

Enable SEO URL in WebSphere Portal

Michele Buccarello

2/4/2015

Enable SEO URL in WebSphere Portal

Table of Contents

Abstract	2
REMOVING THE RICH STATE FROM PORTAL AND WCM CONTENT	2
Configuring portal for friendly url	3
Enable base url in theme	5
Edit the navigation jsp	6
Create a java filter for WCM url	7
Conclusion about rich state	9
Remove the context url /wps/portal.....	10
Create an apache reverse proxy.....	10
Enable modules	10
Configure Apache as reverse proxy.....	10
How it works.....	11
Routing rules.....	11
Payload management.....	12
Rewrite rule explanation	12
Proxy Pass and mod_substitute	13
CONCLUSION	13

Enable SEO URL in WebSphere Portal

Abstract

In WebSphere Portal the concept around the URL are complex because are not simply URL but are RICH URL. Portal URL are structured in this way:

`/wps/portal/home/shop/shoes/!ut/p/04_SB8K8xLLM9MS...`

Context Friendly path prefix Rich state suffix

Rich state is an xml that contain the portal state, this xml is gzipped and encoded with a modified base64 codec, the only way to decode the state in xml format is use the portal poc (piece of content) servlet. In this presentation <http://www.socialbizug.org/files/app#/file/443ae822-48ad-40d1-b106-8750f0e79447> you could find all information about all concept around the WebSphere Portal URL, is important read this pdf to understand what you lose if a customer want full friendly SEO url and not partial like normal Portal url.

What we do in this article?

We want display Portal url without context and rich state, in other words :

- From `/wps/portal/home/shop/shoes/!ut/p/104_05dsffds90..`
- To `/home/shop/shoes`

What are the assumption we need to archive this goal?

- 1) Portal Page doesn't have portlets that made a post to the same page or other page.
- 2) Portlet wiring attach rich state, don't use it!
- 3) WCM url are cleaned with a WCM url filter, without it all WCM url have rich state.

REMOVING THE RICH STATE FROM PORTAL AND WCM CONTENT

In this step we configure portal to remove the rich state from Portal and WCM content. The basics of this operations are described in this wikis document:

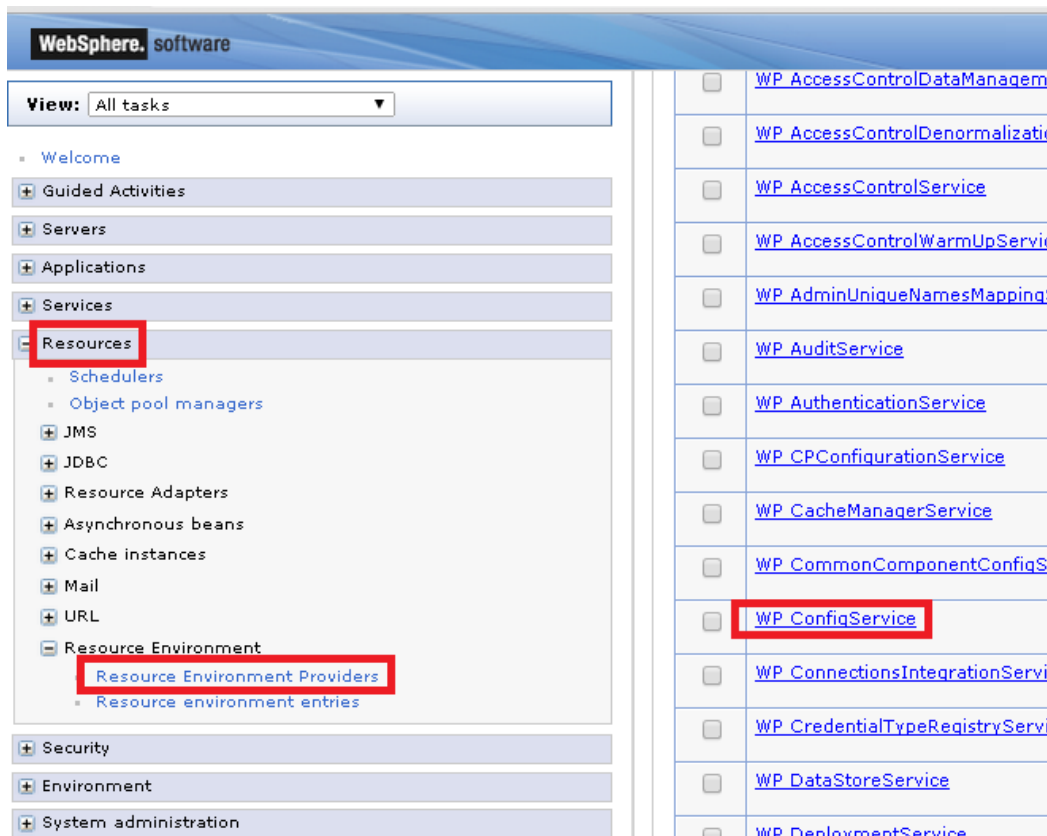
- http://www-01.ibm.com/support/knowledgecenter/SSHRKX_8.5.0/mp/admin-system/mp_friendly_short_url.dita?lang=en
- http://www-10.lotus.com/ldd/portalwiki.nsf/xpDocViewer.xsp?lookupName=IBM+Web+Content+Manager+8+Product+Documentation#action=openDocument&res_title=Example_2_Generate_a_friendly_URL_for_web_content_wcm8&content=pdcontent

