**FilterPlugin**

Substitute and extract information from content by using regular expressions

**Description**

This plugin allows to substitute and extract information from content by using regular expressions. There are three different types of new functions:

1. FORMATLIST: manipulate a list of items; it is highly configurable to define what constitutes a list and how to extract items from it
2. SUBST, STARTSUBST/STOPSUBST: substitute a pattern in a chunk of text
3. EXTRACT, STARTEXTRACT/STOPEXTRACT: extract a pattern from a text

While the START-STOP versions of SUBST and EXTRACT work on inline text, the normal versions process a source topic before including it into the current one.

**Syntax Rules**

**SUBST**

**Syntax:** `%SUBST{topic="..." ...}%`

Insert a topic by processing its content.

- `topic="..."`: name of the topic text to be processed
- `text="..."`: text to be processed (has got higher precedence than 'topic')
- `pattern="..."`: pattern to be extracted or substituted
- `format="..."`: format expression or pattern substitute
- `header="..."`: header string prepended to output
- `footer="..."`: footer string appended to output
- `limit="<n>"`: maximum number of occurrences to extract or substitute counted from the start of the text (defaults to 100000 aka all hits)
- `skip="<n>"`: skip the first n occurrences
- `exclude="..."`: skip occurrences that match this regular expression
- `sort="on,off,alpha,num"`: order of the formatted items (default "off")
- `expand="on,off"`: toggle expansion of TWiki markup before filtering (defaults to on)

**STARTSUBST, STOPSUBST**

**Syntax:**

```plaintext
%STARTSUBST{...}%
...
%STOPSUBST%
```

Substitute text given inline. see SUBST.

**EXTRACT**

**Syntax:** `%EXTRACT{topic="..." ...}%`
Extract text from a topic. see SUBST.

**STARTEXTRACT, STOPEXTRACT**

Syntax:

```
%STARTEXTRACT{...}%
...
%STOPEXTRACT%
```

Extract content given inline. see SUBST.

**FORMATLIST**

Syntax: `%FORMATLIST{"<list>" ...}%`

Formats a list of items. The `<list>` argument is separated into items by using a split expression; each item is matched against a pattern and then formatted using a format string while being separated by a separator string; the result is prefixed with a header and appended with a footer in case the list is not empty.

- `<list>`: the list
- `split="...":` the split expression (default ",")
- `pattern="...":` pattern applied to each item (default "s(\*)s")
- `format="...":` the format string for each item (default "$1")
- `header="...":` header string
- `footer="...":` footer string
- `separator="...":` string to be inserted between list items
- `limit="...":` max number of items to be taken out of the list (default ",-1")
- `skip="...":` number of list items to skip, not adding them to the result
- `sort="on,off,alpha,num":` order of the formatted items
- `reverse="on,off":` reverse the sortion of the list
- `unique="on,off":` remove duplicates from the list
- `exclude="...":` remove list items that match this regular expression

The pattern string shall group matching substrings in the list item to which you can refer to by using $1, $2, ... in the format string. Any format string (format, header, footer) may contain variables $percnt$, $nop, $dollar and $n. The variable $index referse to the position number within the list being formatted; using $count in the footer or header argument refers to the total number of list elements.

**MAKEINDEX**

Syntax: `%MAKEINDEX{"<list>" ...}%`

Formats a list into a multi-column index like in MediaWiki's category topics. MAKEINDEX insert capitals as headlines to groups of sorted items. It will try to balance all columns equally, and keep track of breaks to prevent "schusterkinder", that is avoid isolated headlines at the bottom of a column.

Parameters:

- `<list>`: the list of items
- `split="...":` the split expression to separate the `<list>` into items (default ",")
- `pattern="...":` pattern applied to each item (default ".(*)")
- `cols="...":` maximum number of cols to split the list into
- `format="...":` format of each list item (default "$item")
• sort="on/off": sort the list (default "on")
• unique="on/off": removed duplicates (default "off")
• exclude="...": pattern to check against items in the list to be excluded
• reverse="on/off": reverse the list (default "off")
• header="...": format string to prepend to the result
• footer="...": format string to be appended to the result

Like in FORMATLIST the format parameter can make use of $1, $2, ... variables to match the groupings defined in the pattern argument (like in pattern="(.*) ;(.*) ;(.*)"). The first matched grouping $1 will be used as the $item to sort the list.

