

Documentation

for XPather 1.4.* (July 1st 2009)

How to use the XPather?

Once you [installed it](#), learn about it in:

1. [A quick kick-off tutorial](#)
2. [A full user guide](#)
3. [A cheatsheet that is also built in the XPather](#)

XPather Kick-off Tutorial

There are two main cases how to start to use the XPahter:

- **Case A:** *I have [DOM Inspector](#) in my Firefox*
 1. Open the DOM Inspector from your browser (Ctrl+Shift+i)
 2. Move around in the DOM tree (arrows, clicking, find-by-click). Your actual tree navigation is displayed in the XPather toolbar.
 3. Want to modify the xpath/test your own xpath?
Edit the XPath (Alt+p). Red means error. Click Enter.
 4. By default an XPather Browser pops up and stays on top of all mozilla wins.
Inspect the xpath matches, their locations, contents, position in inspector/browser.
 5. Do you want extended matching, regexps, customize xpath generation, generate relative xpaths, extract information, sync with browser ... [read the Guide](#).
- **Case B:** *I don't have [DOM Inspector](#) (you should get it ;), as of FF3b4 you need to download it as a [separate extension](#)), or I want to go directly from browser:*
 1. Invoke context menu (right click) on the object in the browsed document you are interested in.
 2. Choose "Show in XPather"
 3. Go to [point 4 in scenario A](#)

XPather Guide

Operation Modes

XPather operates in several modes (around 2 and half ;))

If you work in DOM Inspector (Case A):

- it generates XPath as you browse the DOM tree of the inspected document (the generator is customizable)
- you can also use find-by-click option of the DOM Inspector

- allows you manually enter or modify the XPath (with syntax check), evaluate it and see the matching nodes
 1. either as selections in the DOM tree
 2. or in a separate tool, [XPather Browser](#), in a new window

If you are in browser (Case B):

- you can invoke the [XPather Browser](#) from the document context menu. The document element you click on gets displayed

Once you invoke the XPather Browser, you can further work with the XPath, inspect the results, their content, and do simple extraction. It keeps coupled with the original browser/inspector/document. In a way:

- it tries to be compact and to stays above the the other mozilla windows so you can use along with the document you inspect/browse.
- there are actions to see what you do or to control the underlying application
- you can anytime update it by new evaluation in DOM Inspector, or clicking from the browser

XPath Toolbar

This toolbar stands for the main navigation toolbar of the application. It appears both in DOM Inspector and XPather Browser. It:

- displays generated XPath (while navigating DOM tree, using find-by-click)
- allows to manually modify or enter an own xpath (edit it, Alt+p)
- checks xpath syntax while typing (turns red if error)
- evaluates the xpath (Eval button, Enter)
- allows customize [XPather settings](#) (the leftmost XPath menu button)
- displays help and cheatsheet (rightmost ? button)

You can customize: the generation of XPath; usage of additional tools; how the evaluation results are presented. Learn about it in the [next section](#).

XPather Settings Menu

There are several settings and tools you can customize in the XPather. These can be altered in the menu that pops up by clicking on the XPath button (on the XPather toolbar. The options are the following:

XPath generation options (they can be used in any combination)

- Show 'id' - use node id in xpath conditions
- Show 'class' - use class attribute of the node in conditions
- Show Namespaces - show node namespace prefix when generating the XPath. If the document defines default namespace, the XPather generates **default** prefix for all the non-qualified nodes. This is a necessary workaround to query for default namespaces using XPath 1.0 spec.

- To Lowercase - lowercase all node names

XPath evaluation settings

- Cross-frame eval - Cross-frame XPath evaluation. If this setting is active, you can evaluate your XPaths against all frames/iframes/etc in the document. (in this case relative/parent XPaths cannot be supported)

Additional tools:

- Parent Toolbar - toggle a toolbar and functionality that allows to generate relative XPaths. See the [Parent Toolbar](#)
- Regexp View - toggle a toolbar and functionality that allows content text based filtration and extraction (available only in XPather Browser). See the [RegExp Toolbar](#)

Evaluation options:

- Select Results- Display the XPath evaluation results (matching nodes) as selection in the DOM tree.
- Browse Results - Display the XPath evaluation results (matching nodes) in the XPath Browser.

The settings for XPather toolbar in DOM Inspector and XPather Browser are independent. Your options are persisted.

Note: Settings changes are applied immediately. So all the automatically generated content (XPaths) is regenerated. However, any manually modified XPath cannot be regenerated (for obvious reasons)!

XPather Browser (evaluation results)

XPather Browser is a tool that opens up in a separate window, invoked by either DOM Inspector or Browser to work with XPaths and matching nodes in details. It takes an xpath and entire context from the underlying DOM Inspector/Browser and evaluates it. If additional tools (toolbars) are activated, they influence the evaluation results as well. You can inspect, browse, manipulate the results here, and further modify the existing, or test any new XPath.