Enable SEO URL in WebSphere Portal

- [http://www-10.lotus.com/ldd/portalwiki.nsf/dx/Implementing a clean URL in IBM WebSphere Portal 8-based WCM rendering](http://www-10.lotus.com/ldd/portalwiki.nsf/dx/Implementing_a_clean_URL_in_IBM_WebSphere_Portal_8-based_WCM_rendering)

Configuring portal for friendly url

In the administrative console under resource environment provider click on WP ConfigService



Click on custom properties

[Resource environment providers > WP ConfigService](#)

Use this page to configure a resource environment provider, which encapsulates the referenceables that convert resource environment entry resource objects. These objects can then be accessed by applications.

Configuration

General Properties	Additional Properties
<p>Scope cells:portalCell:nodes:portalNode:servers:WebSphere_Portal</p> <p>Name WP ConfigService</p> <p>Description <input type="text"/></p> <p>Apply OK Reset Cancel</p>	<ul style="list-style-type: none">ReferenceablesResource environment entriesCustom properties

Enable SEO URL in WebSphere Portal

Click on button new...

Resource environment providers > WP ConfigService > Custom properties

Use this page to specify custom properties that your enterprise information system (EIS) vendors require additional custom properties for data sources that access the database

Preferences

Maximum rows

200

Retain filter criteria

Show items at the following authorization group level:

All Roles

Apply

Reset

New...	Delete		
			
Select	Name	Value	

Add the property into the was

Resource environment providers

Resource environment providers > WP ConfigService > Custom properties > New...

Use this page to specify custom properties that your enterprise information system (EIS) requires for configure. For example, most database vendors require additional custom properties for data sources

Configuration

General Properties

Scope
cells:portalCell:nodes:portalNode:servers:WebSphere_Portal

Name
friendly.redirect.enabled

Value
false

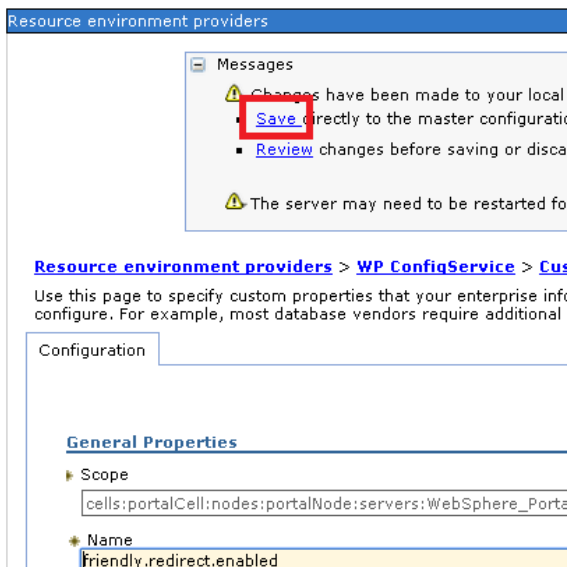
Description
If true send a redirect if the friendly prefix does not match the addressed page

