### VietSpider Quick Guide – Step by Step – extract data from Android Market

### Part 1: Create new channel

### Launch VietSpiderD:\Documents\Guide\1.png

### Open VietSpider folder and double click on VietSpider.exe

### Open Create Channel Wizard

### D:\Documents\Guide\2.png

### Click on Create New Channel icon

### Browse websiteD:\Documents\Guide\3.png

Input your address into the Address text box.

### Select a data page

### D:\Documents\Guide\4.png

### Select a sample data page. This page will be used for selecting extract data area.

### Go to Next step

### D:\Documents\Guide\5.png

### Select Character Encoding for this website

### D:\Documents\Guide\6.png

### The character encoding selected here will be used to crawl data from this channel. If the web page displays properly with character encoding, please click on Next button to go to next step.

### Define data areas on the data page

### D:\Documents\Guide\7.png

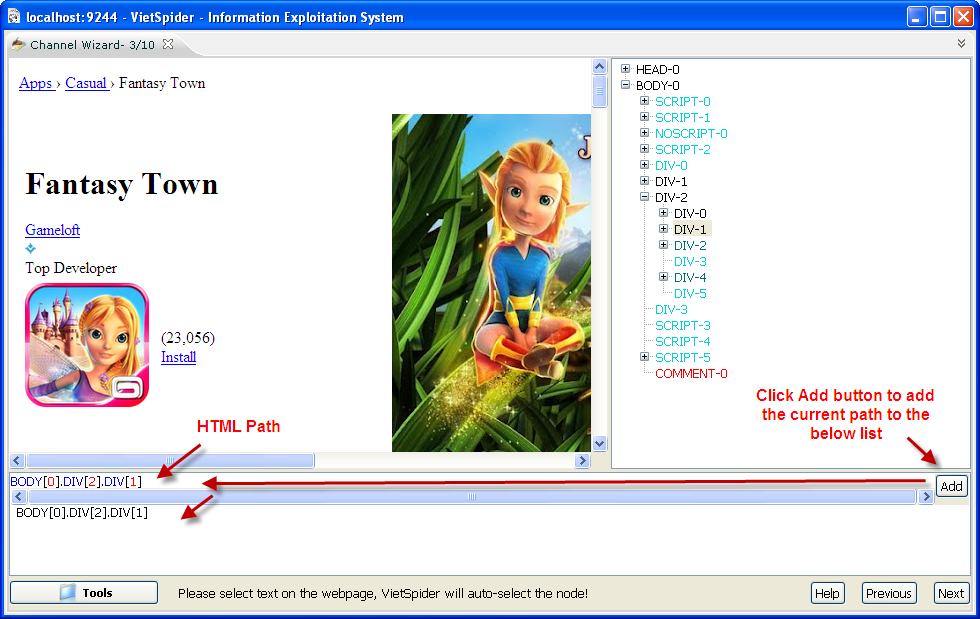
### Click on HTML Node to select data area that you want VietSpider extracts information.

### Auto select html node

### D:\Documents\Guide\8.png

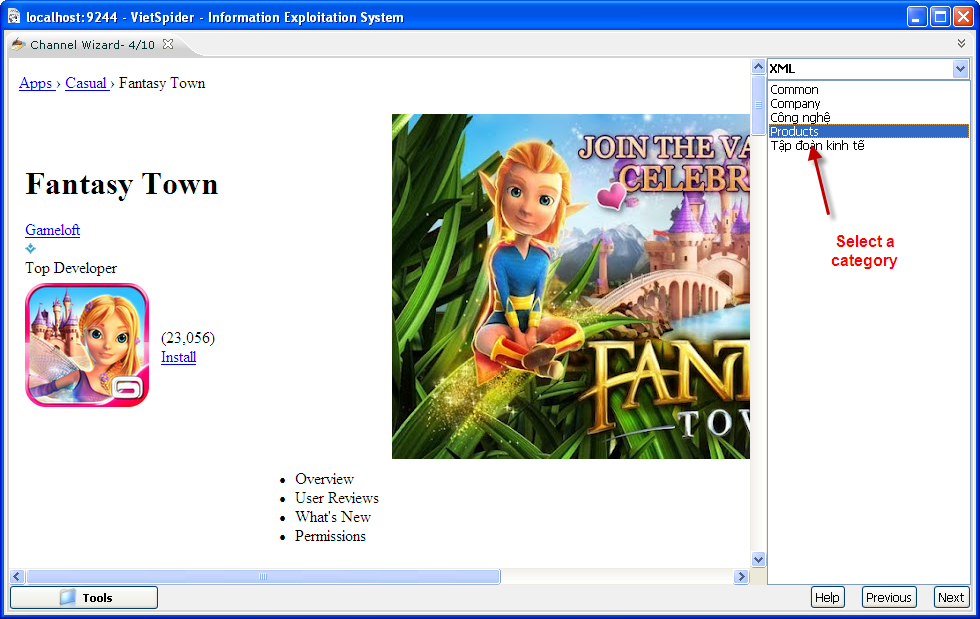
### Please block a passage of text, VietSpider will suggest the node that contains the equivalent data.

