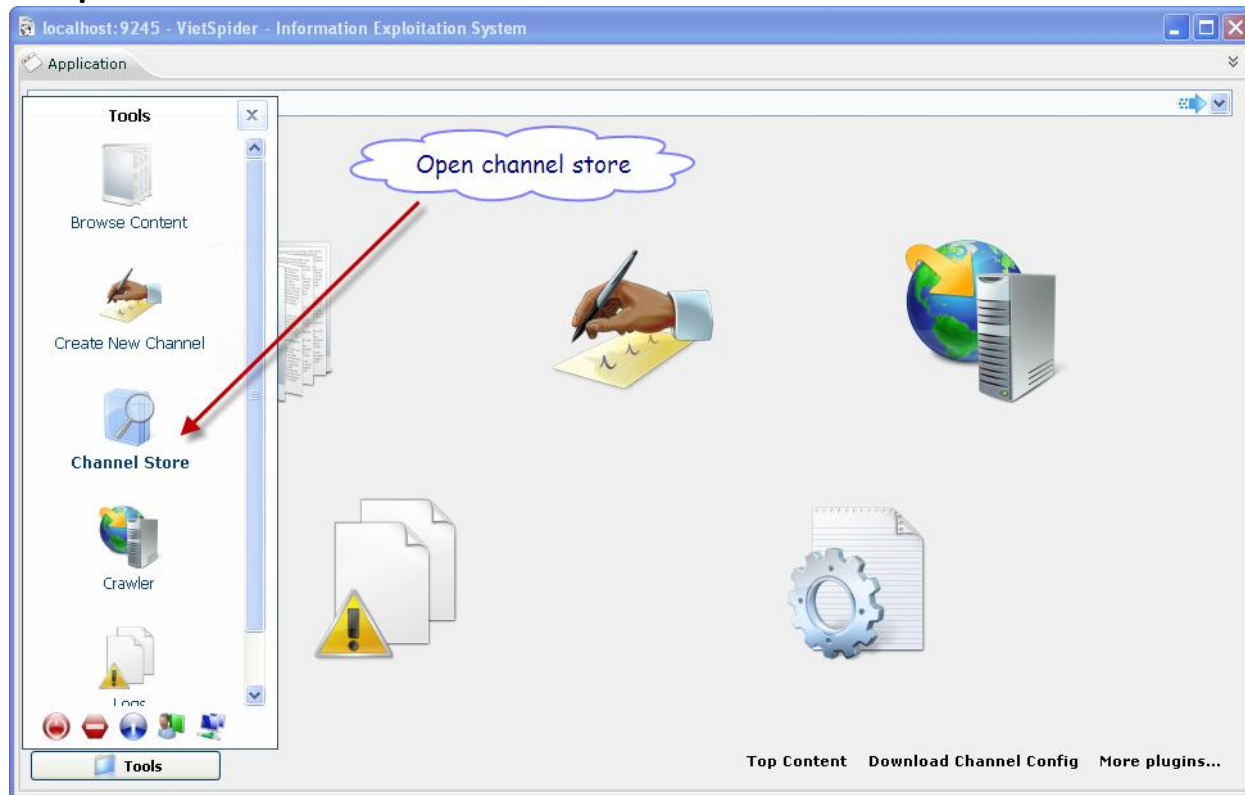


## VietSpider Quick Guide – Step by Step – create channel by Channel Store

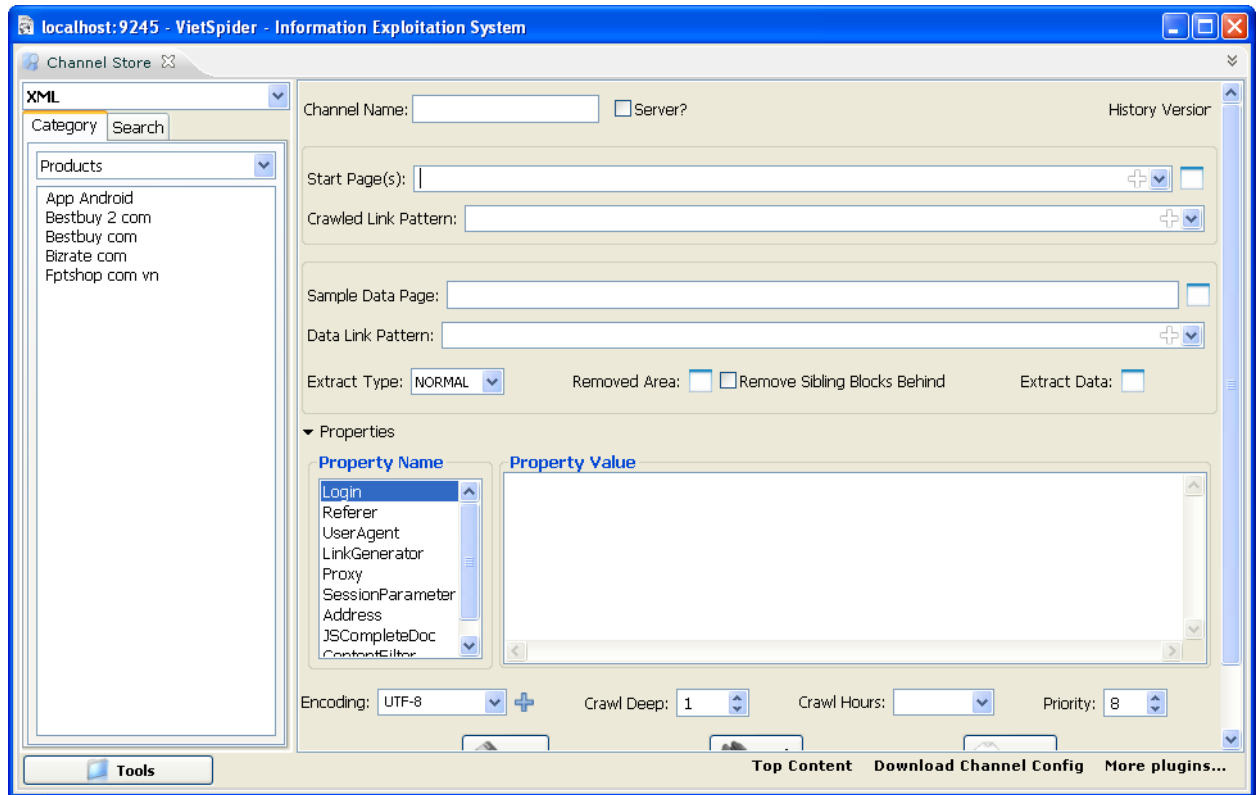
### 1. Launch VietSpider

Goto **VietSpider** folder and double click on **VietSpider.exe**

### 2. Open Channel Store

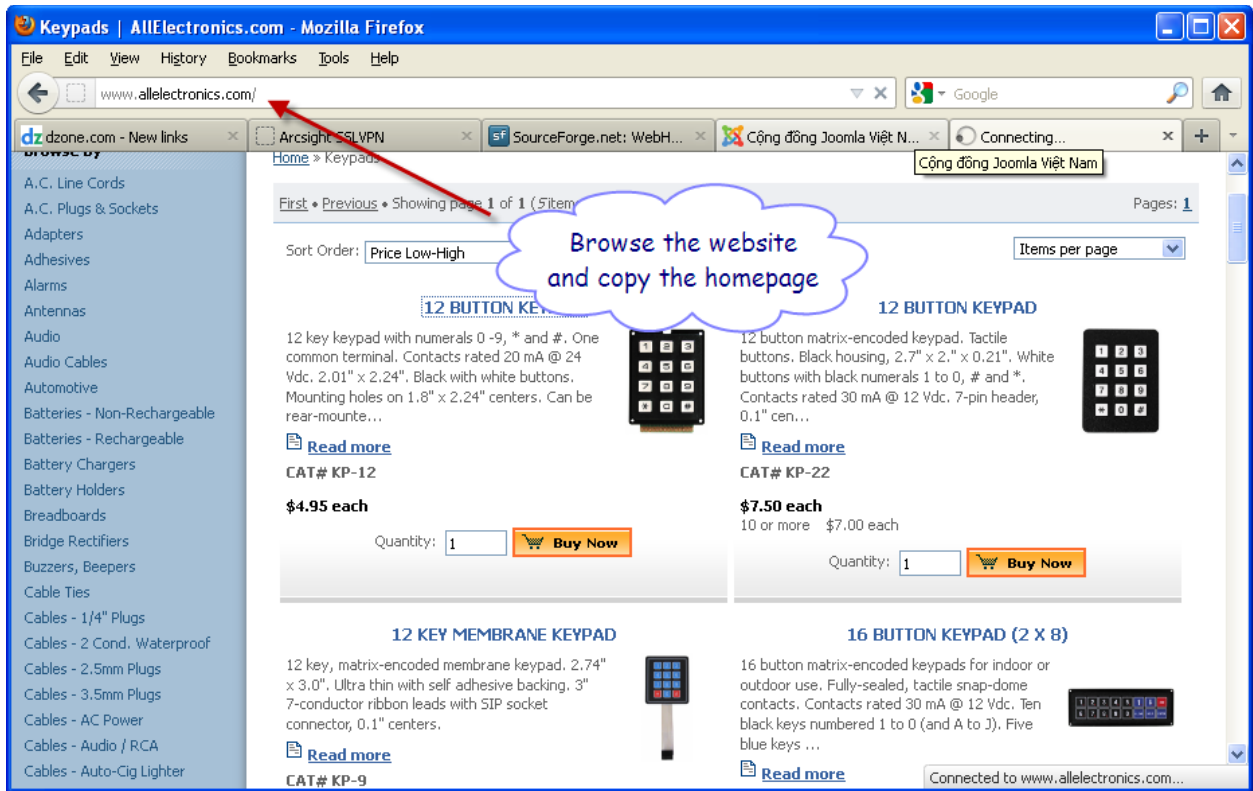


