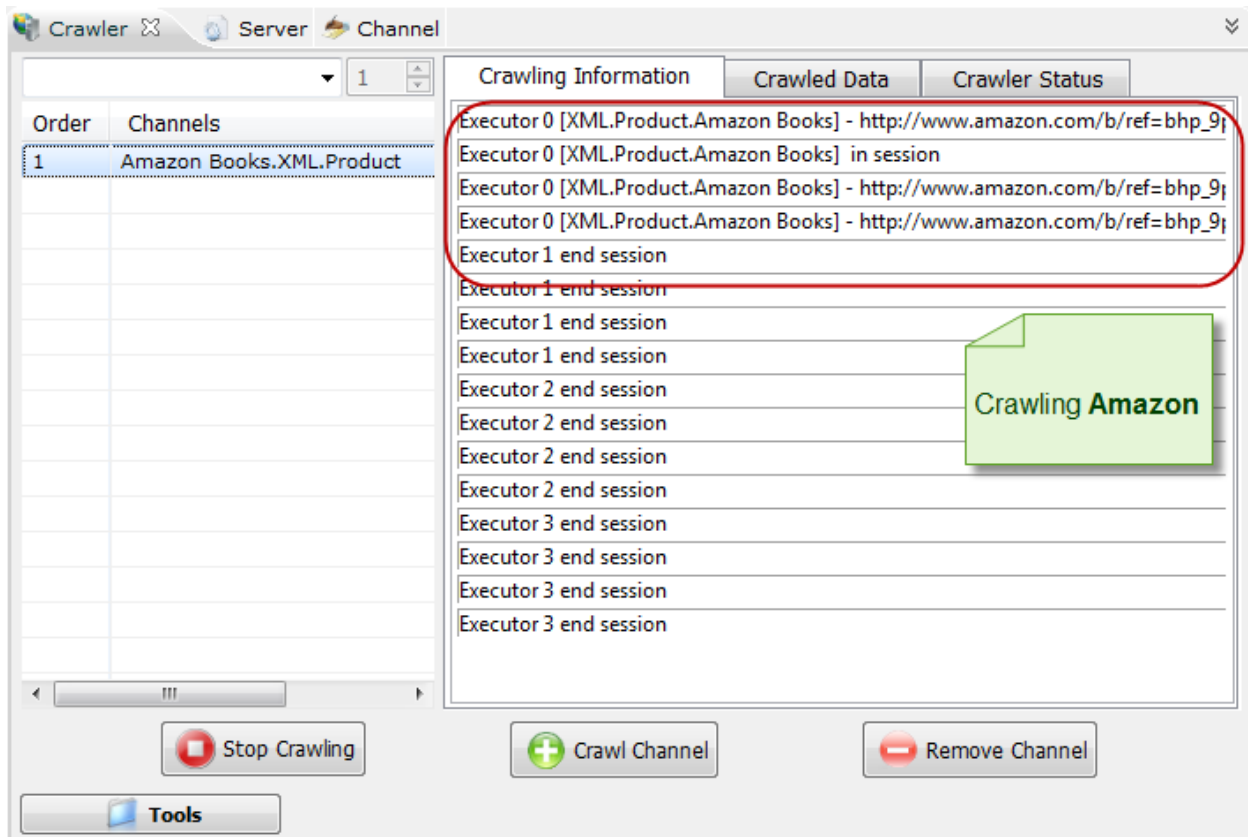
13. Please click on to add channels from channel list.
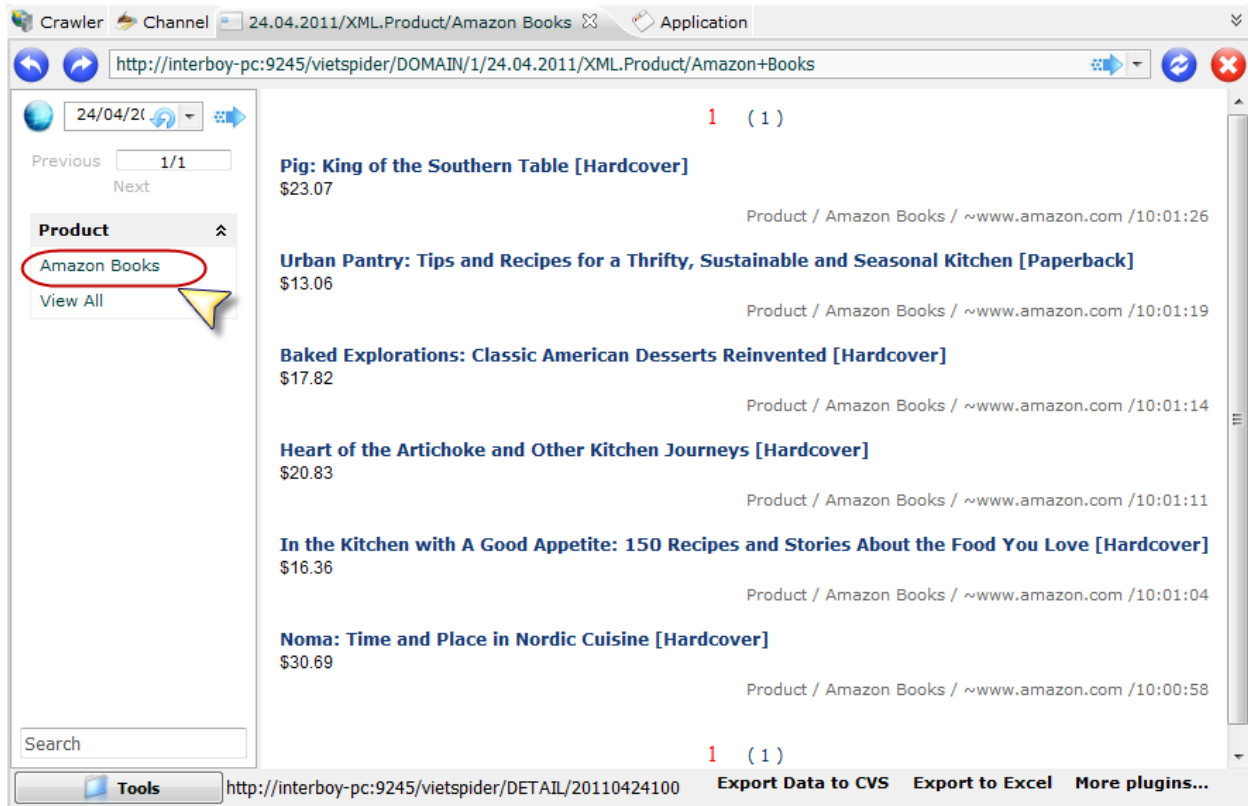14. You will see the interface like this:

15. On **Group** field, select **XML**, choose *Product* in **Category** and take *Amazon com* in **Source**, then click **Add Sources** to add to **Crawled list.**
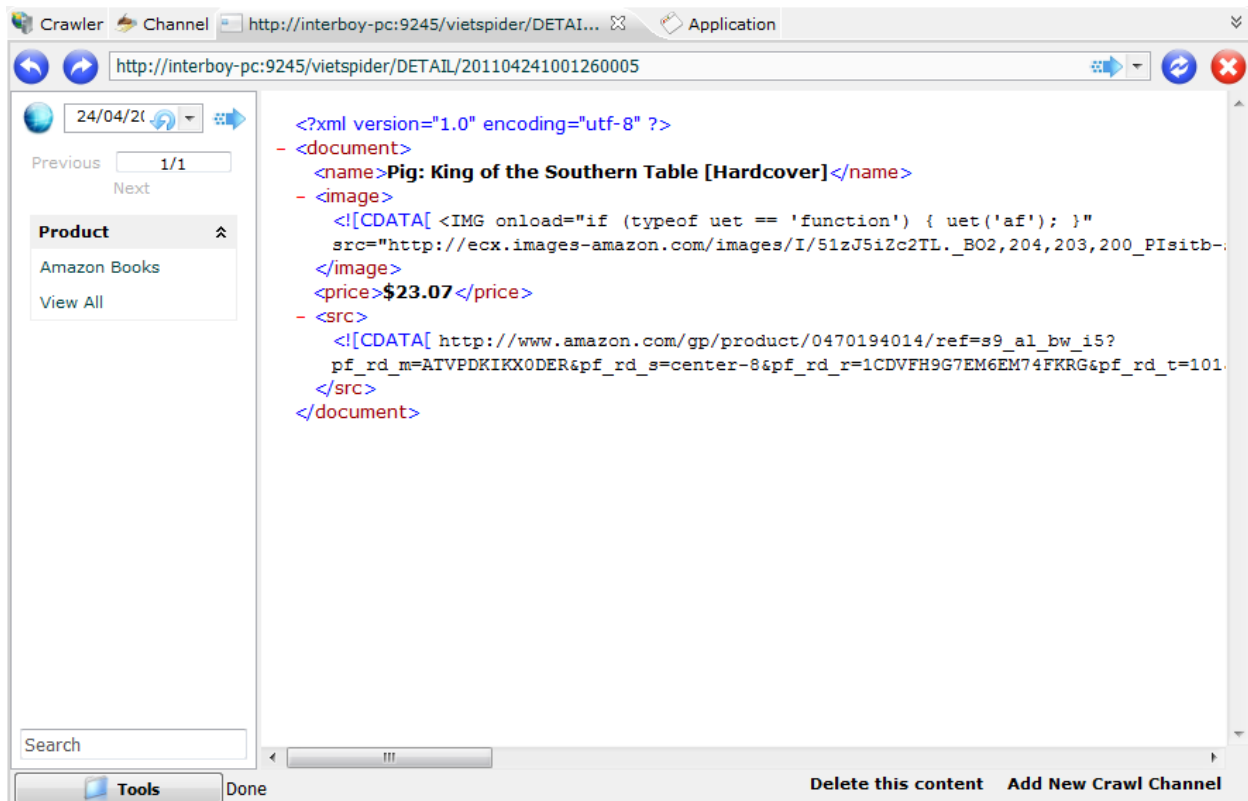
16. Then click on to start crawling data.