XPather Browser contains the same set of toolbars as in the DOM Inspector (RegExp Toolbar is available here additionally), so the shortcuts and functionality is analogical.

The tool window is intentionally quite compact and always above other mozilla windows so they can be used in cooperation. The results of the match can be reflected to the DOM Inspector (selection) or Browser (marking), as well as they can always command the already opened tool to work with another XPath or document.

Results Table

Each time the expression is evaluated the results, i.e. matched nodes are listed in a table below the toolbars. For each node a result number, text content preview (modified), and its full XPath can be provided. When you select some node, its content is displayed in the content viewer tabs at the

bottom. You can also select multiple results. The extracted content from all of them is 'concatenated' accordingly in content viewers.

You can use context menu to fire an action upon the selected node (for many actions just single selection makes sense). The actions are as follows:

- Copy XPath- copy the full XPath into clipboard
- Blink the Node- blink (highlight) the node in original browser (available only if DOMInspector is installed at the moment)
- Set to Toolbar- set xpath of the node to the XPath toolbar
- Set Node to Parent- set the selected node to become parent for relative XPath generation (Parent Toolbar has to be activated in XPath menu)
- Select in DOM Tree- selects and focuses the node(s) in the DOM tree (if DOM Inspector was the opener)

Content Viewers

Content viewers consist of several tab panes, each displaying a different aspect of the node content. The following views are available at the moment:

- Text

This tab displays the textual content of the node, including all its descendants, i.e. `textContent`. Multiple contents are joined by comma. If content substitution/textual extraction is activated, this is the only tab which gets affected. It displays the transformed content.

- Inner HTML

The inner HTML text of the selected nodes is displayed here. Multiple content is separated by XML comments.

- Web Clipping

This tries to display the rendered content as it shows up in the browser. Note, the content may not appear the same formatted or shaped as in the original document. Moreover, only HTML content can be displayed here. With JS, redirects, subframes ...disabled. Multiple content is displayed as a vertical sequence.

- XPaths

This tab shows the XPath of the selected node(s). Multiple content is displayed each on new line. If the Parent Toolbar is activated, and parent node is selected, a full and also relative path is generated for the each node.

- Info

Info tab contains general info about actual document, frame being under inspection. Also warnings and other messages may appear here.

Tip: if you don't want to see all the evaluation results, but just their count matters, use `count ()` function. For example: `count (//li)` results in an alert window showing the number of list items in the document.

Parent Toolbar

Often it is useful work with relative xpaths to some node(e.g. `./td[3]`). If you want to write such xpaths and you want XPather to generate them, activate the XPath>>Parent Toolbar, and set the desired 'parent' node. To set a node as parent: If you are in the DOM Inspector, set the parent node by Set Node as Parent option in the context menu of the selected node in the DOM tree. In XPather Browser, action the similar command from context menu upon the selected result in the table.

To clear the parent node, use the Clear button on the Parent Toolbar.

Note: Both the Parent Toolbar has to be activated and the parent node set so the relative paths are generated/can be used.

RegExp Toolbar

This is an additional toolbar that can provide two services in context of XPath evaluation:

- Postprocessing content text based filter.
- Textual extraction and substitution.

The tool is operable only when it is activated (XPath menu>>RegExp Toolbar), and RegExp, Subst fields are correctly filled respectively.

Note: This toolbar is not available in DOM Inspector toolbar set.

RegExp Filter

If a regular expression is given, the textContent of all the XPath matching nodes is tested against it. Thus you can select only those nodes from the structure that contains an information of your interest. In case the regexp filter is used the matching nodes count has a format (count: X from Y). This means that Y nodes came out of XPath and X of them passed the regexp.

XPather requires regexps in JavaScript literal format (so that processing flags can be used easily), that is:

- `/.*/` matches everything
- `/ab+c/i` matches `abc`, `abbc`, `ABbc`

Similarly to the XPath toolbar also this toolbar checks the syntax validity as you type. An invalid expression turns red. Empty expression means pass-by which is effectively the same as match all.

Tip: If you click on the toolbar heading (i.e. **RegExp**) a default, match-all-and-remember pattern is set into the regexp textbox. That is `/(.*)/`.

Extraction/Substitutions

Additionally to regexp filter, the expression can be used to extract/substitute relevant text data.

Enter the output pattern in the Subst textbox in form of JavaScript string replace, i.e. `$n` is a variable

that denotes n-th substring (group) defined by the regexp; all other characters are kept. For example a regexp `/.*\$(\d+).*/` and subst: `$1` would result in extraction of all prices within given DOM structure (see screenshot).

This transformation applies only on the text content of the nodes, thus it is visible only in the Text content viewer.

Tip: Analogically to RegExp default button, you can click also to **Subst**. If you do so, a default substitution is generated and set to the textbox. Default has the form `$1 $2 . . . $n` according to number of groups defined in the regexp. If no groups are defined, the default value is not set.