Select **Tools** -> **Channel Store**

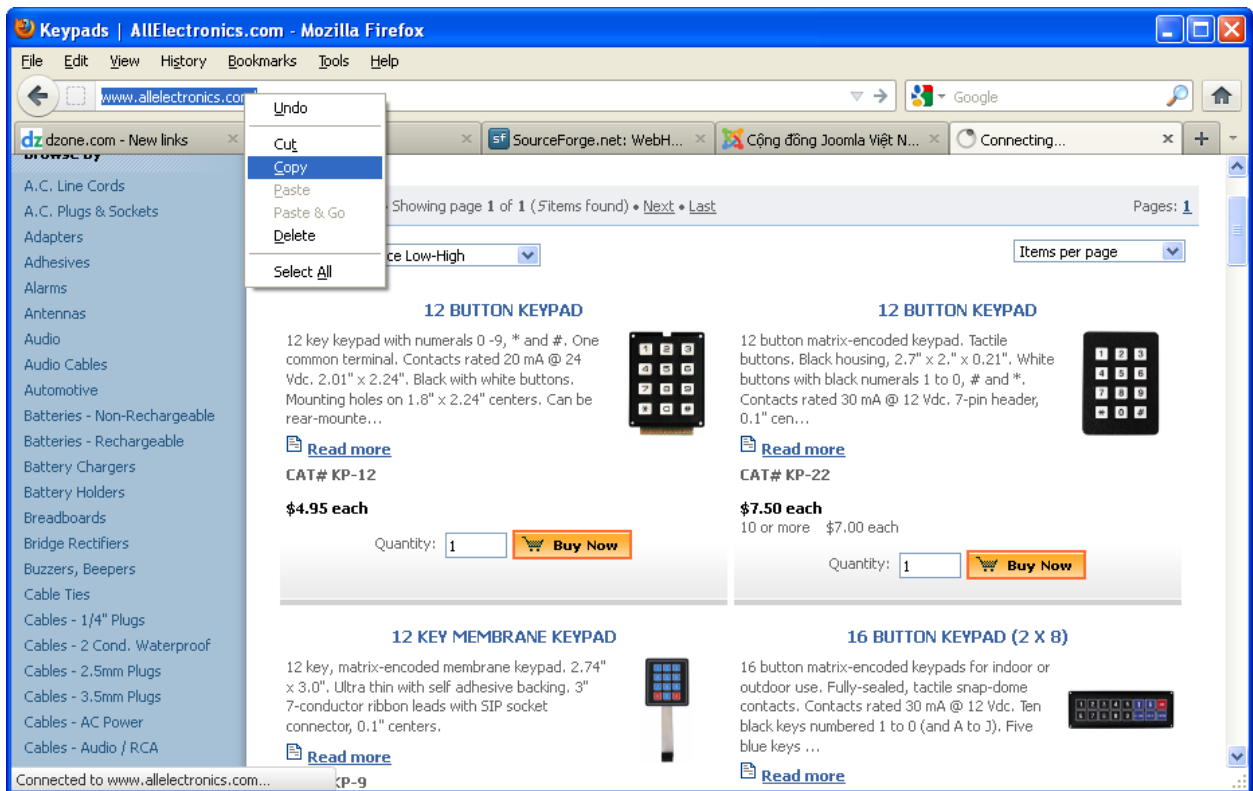


### 3. Add Start Page(s)

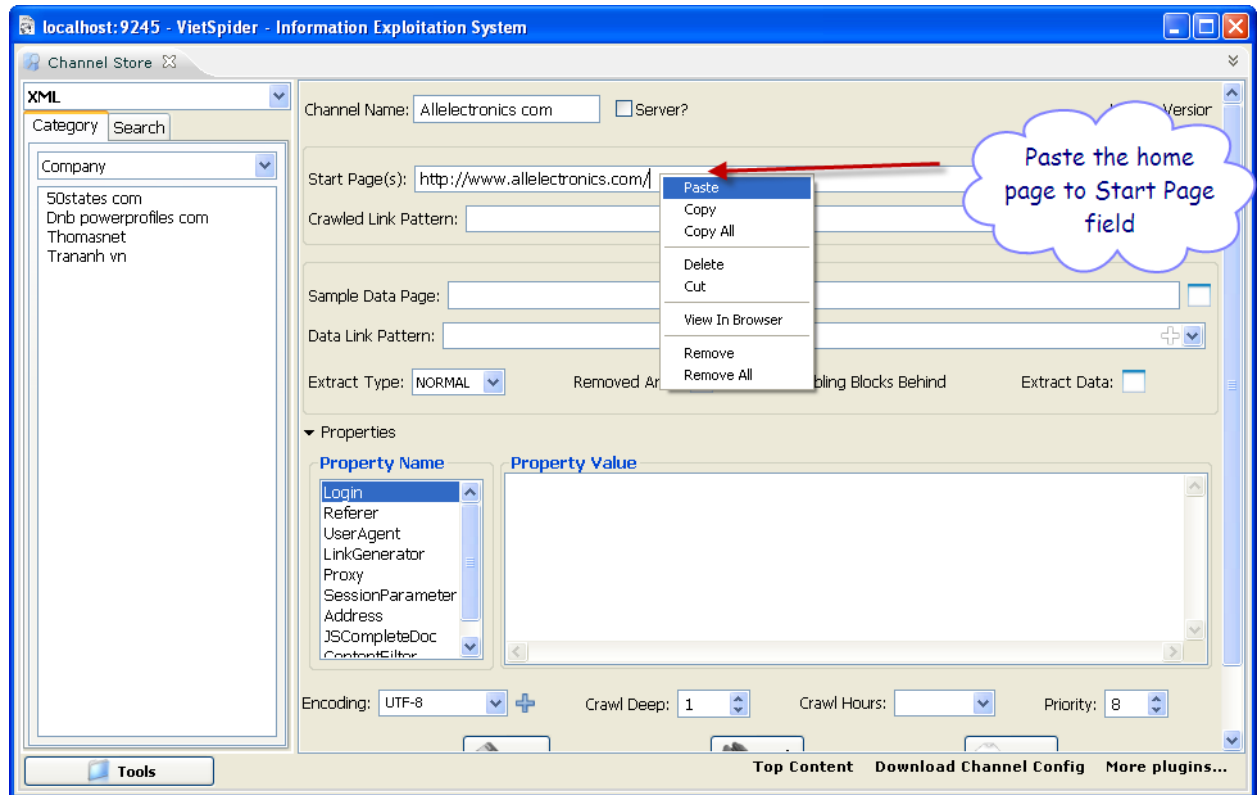
Browse the website by your browser



And copy the home page

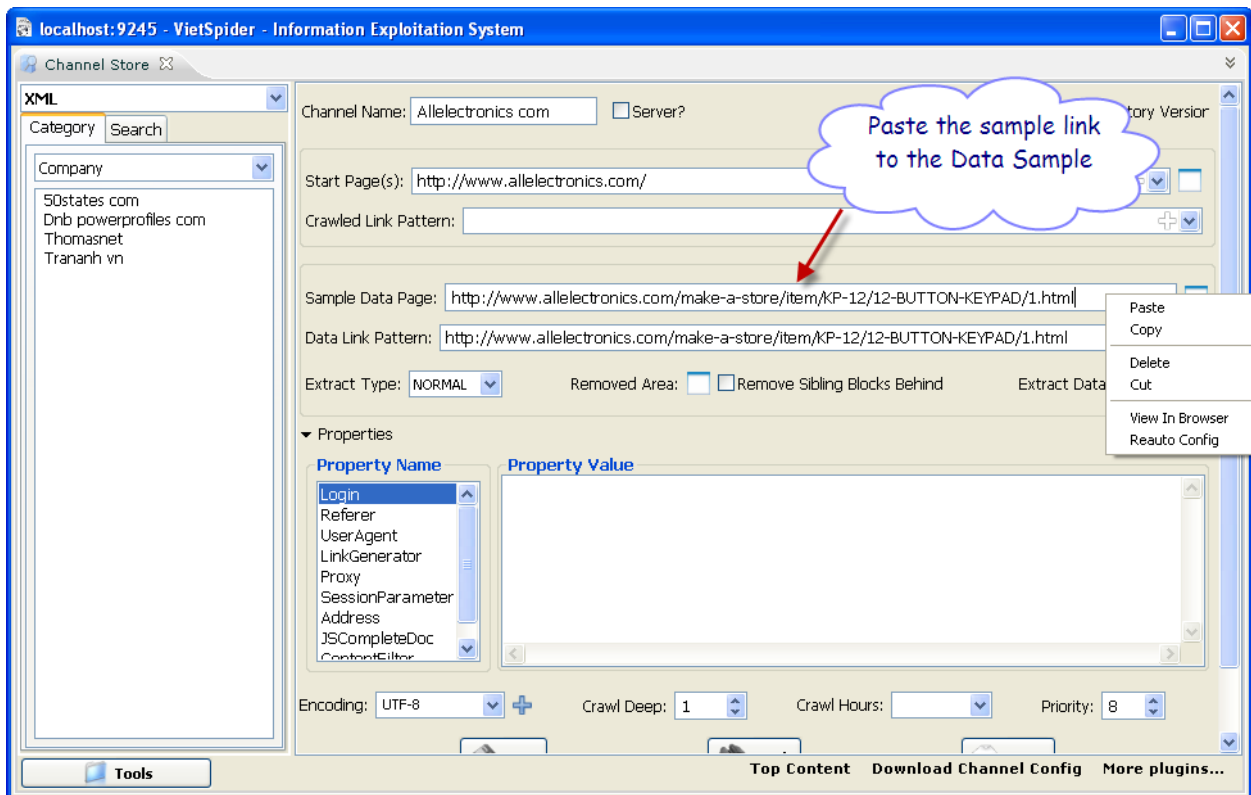
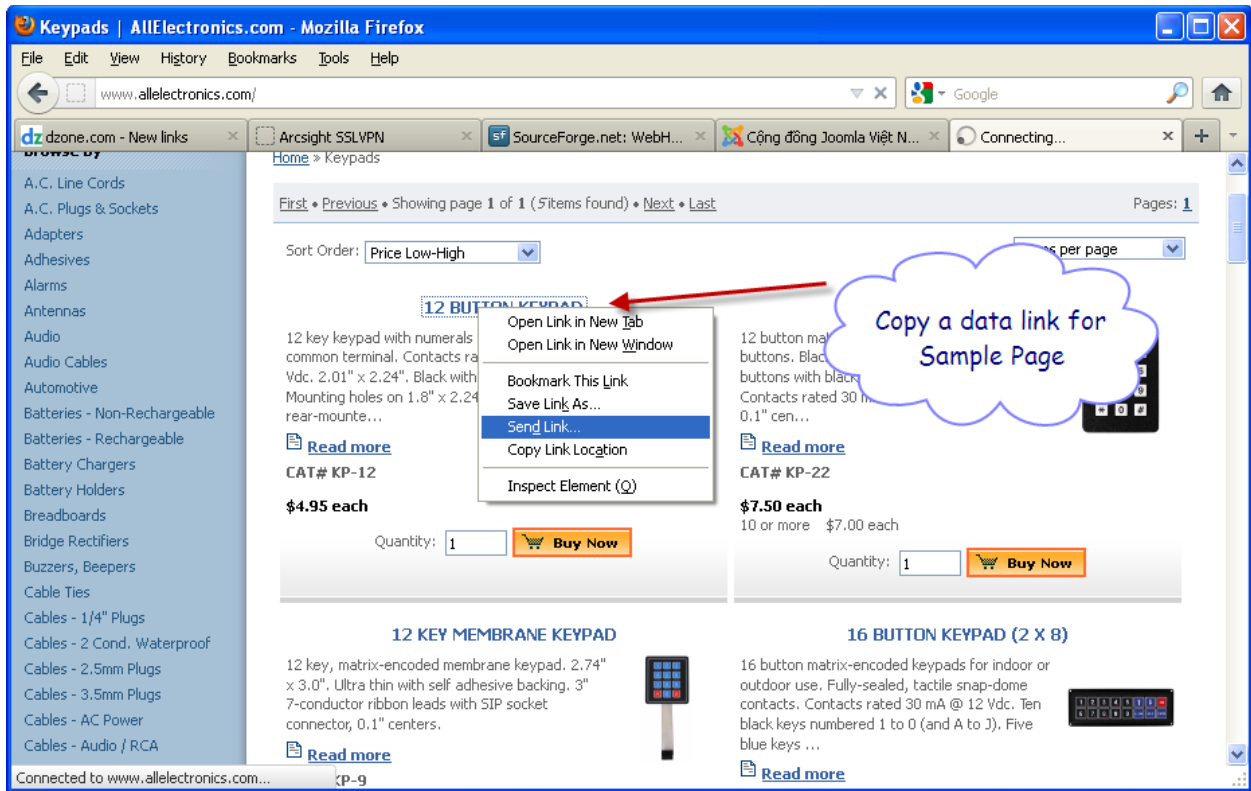