### Add extract data node(s)



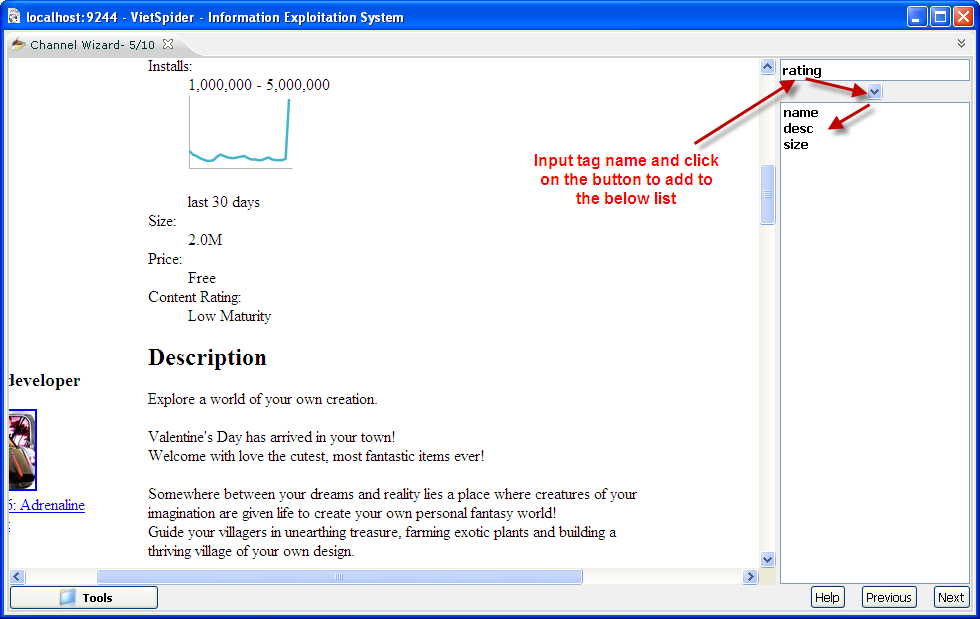
Click Add button to add the current path (from selected HTML Node) to the HTML Path list. The list contains path(s) to data area(s) that you want to extract. Click Next button to go to next step.

### Select category for channel.



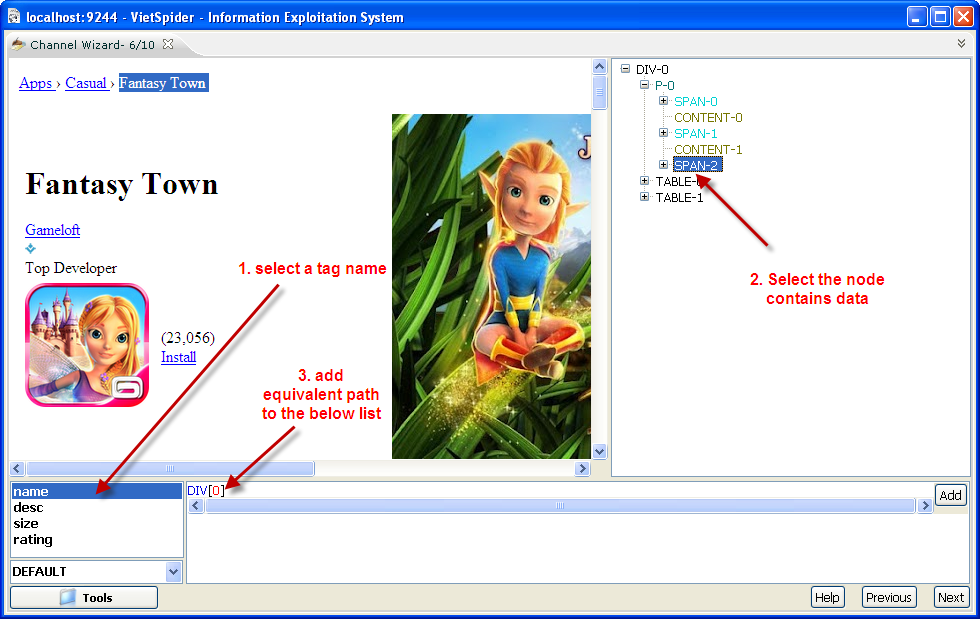
Select a category in the list and click Next button.

### Define XML elements of data document.



You must define XML schame for extracted data document. Please define XML elements by input tag name and add it to the list. Click on Next button when you done.

