

Locationmaps

Table of contents

1 About Locationmaps.....	2
2 Locationmap Examples.....	2
2.1 Retrieving an XDoc via HTTP.....	2
2.2 Retrieving HTML from a CMS.....	3
2.3 Link Rewriting.....	4

1. About Locationmaps

A locationmap defines a mapping from requests to location strings.

It was conceived to:

- Provide a more powerful means for semantic linking.
- Enable Forrest with a standard configuration override mechanism.
- Decouple the conceptual source space used by Cocoon from the concrete source space, so that a change in the concrete sources does not impact on the sitemap.

In other words, the locationmap enables content to be retrieved from a location that is defined in a locationmap file (located at `PROJECT_HOME/src/documentation/content/locationmap.xml` file. The advantage of this is that the URL seen by the user need bear no relation to the location of the source document, thus Forrest can separate the client URL space from the source document URL space. Thus, using the locationmap it is possible to pull together documents from many different locations into a single uniform site.

In addition, since the user URL space is now not connected to the source URL space it is possible to move source documents without breaking any existing user links.

The syntax of a locationmap resembles that of the sitemap, in that it also makes use of Matchers and Selectors to traverse a tree of nodes towards a leaf. In the case of the locationmap however the leaf does not identify a pipeline, but instead identifies a location string.

Apache Forrest looks in the standard location for the source file first (by default `PROJECT_HOME/src/documentation/content/xdocs/...`), if a file is found in this location then the locationmap is not consulted. However, if one is not found then the locationmap is used to resolve the source file.

2. Locationmap Examples

2.1. Retrieving an XDoc via HTTP

Normally files are generated from `{project:content.xdocs}`. Using the Locationmap it is possible to make these files come from elsewhere. This is useful if you want to pull files from different directory structures, or even remote repositories. For example, the following location match will match any request for a document below "remoteDemo" and will retrieve the source file from the Apache Forrest SVN repository (directly from the ASF's SVN webserver). This is an ideal way to ensure that your published docs are always up-to-date.

```
<match pattern="remoteDemo/**/*.xml">
  <location
src="http://svn.apache.org/repos/asf/forrest/trunk/site-author/content/xdocs/{1}.xml"
/>
</match>
```

Note that because this is a wildcard matcher you can request any page from the svn server simply by requesting `/remoteDemo/PATH/TO/FILE/FILENAME.html`. In addition, we can request any other output format available via Forrest plugins.

When including resources from remote repositories one has to be careful about things like `site` and `ext` linking. If the targets are not defined in the local `site.xml` file then these links will be broken.

Warning:

Because of the above limitation many of the links in the page generated from the above example are broken.

2.2. Retrieving HTML from a CMS

Using the locationmap you can use Forrest to retrieve data from a Content Management System (CMS), wither local or remote. As an example we will consider Apache Lenya.

[Apache Lenya](#) is a Java Open-Source Content Management System based on open standards such as XML and XSLT and the Apache Software Stack, in particular the XML publishing framework Apache Cocoon.

The following locationmap matcher will instruct Forrest to retrieve content from `http://lenya.zones.apache.org:8888/default/live/*.html?raw=true`, whenever a local URL of `lenya/**` is encountered.

```
<match pattern="lenya/**/*.xml">
  <location
src="http://lenya.zones.apache.org:8888/default/live/{1}.html?raw=true" />
</match>
```

However, since the source returned by this match is HTML and not XDoc we must convert this to our internal XDoc format before we can use it. We do this by adding the match below to our project's `sitemap.xmap` file.

```
<map:match pattern="lenya/**/*.xml">
  <map:generate type="html" src="{lm:{0}}"/>
  <map:transform src="{forrest:stylesheets}/html2document.xsl" />
  <map:serialize type="xml"/>
</map:match>
```

Since this snippet uses the HTML generator you must also ensure that your sitemap has the HTML generator component defined. That is, your sitemap must also include:

```
<map:components>
  <map:generators default="file">
    <map:generator name="html"
      src="org.apache.cocoon.generation.HTMLGenerator">
      <jtidy-config>WEB-INF/jtidy.properties</jtidy-config>
    </map:generator>
  </map:generators>
</map:components>
```

Since the HTML generator uses JTidy we need to make available a JTidy configuration file. This is placed in `PROJECT_HOME/src/documentation/WEB-INF/jtidy.properties` (the location can be changed in the above sitemap snippet). A sample config file is given below:

```
indent=yes
indent-spaces=8
wrap=72
markup=no
output-xml=no
input-xml=no
```

```
show-warnings=yes
numeric-entities=yes
quote-marks=yes
quote-nbsp=yes
quote-ampersand=yes
break-before-br=yes
uppercase-tags=no
uppercase-attributes=no
char-encoding=latin1
```

Note:

This requirement to add items to your project sitemap will be removed in a future version either by Lenya outputting XDoc, or by Forrest switching to using XHTML as its internal format, or through the development of a plugin for Lenya that will include the necessary processing (whichever comes first).

Warning:

This demo is an example only, it does not work fully work at this time. For example, absolute URLs in the source document need to be rewritten to ensure that they are matched by the locationmap.

2.3. Link Rewriting

The locationmap can be used to rewrite URLs when the page is generated. For example, when the locationmap has:

```
<match pattern="rewriteDemo/**">
  <location src="http://www.domain.org/{1}.xml" />
</match>
```

With the above match in the locationmap a link which has `href="lm:rewriteDemo/index"` will be rewritten to an offsite address at `domain.org`.