Paste the home page to **Start Page** field

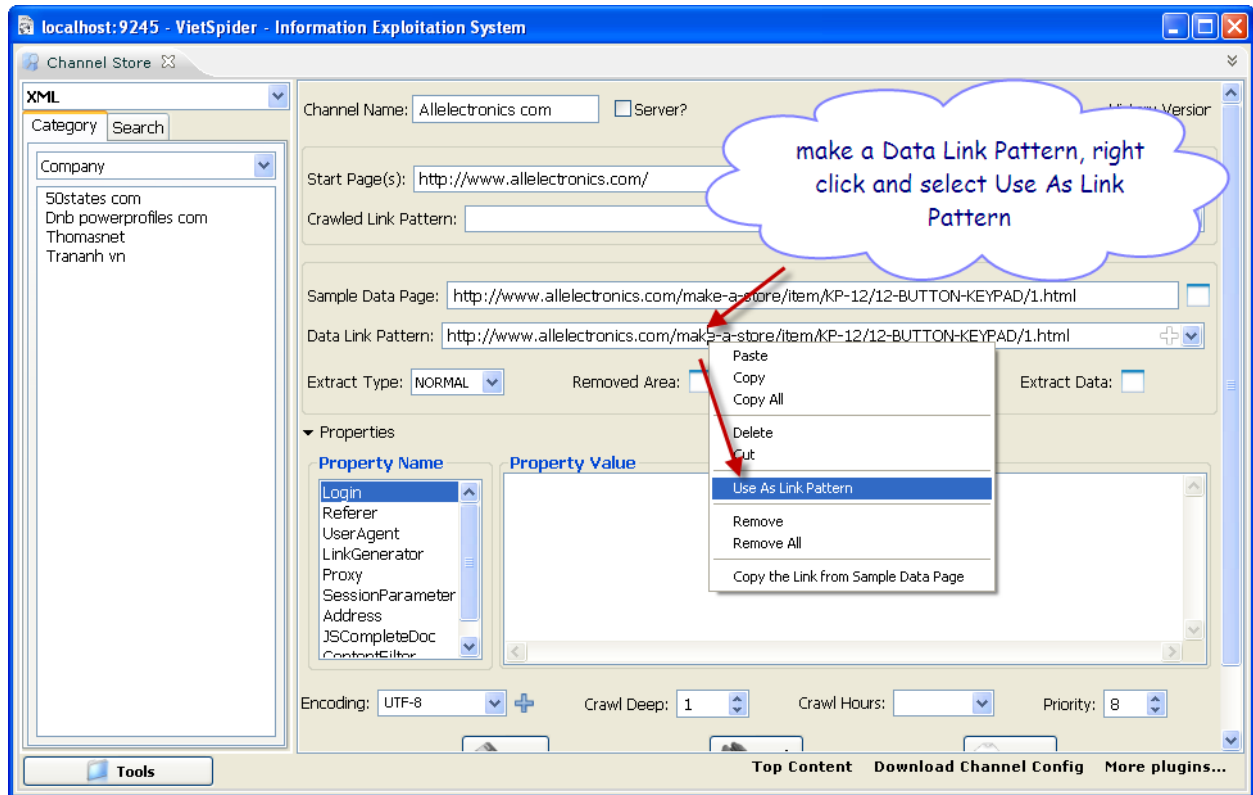


#### 4. Input Sample Data Page

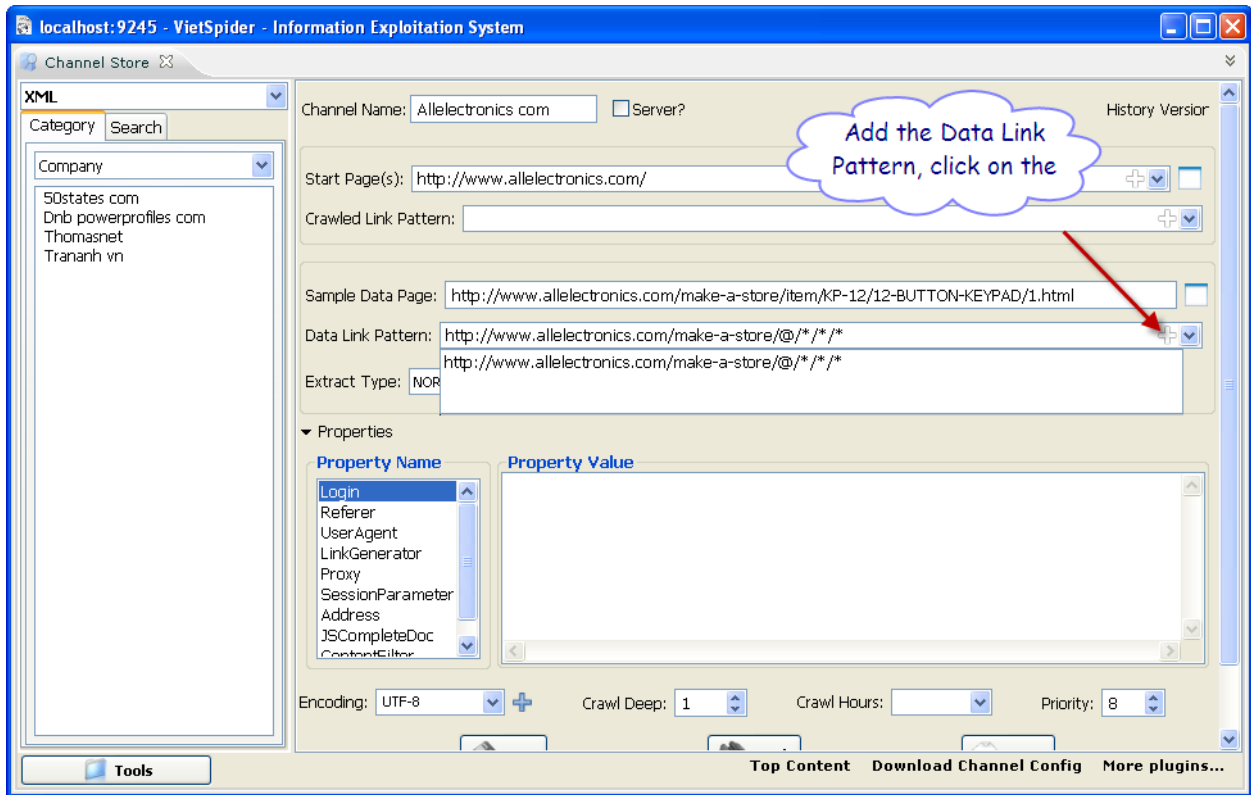
Back to **Web Browser** and copy a data link



## 5. Make Data Link Pattern.

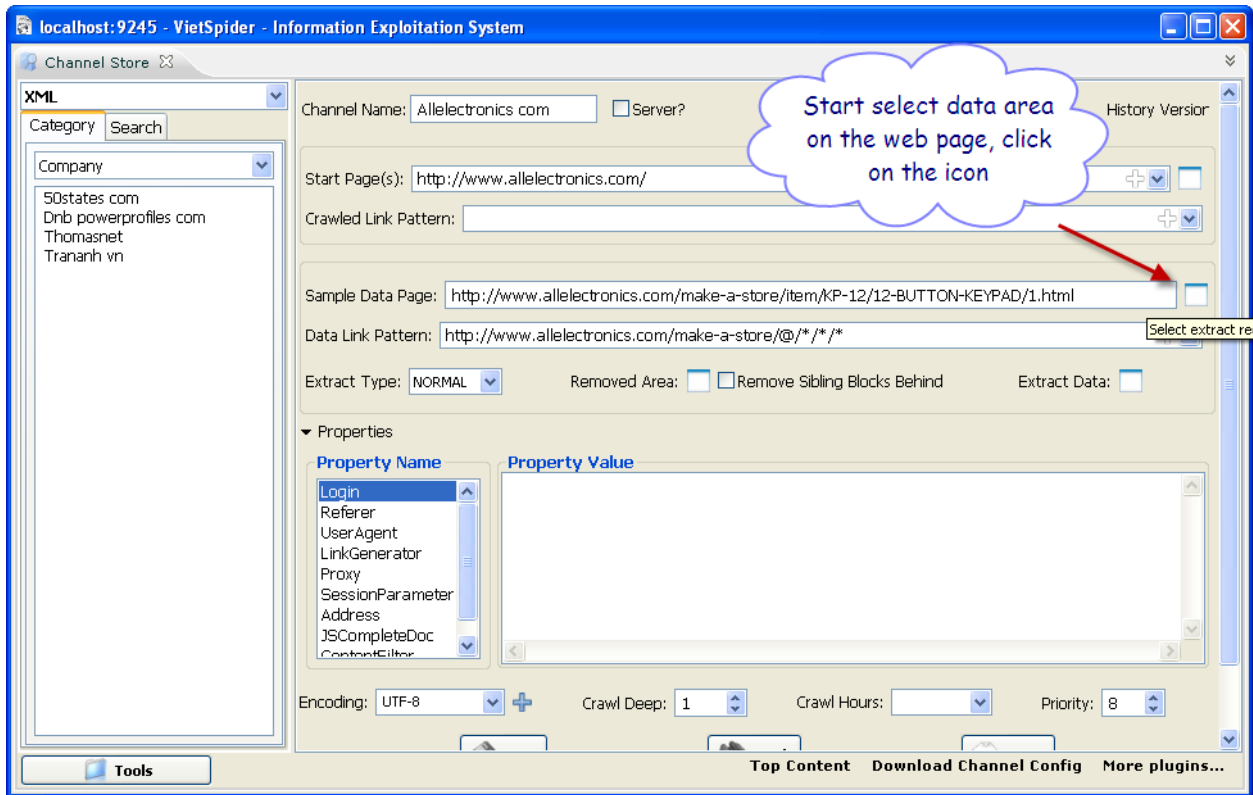


Focus to the appropriate location in the **Data Link Pattern** field, right click and select **Use as Link Pattern**.