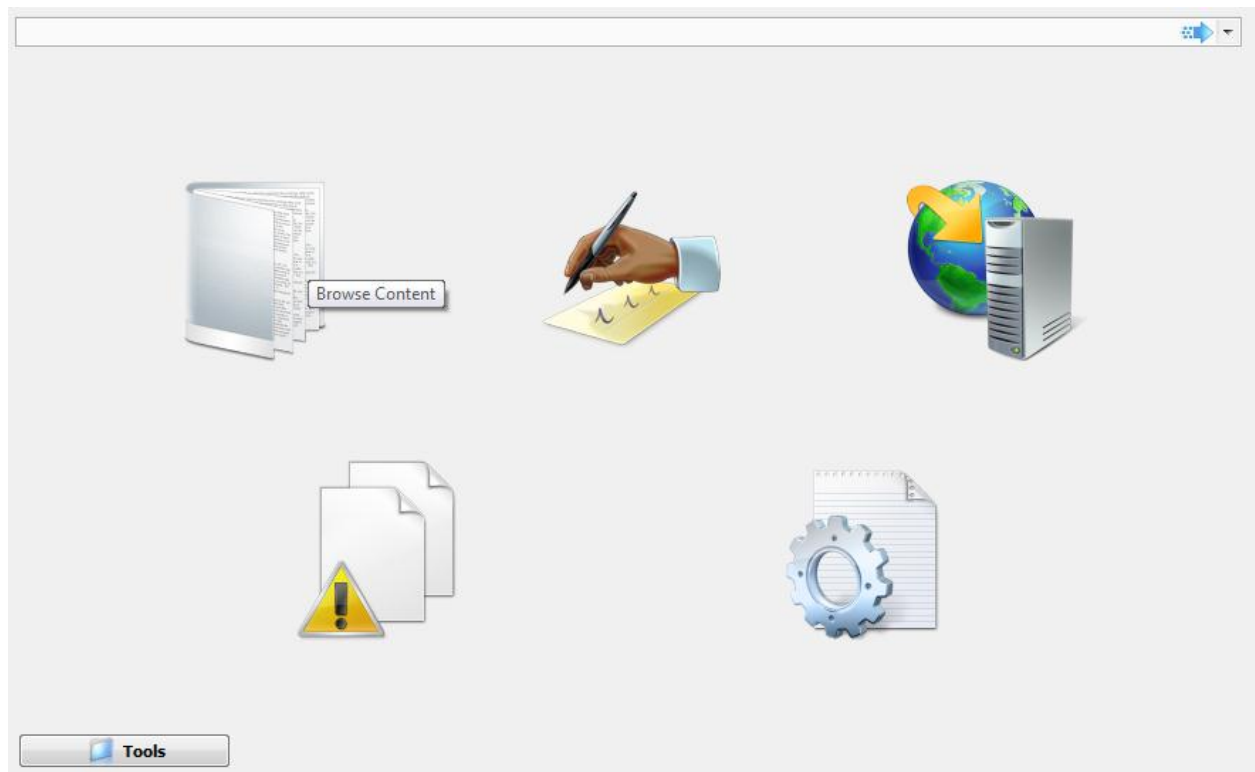The interface when crawling :

Crawling **Amazon**

17. Alright, data now in database, please hit in main interface to browse what we got.

18. Many product on our database, click on a product to view detail:

Get back to main interface.

That comes to the end of this quick guide for Vietspider. If you have and question related to this, please feel free to send email to nhudinhthuan@yahoo.com .