Type
java.lang.String

Apply OK Reset Cancel

Enable SEO URL in WebSphere Portal

Click save



Enable base url in theme

In the Portal theme definition add the property

```
<parameter name="com.ibm.portal.theme.hasBaseURL" type="string" update="set">true</parameter>
```

add in the xmlaccess register script this parameter

```
<theme action="update" active="true" context-root="/CustomThemeDynamic" default="false" defaultskinref="com.ibm.portal.custom.theme.skin85">
  <localedata locale="en">
    <title>Custom Theme</title>
    <description/>
  </localedata>
  <allowed-skin skin="com.ibm.portal.custom.theme.skin85.hidden" update="set" />
  <allowed-skin skin="com.ibm.portal.custom.theme.skin85.hiddenplus" update="set" />
  <allowed-skin skin="com.ibm.portal.custom.theme.skin85.noskin" update="set" />
  <allowed-skin skin="com.ibm.portal.custom.theme.skin85.standard" update="set" />
  <parameter name="com.ibm.portal.theme.hasBaseURL" type="string" update="set">true</parameter>
  <parameter name="com.ibm.portal.friendly.name" type="string" update="set"><![CDATA[CustomThemeDynamic]]></parameter>
  <parameter name="com.ibm.portal.layout.template.href" type="string" update="set"><![CDATA[war:CustomThemeStatic/themes/CustomTheme85/1e]]></parameter>
  <parameter name="com.ibm.portal.theme.template.ref" type="string" update="set"><![CDATA[war:CustomThemeStatic/themes/CustomTheme85/]]></parameter>
  <parameter name="com.ibm.portal.theme.aggregationnodes" type="string" update="set"><![CDATA[ssa]]></parameter>
  <parameter name="resourceaggregation.profile" type="string" update="set"><![CDATA[profiles/profile_deferred.json]]></parameter>
</theme>
```

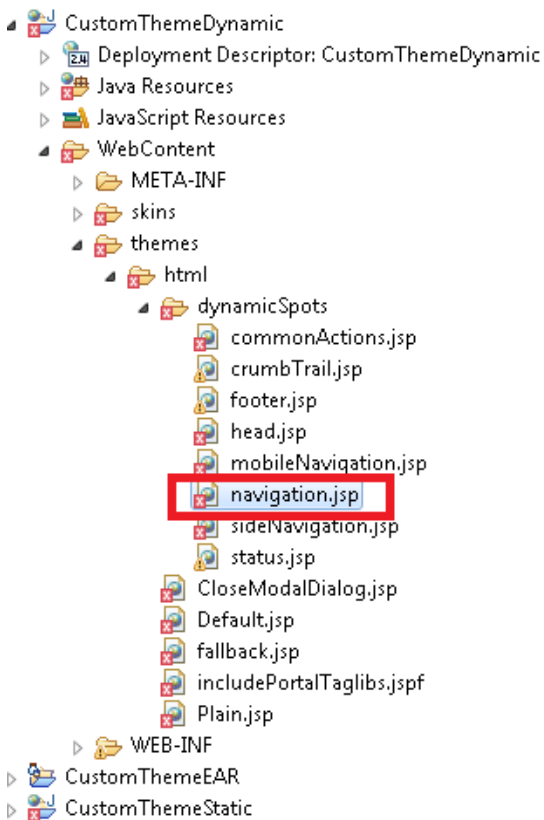
If you're theme is already registered retrieve the unique name and run the xml access below

```
<?xml version="1.0" encoding="UTF-8"?>
<request xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="PortalConfig_8.5.0.xsd" type="update">
  <portal action="locate">
    <theme action="update" uniquename="com.ibm.portal.custom.theme.theme85" >
      <parameter name="com.ibm.portal.theme.hasBaseURL" type="string" update="set">true</parameter>
    </theme>
  </portal>
</request>
```

Enable SEO URL in WebSphere Portal

Edit the navigation.jsp

In the navigation.jsp add keepNavigationalState="false" in the portal navigation tags, this particular property remove the rich state from the portal navigation URL.