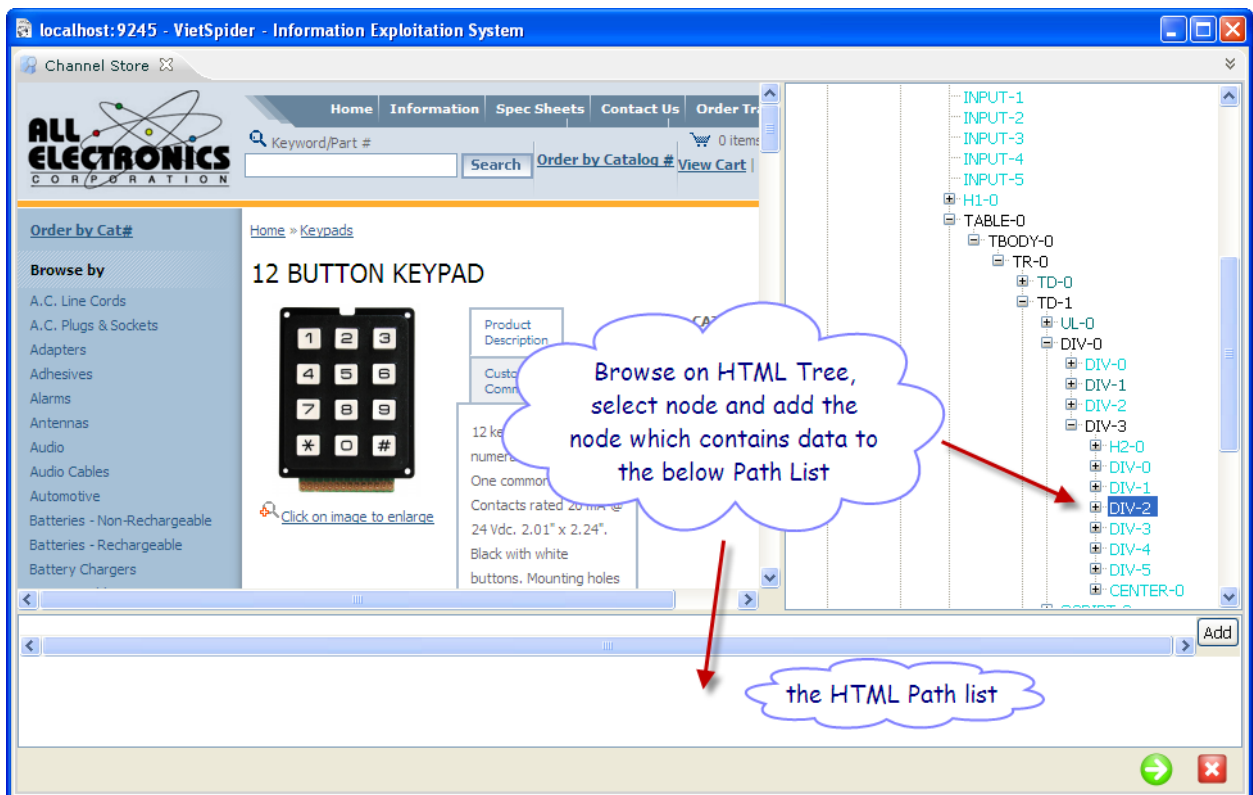


Click on add icon to add url pattern to the list.

**6. Define extract area on the Sample Data Page.**

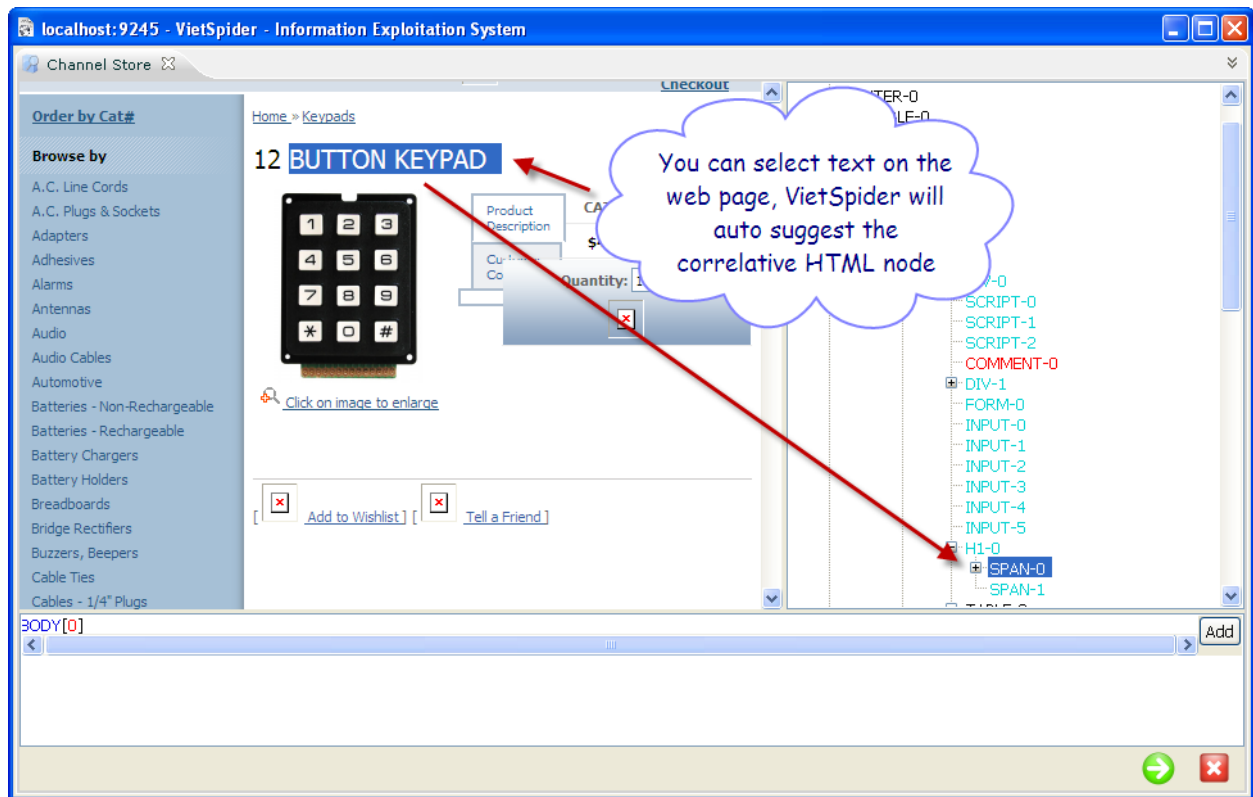


Click the icon is right at the end of the **Sample Data Page** field.

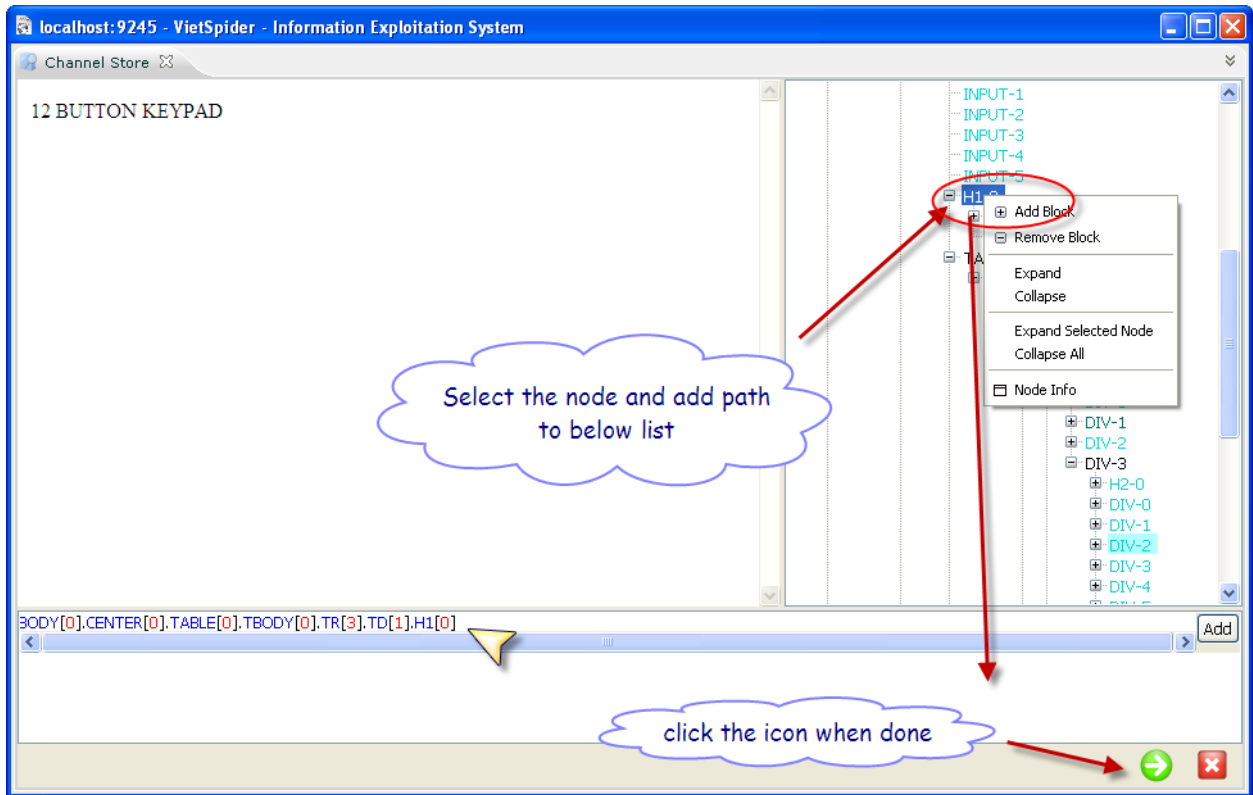




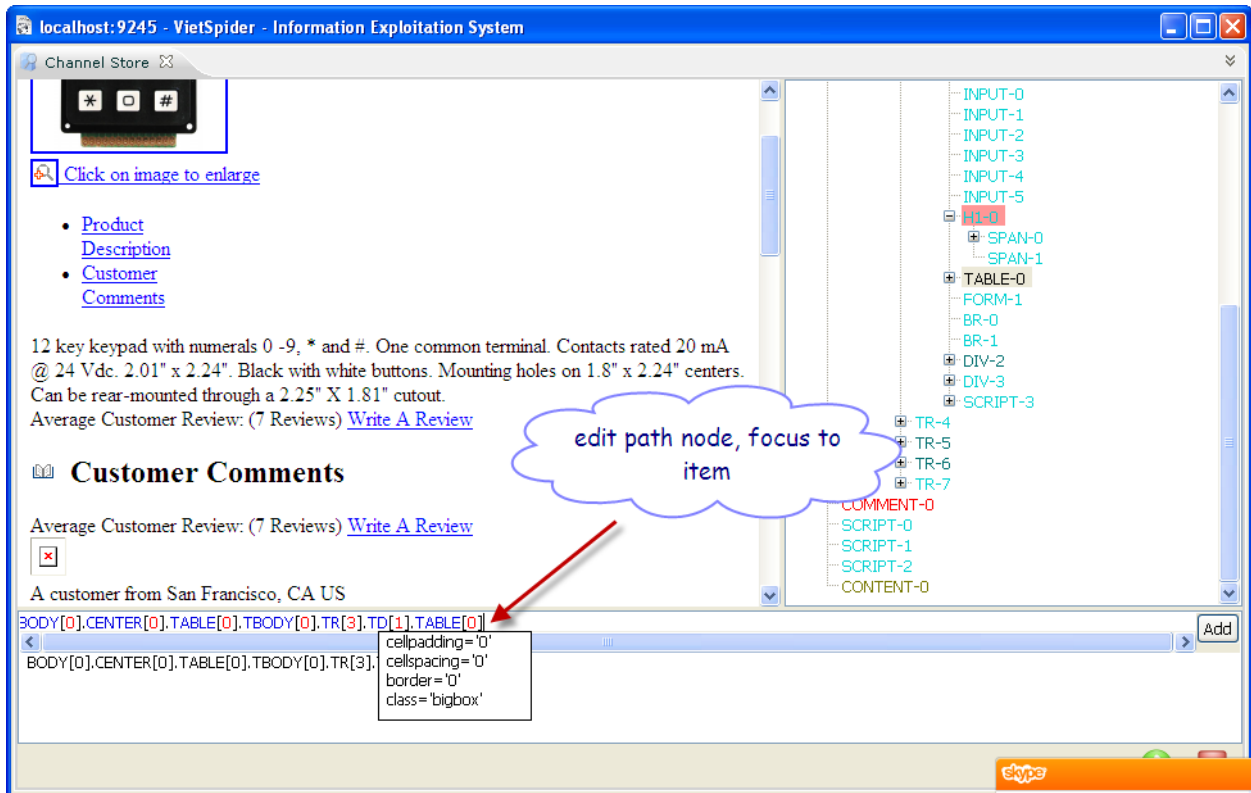
Browse data area of the page by click on the node of HTML Tree



You can select sample text, **VietSpider** will suggest correlative node.