Examples

Secure Html

<a href="javascript:window.alert('Pop me up');">Pop me up</a>

Format Comments

<table>
<thead>
<tr>
<th>Date</th>
<th>Author</th>
<th>Headline</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 Aug 2005</td>
<td>TWiki:Main.MichaelDaum</td>
<td>This is a first comment. ...</td>
</tr>
<tr>
<td>22 Aug 2005</td>
<td>TWiki:Main.MichaelDaum</td>
<td>This is a second comment. ...</td>
</tr>
</tbody>
</table>

This is a first comment.
-- TWiki:Main.MichaelDaum on 22 Aug 2005

This is a second comment.
-- TWiki:Main.MichaelDaum on 22 Aug 2005

Extract table data

<table>
<thead>
<tr>
<th>Pos</th>
<th>Description</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>onsite troubleshooting</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>normalizing data to new format</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>testing server performace</td>
<td>5</td>
</tr>
</tbody>
</table>

• it took 3 hours onsite troubleshooting
• it took 10 hours normalizing data to new format
• it took 5 hours testing server performace

MAKEINDEX example

Compare with Philosophy articles needing attention

A
• Absolute (philosophy)
• Accident (philosophy)
• Actualism

F
• Futilitarianism

N
• Non-rigid designator

H
• Hermeneutics
• Hypokeimenon

O
• Object (philosophy)
• Ontic
Plugin Installation Instructions

- Download the ZIP file from twiki.org (see below)
- Unzip it in your twiki installation directory. Content:

<table>
<thead>
<tr>
<th>File:</th>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>data/TWiki/FilterPlugin.txt</td>
<td></td>
</tr>
<tr>
<td>lib/TWiki/Plugins/FilterPlugin/Core.pm</td>
<td></td>
</tr>
<tr>
<td>lib/TWiki/Plugins/FilterPlugin.pm</td>
<td></td>
</tr>
</tbody>
</table>

- Visit configure in your TWiki installation, and enable the plugin in the {Plugins} section.

Plugin Info

<table>
<thead>
<tr>
<th>Plugin Author</th>
<th>TWiki:Main.MichaelDaum</th>
</tr>
</thead>
<tbody>
<tr>
<td>License</td>
<td>GPL (GNU General Public License)</td>
</tr>
<tr>
<td>Plugin Version</td>
<td>2010-07-25</td>
</tr>
<tr>
<td>Change History</td>
<td></td>
</tr>
<tr>
<td>2010-07-25</td>
<td>TWikibug:Item6530 - doc improvements</td>
</tr>
<tr>
<td>2010-04-29</td>
<td>TWikibug:Item6433 - doc improvements</td>
</tr>
<tr>
<td>03 Jul 2008</td>
<td>sorting a list before, not after, formatting it in FORMATLIST</td>
</tr>
<tr>
<td>08 May 2008</td>
<td>added 'text' parameter to SUBST and EXTRACT; fixed SUBST as it was pretty useless before</td>
</tr>
<tr>
<td>07 Dec 2007</td>
<td>added MAKEINDEX, added lazy compilation</td>
</tr>
<tr>
<td>14 Sep 2007</td>
<td>added sorting for EXTRACT and SUBST</td>
</tr>
<tr>
<td>02 May 2007</td>
<td>using registerTagHandler() as far as possible; enhanced parameters to EXTRACT and SUBST</td>
</tr>
<tr>
<td>05 Feb 2007</td>
<td>fixed escapes in format strings; added better default value for max number of hits to prevent deep recursions on bad regexpressions</td>
</tr>
<tr>
<td>22 Jan 2007</td>
<td>fixed SUBST, added skip parameter to FORMATLIST</td>
</tr>
<tr>
<td>18 Dec 2006</td>
<td>using registerTagHandler for FORMATLIST</td>
</tr>
<tr>
<td>13 Oct 2006</td>
<td>fixed limit parameter in FORMATLIST</td>
</tr>
<tr>
<td>31 Aug 2006</td>
<td>added NO_PREFS_IN_TOPIC</td>
</tr>
<tr>
<td>15 Aug 2006</td>
<td>added use strict; and fixed revealed errors</td>
</tr>
<tr>
<td>14 Feb 2006</td>
<td>moved in FORMATLIST from the TWiki:Plugins/NatSkinPlugin; added escape variables to format strings</td>
</tr>
<tr>
<td>06 Dec 2005</td>
<td>fixed SUBST not to cut off the rest of the text</td>
</tr>
<tr>
<td>09 Nov 2005</td>
<td>fixed deep recursion using expand=&quot;on&quot;</td>
</tr>
<tr>
<td>22 Aug 2005</td>
<td>Initial version; added expand toggle</td>
</tr>
<tr>
<td>TWiki Dependency</td>
<td>TWiki::Plugins::VERSION 1.024</td>
</tr>
<tr>
<td>CPAN Dependencies</td>
<td>none</td>
</tr>
<tr>
<td>Other Dependencies</td>
<td>none</td>
</tr>
<tr>
<td>Perl Version</td>
<td>5.005</td>
</tr>
</tbody>
</table>
|TWiki:Plugins/Benchmark| GoodStyle nn%, FormattedSearch mn%, FilterPlugin nn%
|Plugin Home| http://TWiki.org/cgi-bin/view/Plugins/FilterPlugin|