Search this string:

- `<a href="?uri=nm:oid:${nodeID}"`

Replace with this:

- `<portal-navigation:urlGeneration contentNode="${nodeID}" keepNavigationalState="false" allowRelativeURL="true">
<a href="<%wpsURL.write(out);%>"`

See the image below.

```
<ul class="wpthemeNavList">  
</c:if>  
  <li class="wpthemeNavListItem wpthemeLeft"><:if test="${wp.selectionModel[node] != null}>  
    <a href="?uri=nm:oid:${nodeID}" class="wpthemeLeft"><:if test="${childrenStatus.count > 0}>  
      <portal-dynamic:closePage node="${node}"><a class="wpthemeClose wpthemeLeft" href="#">  
    </li>  
  <c:set var="childrenAvailable" value="true"/>  
</c:if>
```

Enable SEO URL in WebSphere Portal

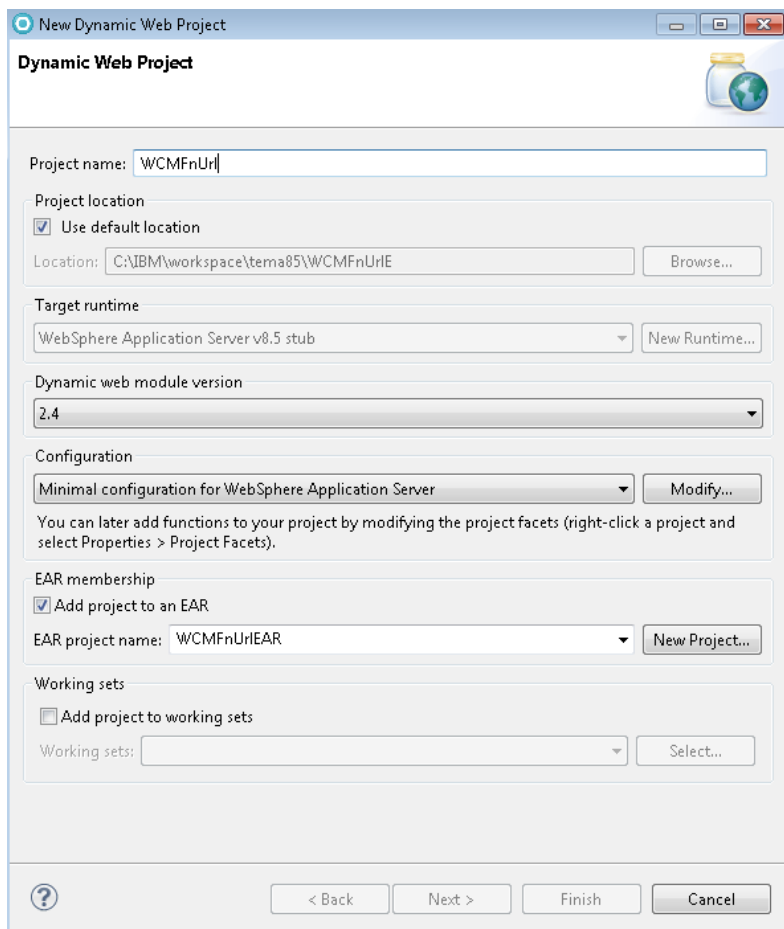
```
<li class="wpthemeNavListItem wpthemeLeft<<:if test="{wp.selectionModel[node] != null}"> wpthemeSe  
<portal-navigation:urlGeneration contentNode="{nodeID}" keepNavigationalState="false" allowRelativeURL="t  
<a href="{wpsURL.write(out);}" class="wpthemeLeft<<:if test="{childrenStatus.count == 1}"> w  
<portal-dynamiccui:closePage node="{node}"><a class="wpthemeClose wpthemeLeft" href="{closePag  
</li>  
::set var="childrenAvailable" value="true"/>
```

After the replaced string search the first "" and replace it with </portal-navigation:urlGeneration>, see the image below.

```
"wpthemeAccess"> <portal-fmt:text key="currently_selected" bundle="nls.commonTheme"/></span></c:if></span> </a>  
  
<li class="wpthemeNavListItem wpthemeLeft<<:if test="{wp.selectionModel[node] != null}">  
<portal-navigation:urlGeneration contentNode="{nodeID}" keepNavigationalState="false" allowRel  
<a href="{wpsURL.write(out);}" class="wpthemeLeft<<:if test="{childrenStatus.co  
</a>  
</portal-navigation:urlGeneration>  
<portal-dynamiccui:closePage node="{node}"><a class="wpthemeClose wpthemeLeft" href  
</li>
```

Create a java filter for WCM url

Create new dynamic project with the same projectname and ear name like the image below.