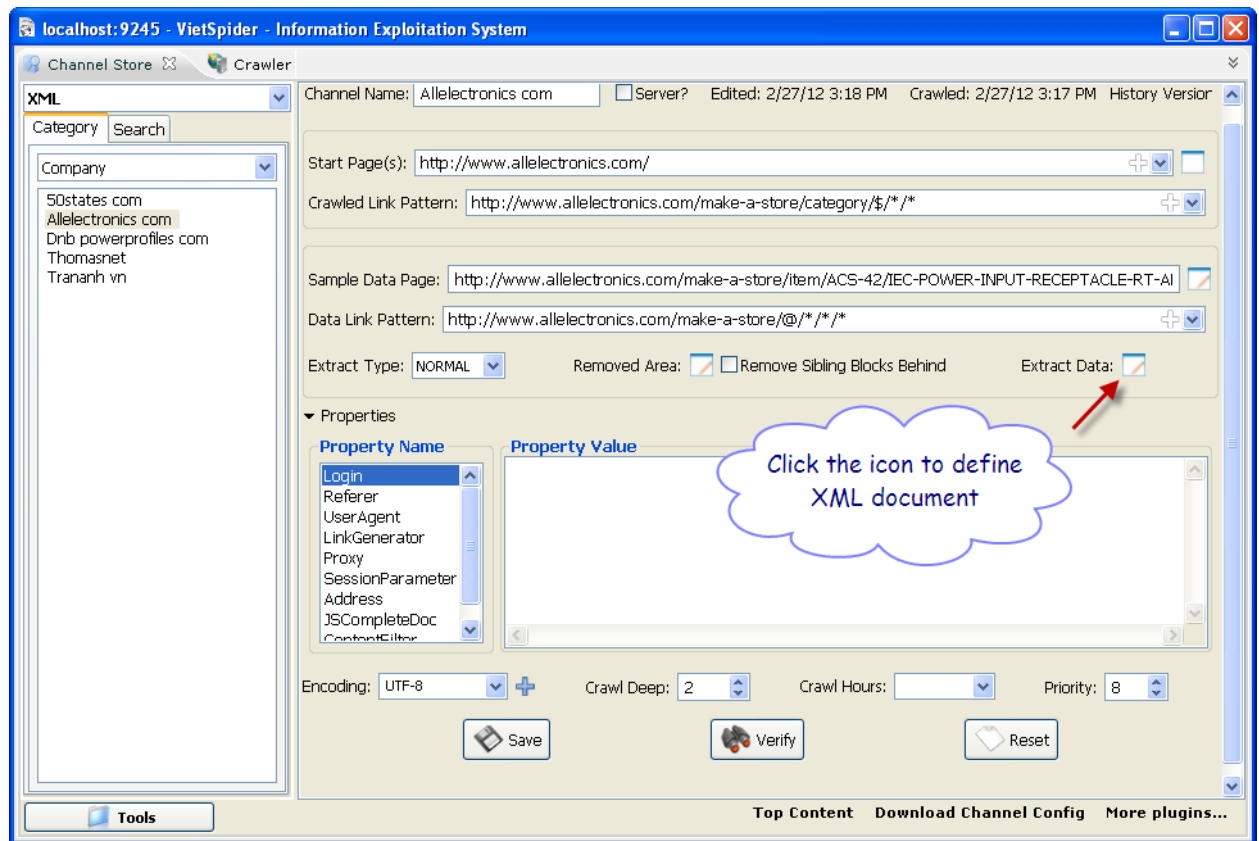


Right click on the node which contains data that you want extract, select **Add Block**.

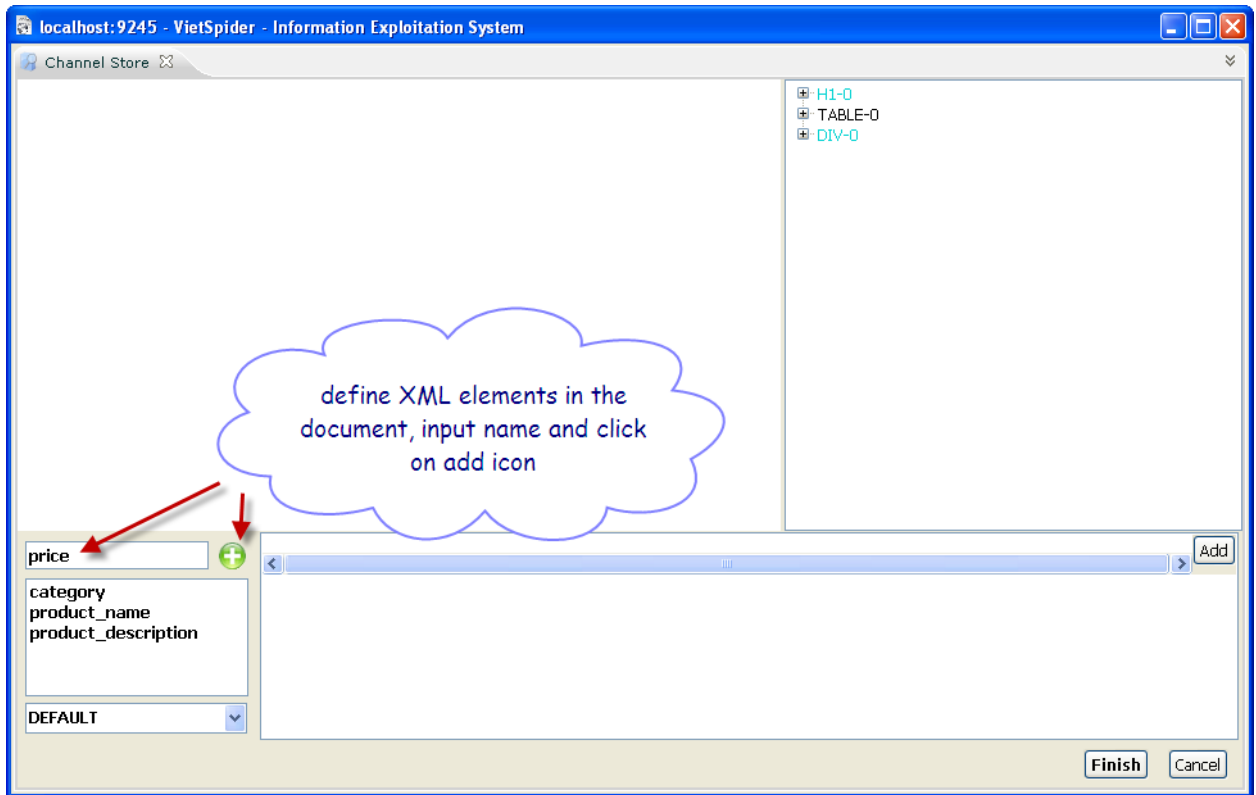


You can edit the HTML node path by focus to a item. Click **Finish** icon when done.

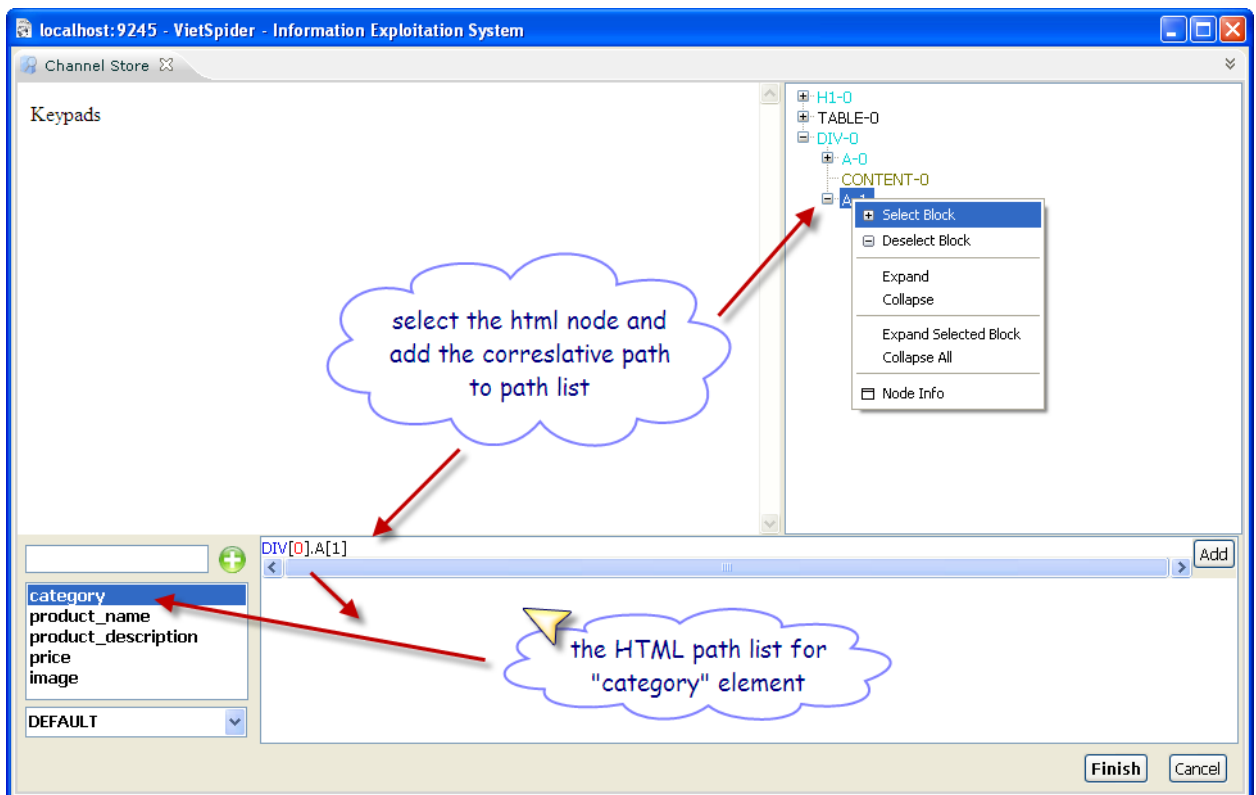
## 7. Define XML schema for data document.