### Define data area for XML elements

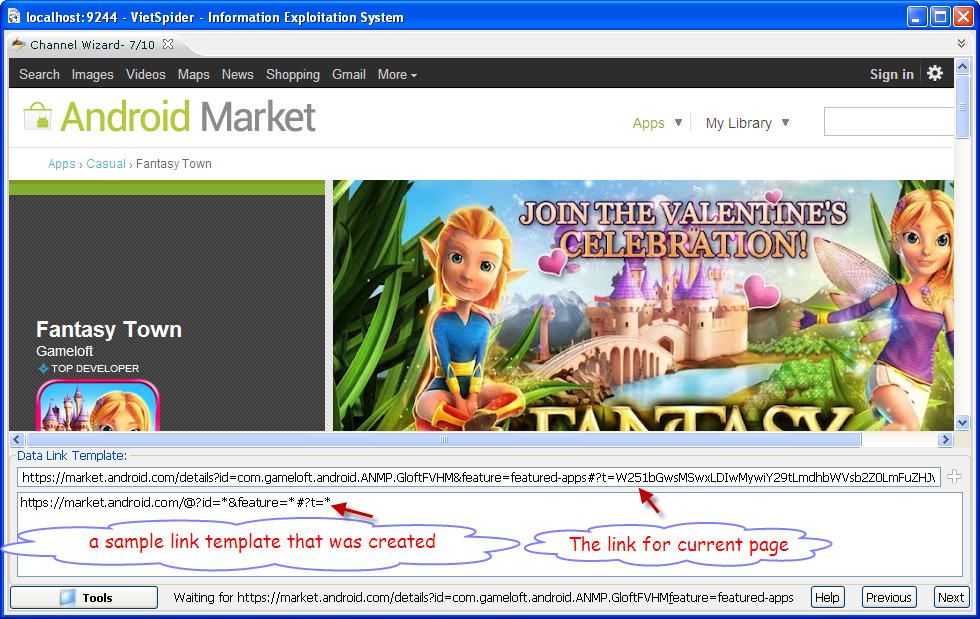


Define data area for XML element, you must select a tag name -> select the node(s) that contains data area then add equivalent path to the list.

There are 4 data type for element:

* Default – VietSpider will auto detect data for equivalent element. That could be CDATA or Text
* Text – the element contains text value.
* CDATA – the element contains HTML tag.
* File - data type is used for image tag or download file link.

### Define data link template.



Data Link Template: this is template that will be used for VietSpider filters data link(s) from link collection when the Crawler crawl data.

Summary of regular-expression constructs

Character – Matches

* \* - any character
* @ - letter character
* $ - digit character

### Add the Start page collection.

### D:\Documents\Guide\14.png

### The Start Page(s): the list of links that VietSpider will use to visit and find data link when crawl data. You can drap and drop the link from the browser.

### Define channel name

### D:\Documents\Guide\15.png

### Input the channel name and click Next button to go to the last step.

### Finish and save the channel

### D:\Documents\Guide\16.png

### If you want to test data extracting, you can copy another data link from the website and paste it to the Sample Data Page text box and click on Test button. Click Done button to save the channel to the Channel store.

### Part 2: Crawl data by the configured channel.

### Open Crawler

### D:\Documents\Guide\18.png

### From VietSpider user interface, click Tools button -> click on Crawler icon to open Crawler.

### Add channel(s) to crawling list.D:\Documents\Guide\19.png

### Click on Crawl Channel button and add the channel(s) that you want to download and extract data. Click Start Crawling button for starting crawling data from website(s).

### Part 3: Repair the channel.

### Open the Channel Store

### D:\Documents\Guide\20.1.png

### Click Tools button then click on Channel Store icon.

### View website in the browser.

### D:\Documents\Guide\20.2.png

### Select the channel that you want repair it. From Start Page(s) text box, right click then select View In Browser.

### Copy a sample data link

### D:\Documents\Guide\21.png

### Copy a sample data link for testing defined extract data area.

### Test and repair Data Link Pattern

### D:\Documents\Guide\22.png

### Paste the link to Sample Data Page text box, you can see the error message from VietSpider. Please add the below patterns to Data Link Pattern to fix it.

### [https://market.android.com/details?id=\*&feature=\*](https://market.android.com/details?id=*&feature=*)

### Add New Data Link Pattern

### D:\Documents\Guide\20.3.png

### Copy and Paste new data link pattern to the text box and click Add icon at the end text box.

### Click Test button

### Test extracted data.

### D:\Documents\Guide\20.4.png

### When you click on Test button, you can see extracted data document is empty.

### Fix extract data for XML element.

### D:\Documents\Guide\20.5.png

### Please review extract data and extract region, click icon.

### Update incorrect path

### D:\Documents\Guide\23.png

### D:\Documents\Guide\24.png

### Review the path of XML elements and the path of extract region, update when you see the path is incorrect. Save the channel and re-crawl data when you done.

### Part 4: Browse the crawled Data

### Click Tools -> Browse Content

### D:\Documents\Guide\25.png

### D:\Documents\Guide\26.png