Enable SEO URL in WebSphere Portal

In src package create a package named:

- com.ibm.workplace.wcm.api.rewrite

create into it two java class:

- FriendlyUrlGenerationFilter.java
- FriendlyUrlGenerationFilterFactory.java

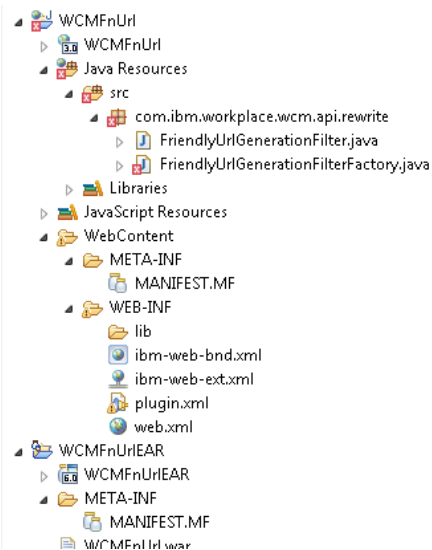
The functional sample content is located in this links:

- http://www-10.lotus.com/ldd/portalwiki.nsf/dx/Example_2_Filter_class_wcm8
- http://www-10.lotus.com/ldd/portalwiki.nsf/dx/Example_2_Filter_factory_class_wcm8

Under the webcontent →WEB-INF create the plugin.xml and write the information below:

```
<?xml version="1.0" encoding="UTF-8"?>
<plugin id="com.ibm.workplace.wcm.api.rewrite.plugin" name="URL generation filter"
version="1.0.0" provider-name="IBM">
  <extension
    point="com.ibm.workplace.wcm.api.ContentUrlGenerationFilter"
    id="UrlGenerationFilter">
    <factory
class="com.ibm.workplace.wcm.api.rewrite.FriendlyUrlGenerationFilterFactory"
weight="4"/>
  </extension>
</plugin>
```

The structure must be like the image below.



At this point you could install the ear in portal and set the startup behavior to a number upper than WCM ear startup number (normally 21 is a right number). We need to do this change to ensure the connections between WCM and the filter extension point.

Enable SEO URL in WebSphere Portal

The screenshot shows the WebSphere Portal configuration interface. On the left is a navigation tree with the following items: Welcome, Guided Activities, Servers, Applications, New Application, Application Type, WebSphere enterprise applications, Business-level applications, Assets, Global deployment settings, Services, Resources, Security, Environment, and System administration. The 'Applications' and 'WebSphere enterprise applications' items are highlighted with red boxes. The main content area shows the breadcrumb path 'Enterprise Applications > WCMFnUrEAR > Startup behavior' at the top, also highlighted with a red box. Below the breadcrumb is a description: 'Use this page to configure settings that determine how quickly an application starts compared'. Under the 'Configuration' tab, the 'General Properties' section contains a 'Startup order' field with the value '24' and a red box around it. There are two checkboxes: 'Launch application before server completes startup' (unchecked) and 'Create MBeans for resources' (checked). At the bottom of the configuration area are four buttons: 'Apply', 'OK', 'Reset', and 'Cancel', with 'Apply' highlighted by a red box.

Conclusion about rich state

With this four step after e Portal restart we successful remove the rich state from all page that use the theme with the customized navigation jsp and from all WCM contents.

Enable SEO URL in WebSphere Portal

Remove the context url /wps/portal

WebSphere Portal for security reason have two area, one for anonymous users and one for authenticated users:

- /wps/portal anonymous
- /wps/myportal authenticated

We remove the context root only for anonymous user to prevent security related problems.

Create an apache reverse proxy

To create an apache reverse proxy we a lot of way, personally I prefer opensuse because have a built in apache with a structured folders, in this guide http://en.opensuse.org/SDB:Apache_installation you could find all information about apache installation. After a successful installation we need to create a virtualhost as documented in the guide.

Enable modules

Apache 2.2 has a lot of built-in modules, opensuse by default don't enable it when you do the first installation, to enable it you need to edit the file `/etc/sysconfig/apache2`, inside it there is a variable called `APACHE_MODULES` append to the end of this variable this modules:

- rewrite http://httpd.apache.org/docs/2.2/mod/mod_rewrite.html
- proxy http://httpd.apache.org/docs/2.2/mod/mod_proxy.html
- proxy_http http://httpd.apache.org/docs/2.2/mod/mod_proxy_http.html
- substitute http://httpd.apache.org/docs/2.2/mod/mod_substitute.html

This modules are required to configure apache to act as a reverse proxy and manipulate http requests and payload. In the apache foundation site you could find all information for every modules.