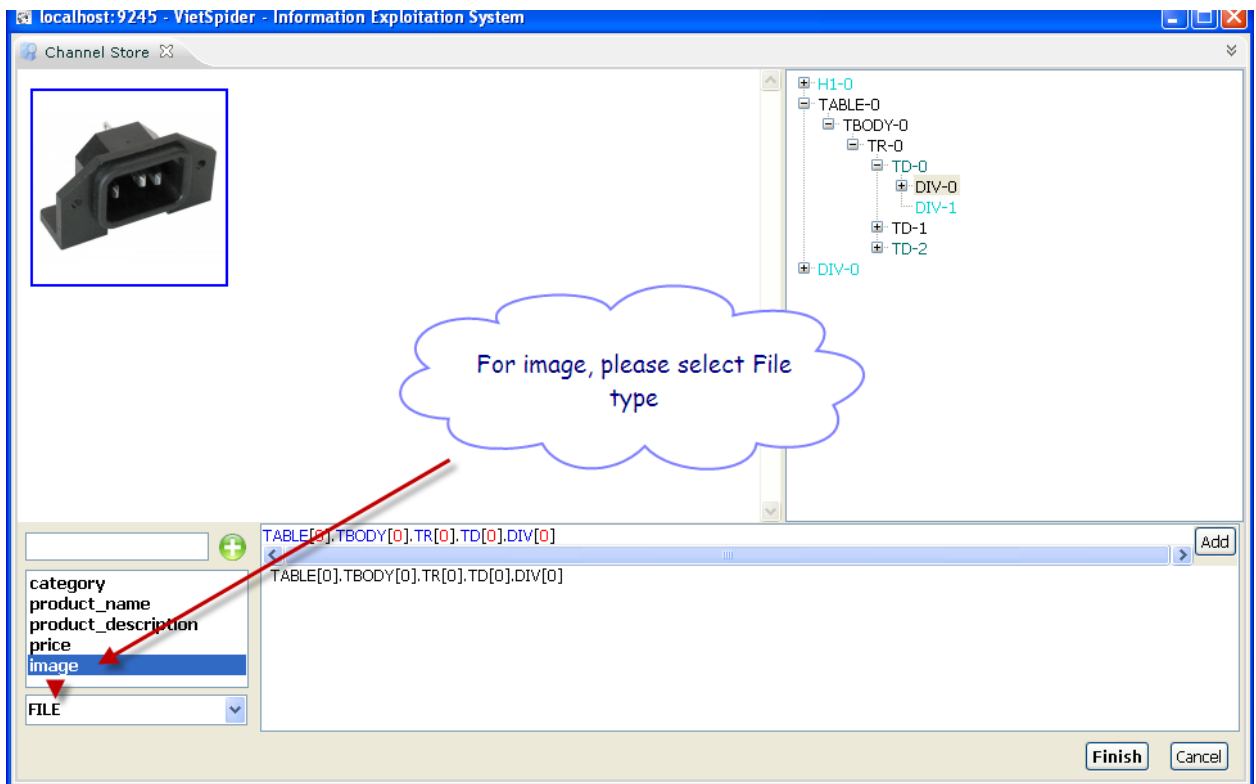
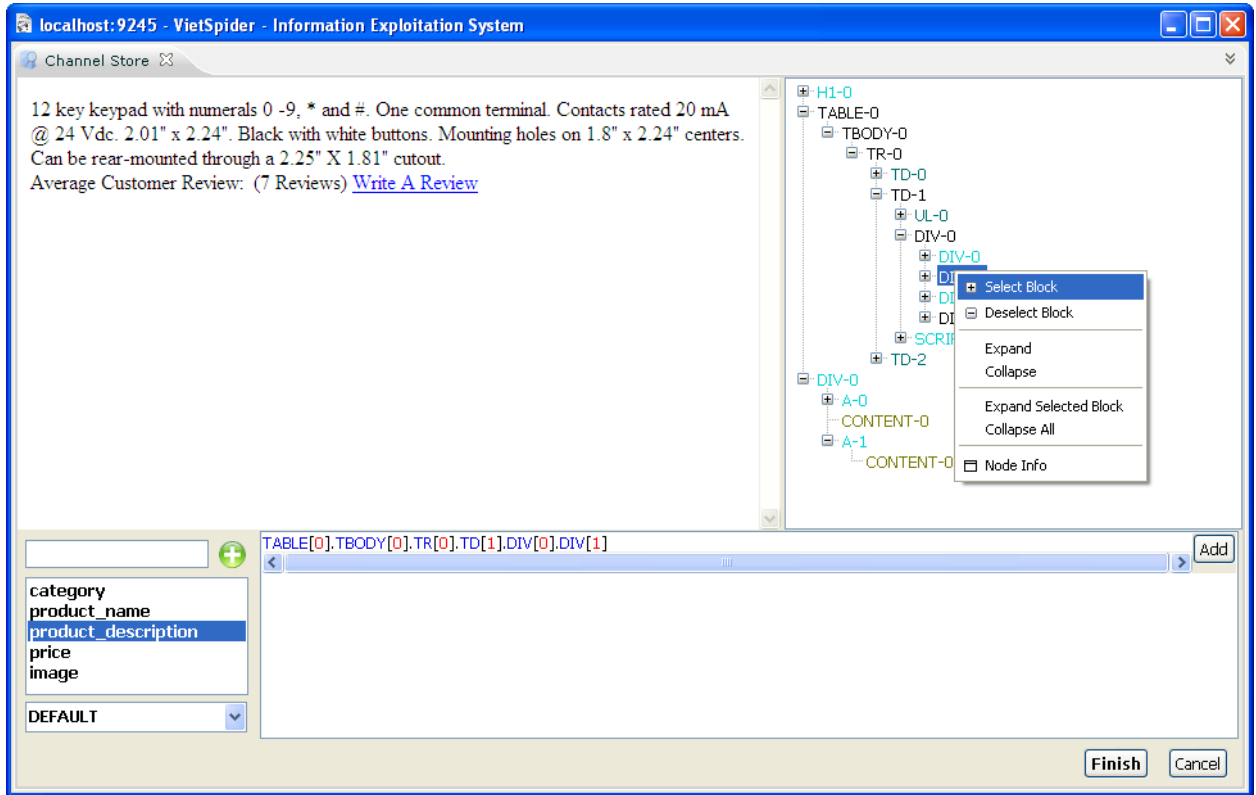
Click on the **Extract Data** icon.



Define XML element of the document by input name and click add icon.

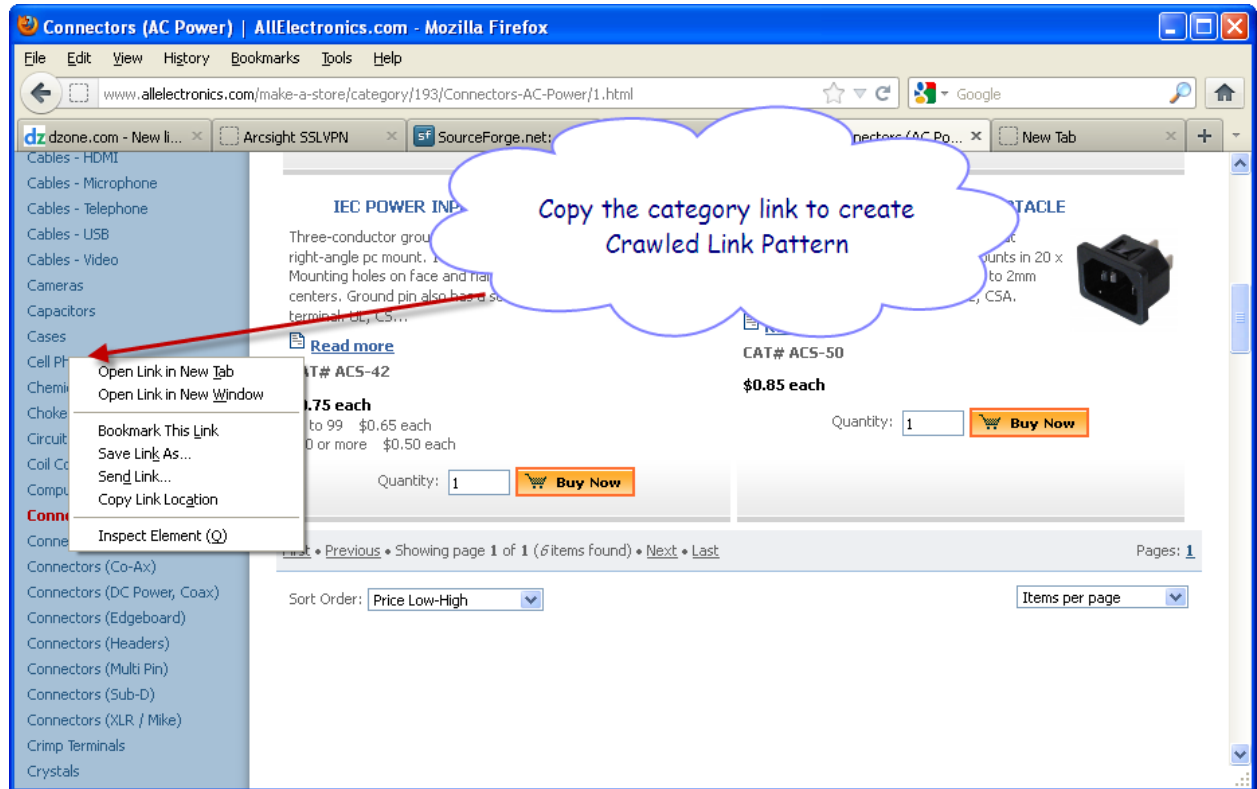


Select element and add data node to the HTML node path list.

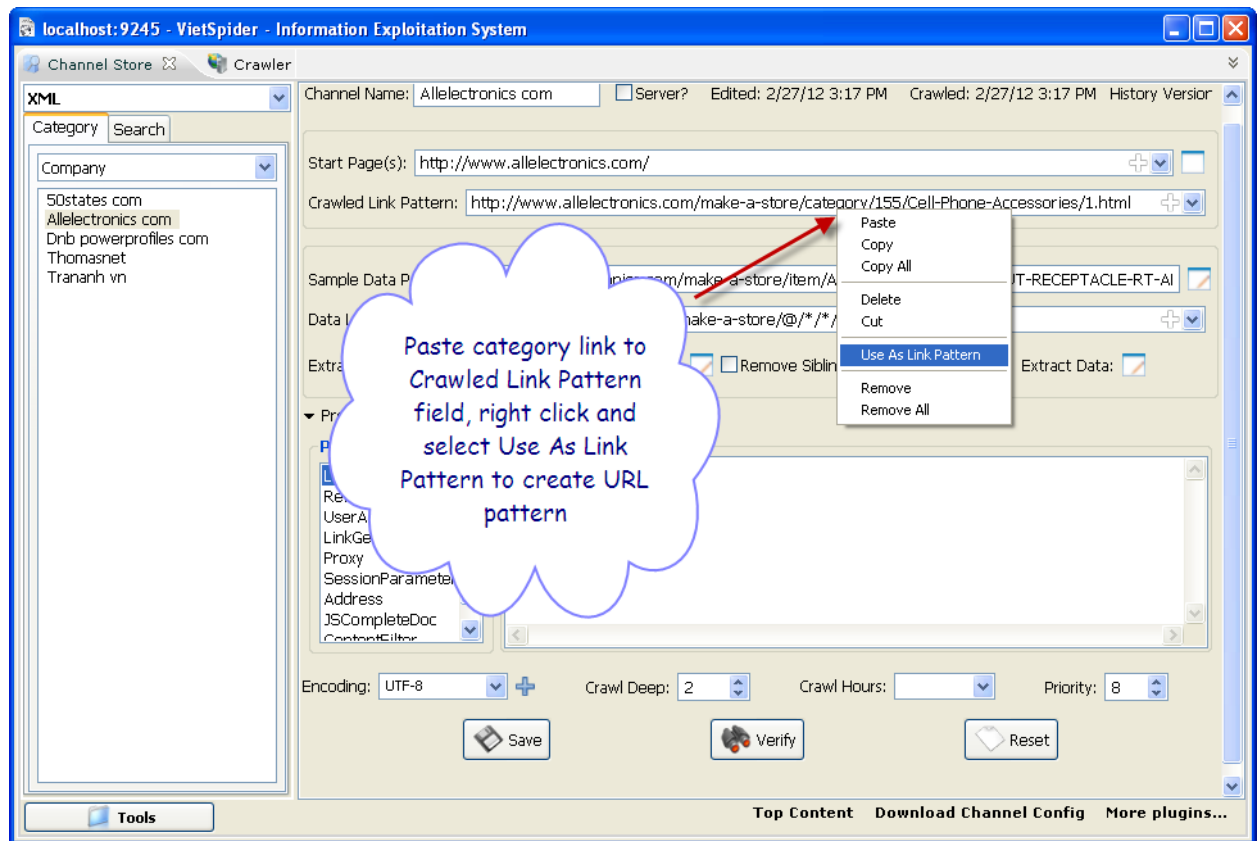


Select data type is **File** if the XML element is image element. Click **Finish** button when done.

