[轉]Tomcat Post / Get 中文編碼處理方法 (中文亂碼問題)

作者是 Administrator
週三, 12 三月 2014 16:34 - 最近更新 週三, 12 三月 2014 16:38

轉貼自：http://minchuanwang.blogspot.tw/2009/03/tomcat-post-get.html

最近在把原本佈署在 Resin 的 Java Web Application移植到 Tomcat 時，有許多原本正常的中文 get / post 功能，都變的異常，只要是透過 get / post 得到的資料，就會變亂碼

剛好也驗證了 Resin 在 Character Encoding的部分做的較好
而看看網路上大家的問題，也可以知道 Tomcat 在編碼處理的部分 真是讓很多人頭痛

主要的原因是，TOMCAT 預設都是用 ISO-8859-1 的編碼方式來傳遞資訊

這個問題，可以解決的方式整理如下：

1. JSP 頁面的編碼宣告需與實際儲存檔案時用的編碼一致

```
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
```

這是個很好玩的問題 ~ 有些人 在頁面宣告用 big5 or UTF-8 or... 字集
但是，檔案實際儲存的方式與該編碼不同，則 頁面當然會出現亂碼的問題

2. 自行轉碼
form method=POST or GET 時 因為 Tomcat預設編碼是 ISO-8859-1的關係
直接用 <%=request.getParameter("firstname")%> 取值，中文字會變亂碼
(你傳的明明是 UTF-8, 但是讀取時 確用 ISO-8859-1來解譯 .. 當然有問題囉)
要使用 <%=new String(request.getParameter("firstname").getBytes("ISO-8859-1"), "UTF-8")%>這種方式 來轉換編碼才會正常 (假設，頁面用的編碼為 UTF-8時)

3. 透過設定 server.xml URIEncoding的方式 來處理 GET 亂碼問題
在 tomcat 的設定檔內 ./conf/server.xml 修改下列設定的話，則透過 GET的方式傳遞參數的話，Tomcat會用你指定的編碼方式來對待用 get方法傳遞的參數，而不是使用預設的 ISO-8859-1 (get 的方式是透過 url parameter傳遞參數)
但是，因為只處理 URIEncoding, 所以若使用 Post 傳遞，還是會有問題

```
<Connector port="8080" protocol="HTTP/1.1"
           connectionTimeout="20000"
           redirectPort="8443" URIEncoding="UTF-8"/>
```

URIEncoding: This specifies the character encoding used to decode the URI bytes, after %xx decoding the URL. If not specified, ISO-8859-1 will be used.

P.S. 若使用 GET Method, 且設定了 URIEncoding 的話就不用再做上述二的自行轉碼動作，如果多加該處理，一樣會有亂碼的問題

4. 使用 request.setCharacterEncoding 來處理 GET / POST 的亂碼問題

作者是 Administrator
週三 , 12 三月 2014 16:34 - 最近更新 週三 , 12 三月 2014 16:38

在處理的頁面，加上下列的處理即可，明確的告訴 Tomcat request的編碼為何，不要一廂情願的使用 ISO-8859-1來解讀 ~

<%request.setCharacterEncoding("UTF-8");%>

javax.servlet.servletRequest.setCharacterEncoding(string env)
Overrides the name of the character encoding used in the body of this request. This method must be called prior to reading request parameters or reading input using getReader(). Otherwise, it has no effect.

5. 使用 filter 來處理 GET / POST 的亂碼問題
這個方法是最方便的，只要在 web.xml 裡面 透過 filter 的設定，讓每個網頁都能
透過 filter 的處理一體適用所有編碼問題的解決
方法如下：

a. 目前 Tomcat 提供的範例裡面，就有現成的 Encoding Filter, 位置如下：
.Tomcat 6.0webappsexamplesWEB-INFclassesfiltersSetCharacterEncodingFilter.class

b. 所以，只要把該 class 檔，copy 到你的 web application 的
.WEB-INFclassesfilter 下，再搭配 web.xml 的 filter 設定，
就可以讓所有符合 url pattern 的網頁，都能自動 apply所想要的編碼設定 (透過 encoding這個 parameter來設定)，是不是很方便呢 ?
(這個功能 就像是自動把上述第四點的 setCharacterEncoding 語句 自動 apply到各頁面囉)

以上 ~ 就從 Resin Porting 到 Tomcat 的經驗，當然是使用 filter 的方式 一次解決 ~ 不然的話 就得每個網頁每個網頁去改囉 ~

延伸資料：
FAQ/CharacterEncoding
http://jim.blogsome.com/2005/05/27/jsp-chinese-character-solution/
http://www.javaworld.com.tw/confluence/pages/viewpage.action?pageId=752