Configure Apache as reverse proxy

Below the piece of configuration you need to configure routing and payload manipulation to the WebSphere Portal Installation request and response.

```
# used for mantain old link in external forum and site and with a 301 redirect  
RewriteRule ^/holidays/old-link?$ /holidays/new-link [L,R=301]
```

```
# this rewrite rule with proxy transparent use a regulare expresion to manage all context root different  
to the main route /wps/portal  
# the /wps escape all route to the servlets poc mypoc contenthandler mycontenthandler and all portlets  
# customThemeStatic and customThemeDynamic are the context root to my custom theme
```

Enable SEO URL in WebSphere Portal

```
RewriteCond %{REQUEST_URI} !/(wps/|customThemeStatic/|customThemeDynamic/)  
RewriteRule ^(.*)$ /wps/portal$1 [L,PT]
```

#this rewrite masquerade the landing page and remove the path to it

```
RewriteCond %{HTTP_HOST} ^www.example.it  
RewriteCond %{REQUEST_URI} ^(/)?$  
RewriteRule ^(/)?$ /wps/portal/home [PT,NC]
```

```
ProxyPass / http://www.example.it/
```

<Location / >

```
ProxyPassReverse http://www.example.it/
```

#this output filter simply parse the HTTP payload and remove /wps/portal from it, finally add /wps/portal only to the html tag base url

#this is required to the portal javascript framework

```
AddOutputFilterByType SUBSTITUTE text/html
```

remove /wps/portal from the payload

```
Substitute "s|/wps/portal| |nfq"
```

add to the base url tag /wps/portal to prevent issue with the portal javascript framework

```
Substitute "s|www.example.it|www.example.it/wps/portal|nfq"
```

</Location>

How it works

In this area we explain how the solution it works.

Routing rules

To archive this result we divide all uri in two groups:

- All uri different from “/wps/portal”
- All uri that match “/wps/portal”

In the first group there are:

- all common servlet like poc (/wps/poc) , mypoc (/wps/mypoc) , contenthandler (/wps/contenthandler), mycontenthandler (/wps/mycontenthandler)
- all authenticated portal uri (/wps/mycontenthandler)
- all portlets (/wps/PA_*)
- all predeployed portlet (/custom_context_root)
- all filters, OSGI plugin and so on (/custom_context_root)

Enable SEO URL in WebSphere Portal

In the second group we identify all uri that must be routed to /wps/portal. To identify this uri we must use a negative approach because with the mod_substitute we remove /wps/portal from the payload.

With these assumptions we made all rewrite rule to setup the right routing.

Payload management

There are two possible way to manage the payload:

- Application server side with a servlet filter
- HTTP Server side with HTTP Filter module

We choose the second way because:

- From apache 2.2 we have a built solution that provide an in memory payload rewrite with the mod_substitute.
- Performance scalability, it is quite simple to install more HTTP server than install more WebSphere Portal installation.
- A servlet filter interact with all HTTP request and response, portal have a set of built-in servlet filter, this means you must understand how to manage the servlet filter chain.
- With a simple regular expression we could manage all routing request directly from the HTTP Server.

Rewrite rule explanation

This rewrite url match with a negative condition all http request that must be routed to /wps/portal

```
RewriteCond %{REQUEST_URI} !/(wps/|customThemeStatic/|customThemeDynamic/)
```

```
RewriteRule ^(.*)$ /wps/portal$1 [L,PT]
```

This rewrite rule masquerade the homepage url

```
RewriteCond %{HTTP_HOST} ^www.example.it
```

```
RewriteCond %{REQUEST_URI} ^(/)?$
```

```
RewriteRule ^(/)?$ /wps/portal/home [PT,NC]
```

Enable SEO URL in WebSphere Portal

Proxy Pass and mod_substitute

All HTTP request skipped in the first rewrite rule are routed with the proxy pass directive

```
ProxyPass / http://www.example.it/
```

All response are manage inside the <location /> tag, inside this you could see the substitute that rewrite the payload.

```
<Location />
```

```
ProxyPassReverse http://www.example.it/
```

```
AddOutputFilterByType SUBSTITUTE text/html
```

```
Substitute "s|/wps/portal| |nfq"
```

```
Substitute "s|www.example.it|www.example.it/wps/portal|nfq"
```

```
</Location>
```

CONCLUSION

In this article you have a way and set of rules to implement SEO url in portal, but remember what we say in the abstract, the /wps/portal don't affect the SEO ranking of the site.