## 8. Make Crawled Link Pattern(s).

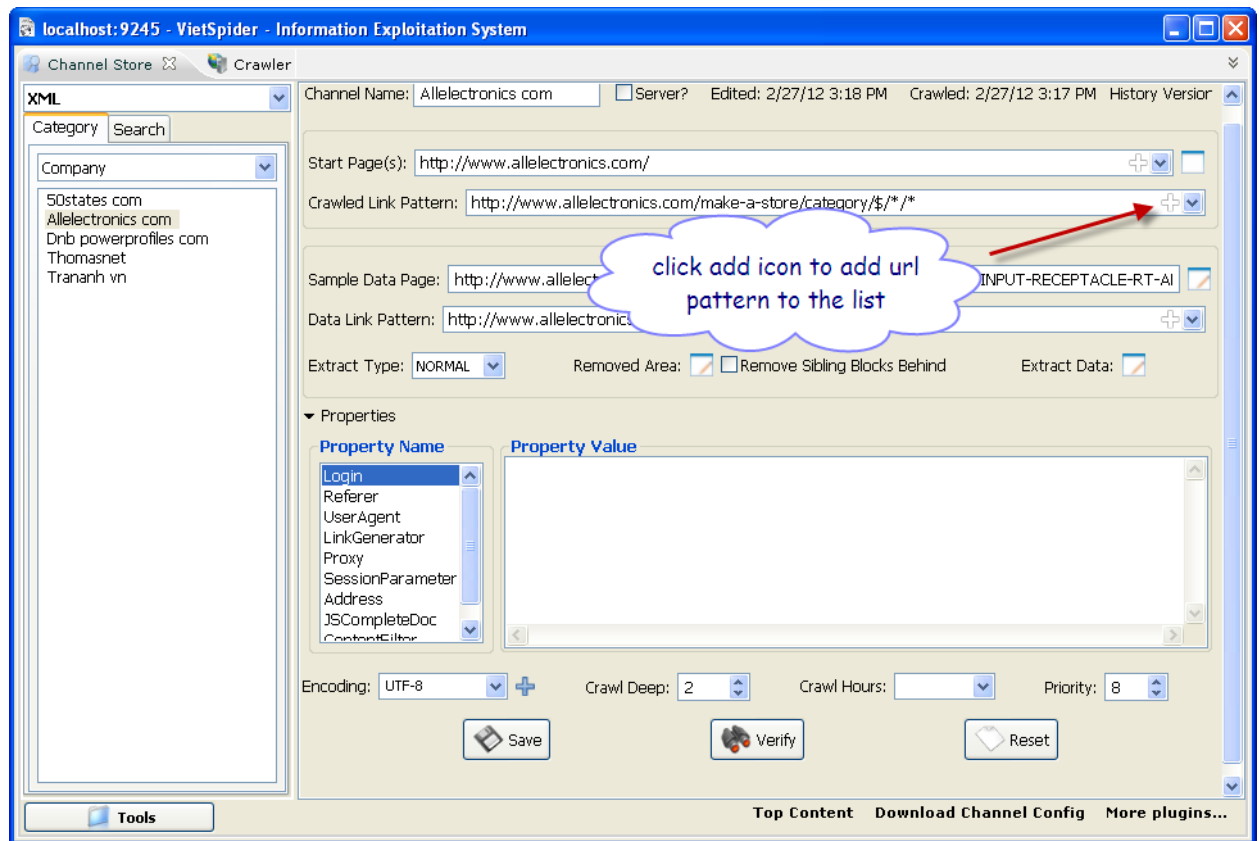


Back to the **Web Browser**, copy a category link.



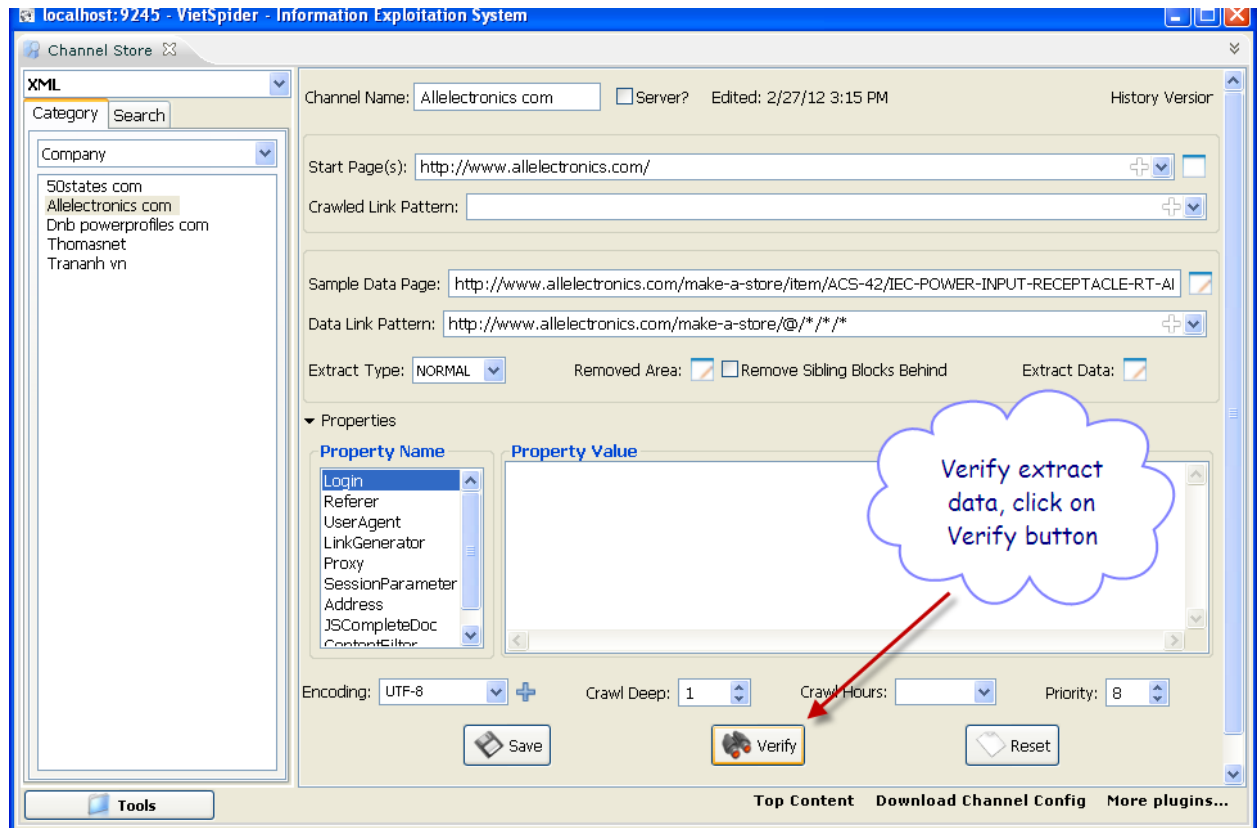
And paste to **Crawled Link Pattern** field to create URL pattern.





Add url pattern to the list by click on add icon.

## 9. Verify the configuration

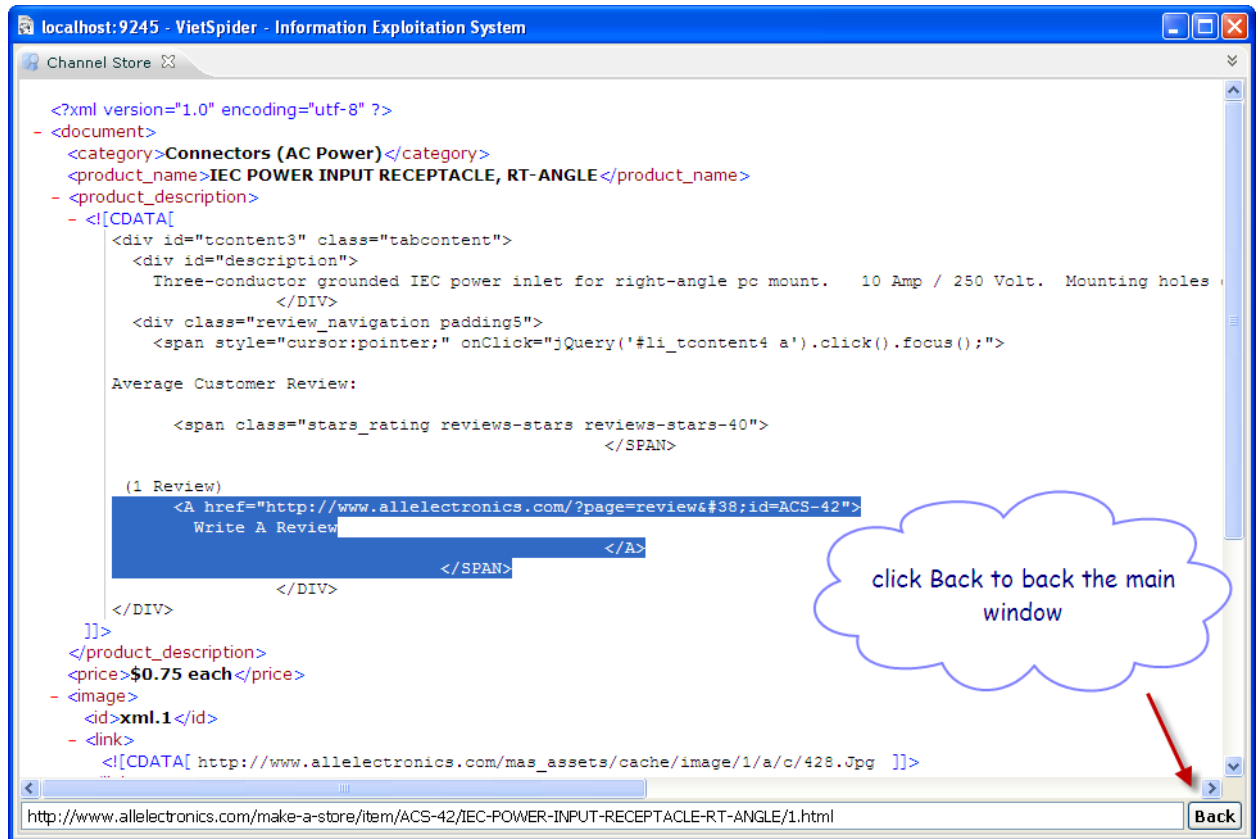


Click on **Verify** button

```
localhost: 9245 - VietSpider - Information Exploitation System
Channel Store
<?xml version="1.0" encoding="utf-8" ?>
- <document>
  <category>Connectors (AC Power)</category>
  <product_name>IEC POWER INPUT RECEPTACLE, RT-ANGLE</product_name>
  - <product_description>
    - <![CDATA[
      <div id="tcontent3" class="tabcontent">
        <div id="description">
          Three-conductor grounded IEC power inlet for right-angle pc mount. 10 Amp / 250 Volt. Mounting holes
        </DIV>
        <div class="review_navigation padding5">
          <span style="cursor:pointer;" onClick="jQuery('#li_tcontent4 a').click().focus();">
            Average Customer Review:
            <span class="stars_rating reviews-stars reviews-stars-40">
              </SPAN>
            (1 Review)
            <A href="http://www.allelectronics.com/?page=review&#38;id=ACS-42">
              Write A Review
            </A>
          </SPAN>
        </DIV>
      </DIV>
    ]]>
  </product_description>
  <price>$0.75 each</price>
  - <image>
    <id>xml.1</id>
    - <link>
      <![CDATA[ http://www.allelectronics.com/mas_assets/cache/image/1/a/c/428.Jpg ]]>
    </link>
  </image>
</document>
http://www.allelectronics.com/make-a-store/item/ACS-42/IEC-POWER-INPUT-RECEPTACLE-RT-ANGLE/1.html
Back
```

click Back to back the main window

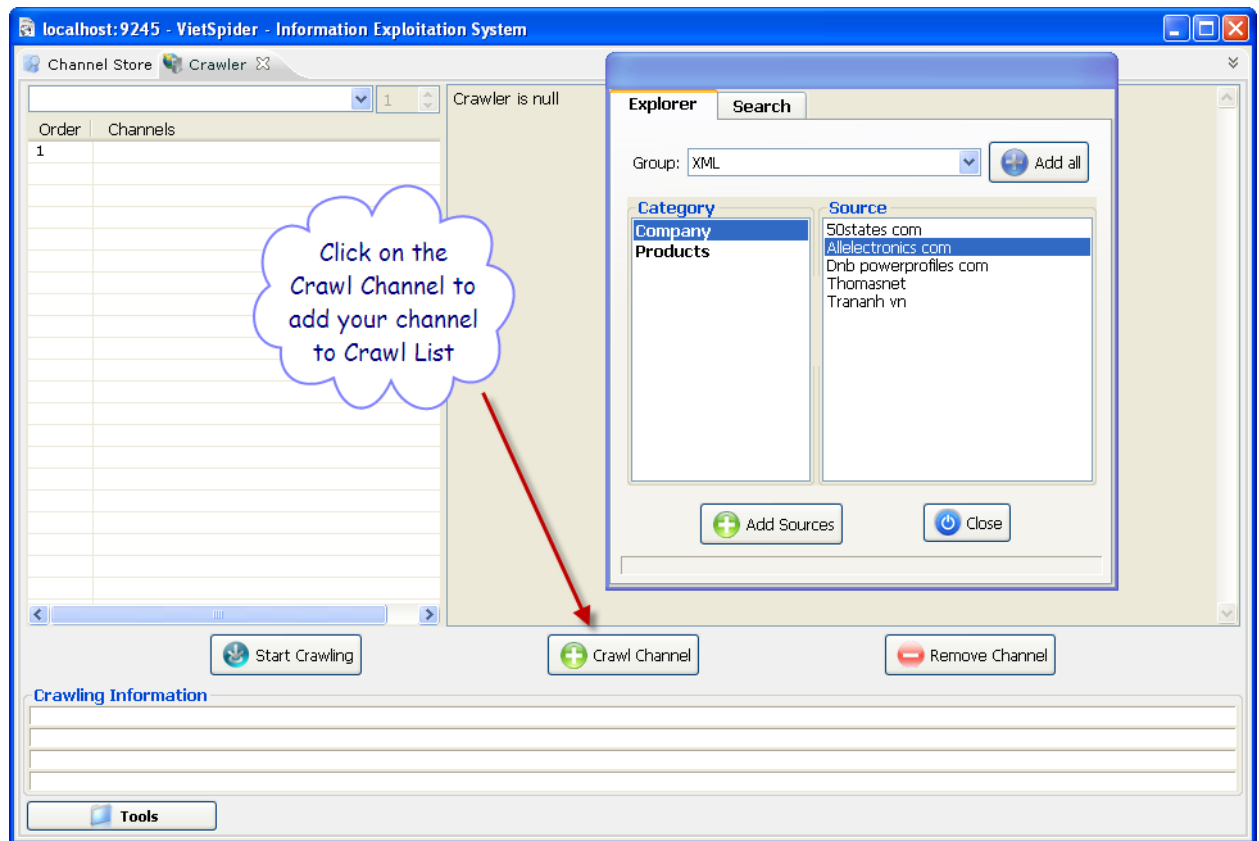
And review sample data after extraction and re-format by XML schema.



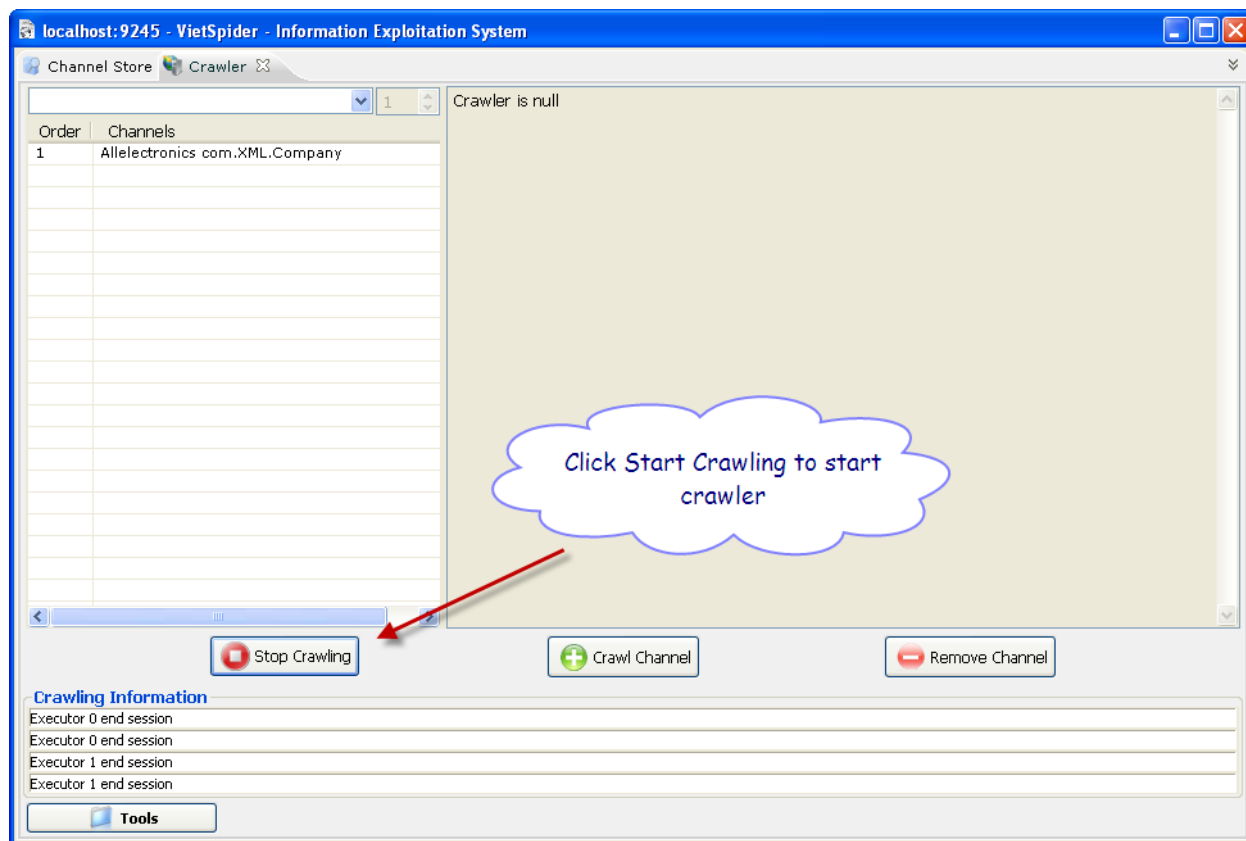
Click **Back** button to back to main window.

## 10. Start Crawling

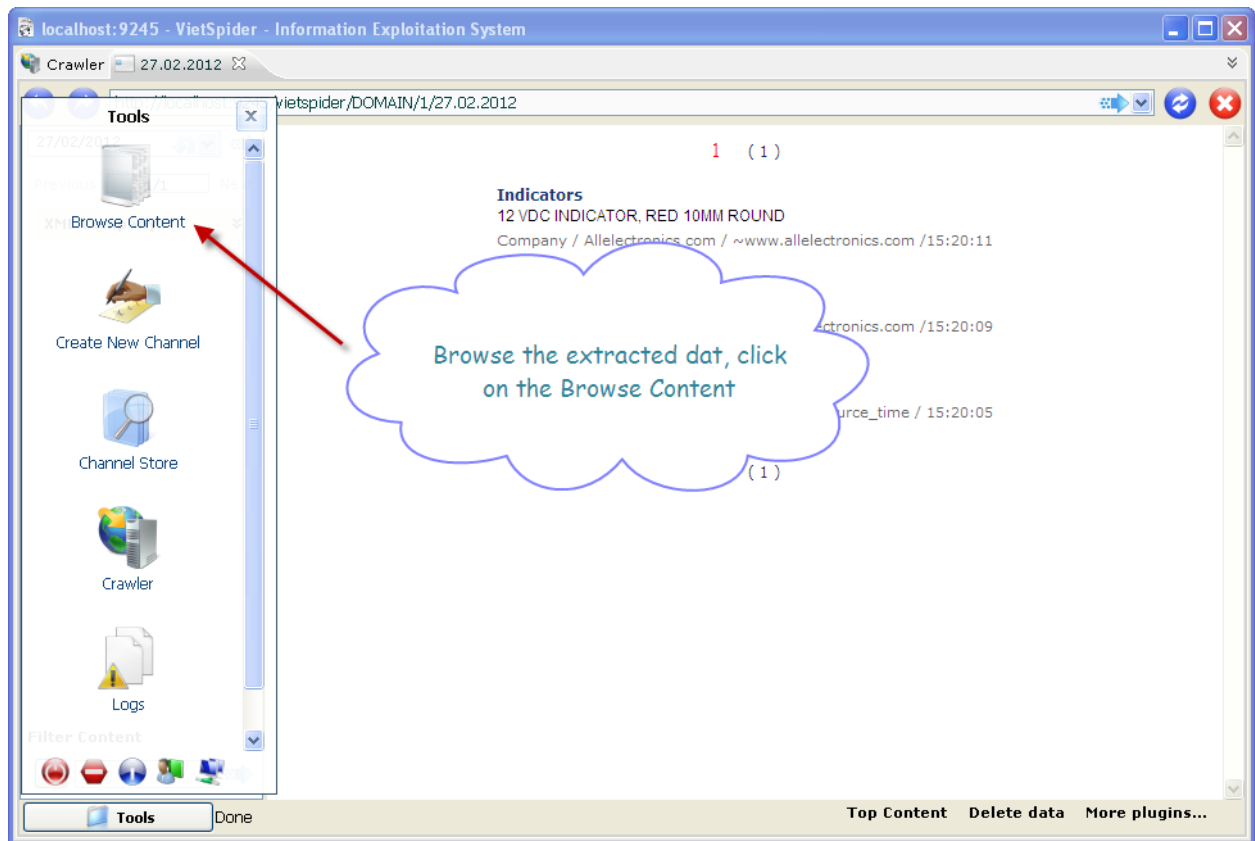
Open **Crawler**, click **Tools** -> select **Crawler**.



Click **Crawl Channel** button to add the configured channel to crawling list.



Click **Start Crawling** button to start crawling the channel.  
**11. Browse downloaded data.**



Click **Tools** -> **Browse Content** to browse crawled data.