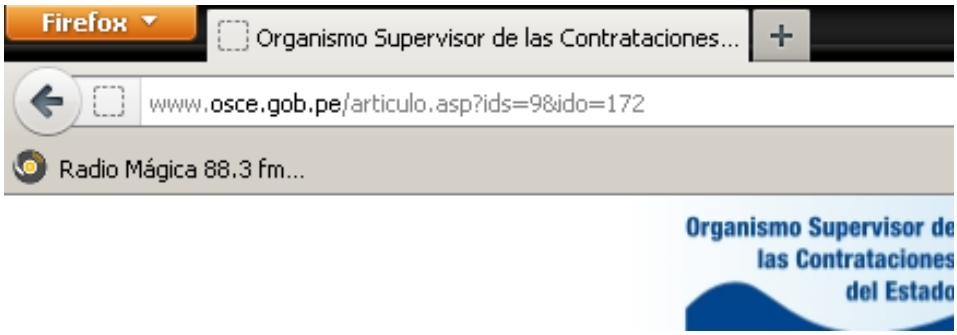
Tecnología de Desarrollo de Software II

Descarga de Resoluciones Emitidas por el Tribunal del Organismo Supervisor de las Contrataciones del Estado

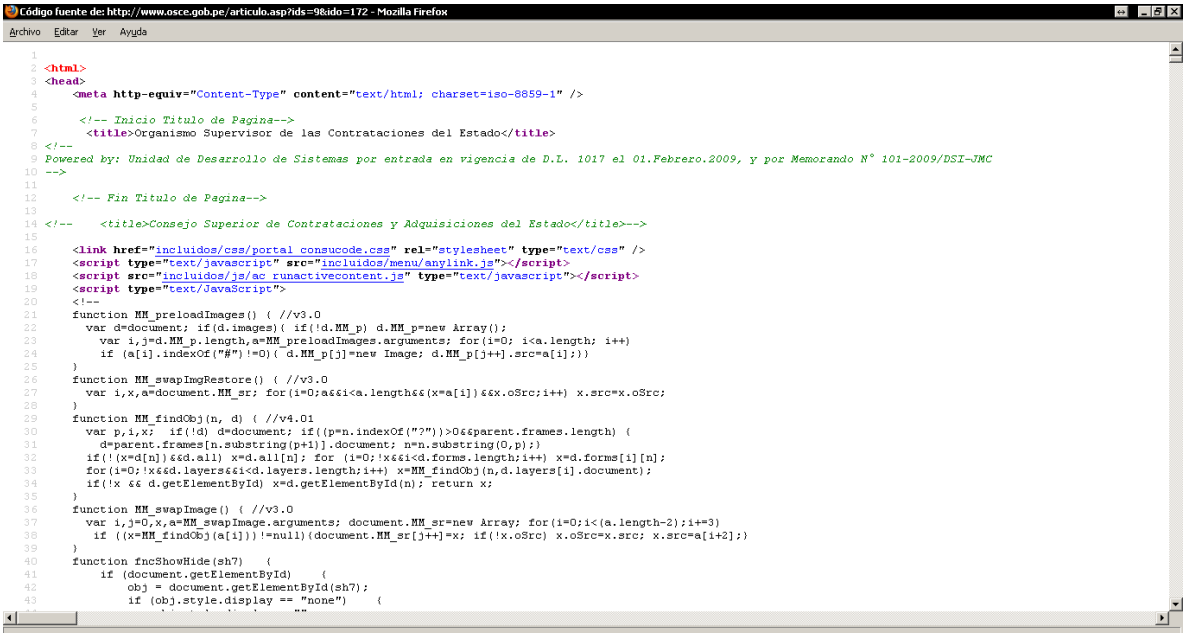
http://www.osce.gob.pe

Integrantes:
Muga Ampuero, Luis
Salinas Francia, Antonio

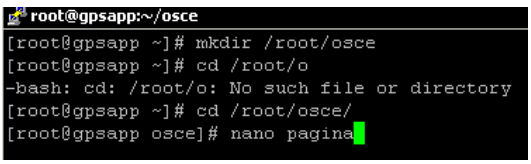
1. Ingresamos a la dirección **http://www.osce.gob.pe/articulo.asp?ids=9&ido=172** desde nuestro navegador Firefox.



2. Presionamos **Ctrl+U** y se nos mostrará el código fuente de la página.



3. Seleccionamos todo el código (**Ctrl+A**) y lo copiamos en un archivo en un nuevo archivo de texto (mkdir /root/osce, nano pagina):



4. Pegamos todo el contenido seleccionado y cerramos el archivo (**Ctrl+X**).

5. Al revisar el código nos fijamos que los enlaces figuran como un 5to campo separado por comillas (“ ”), hacemos uso de awk para separarlo y redirigimos la salida a un archivo llamado “enlaces”:

<td width="386">Desde Nº 001 - Nº 296</td>

Código: awk 'BEGIN {FS="\""} /DescargaHit*/ {print $4}' pagina > enlaces

6. Quedándonos las siguientes líneas en el archivo “enlaces”:

…
DescargaHit.asp?dir=userfiles/archivos/151-200.zip&nom=151-200.zip
DescargaHit.asp?dir=userfiles/archivos/101-150.zip&nom=101-150.zip
DescargaHit.asp?dir=userfiles/archivos/051 - 100.zip&nom=051 - 100.zip
DescargaHit.asp?dir=userfiles/archivos/RT001-050.zip&nom=RT001-050.zip
…

7. Agregamos el nombre de domino a los enlaces y lo guardamos en el archivo “enlacesrevisados”; quedándonos las líneas:

Código: awk '{print "http://www.osce.gob.pe/" $0}' enlaces > enlacesrevisados

http://www.osce.gob.pe/DescargaHit.asp?dir=userfiles/archivos/RESOL. (451-500) 2012.zip&nom=RESOL. (451-500) 2012.zip
http://www.osce.gob.pe/DescargaHit.asp?dir=userfiles/archivos/RESOL. (401-450) 2012.zip&nom=RESOL. (401-450) 2012.zip
http://www.osce.gob.pe/DescargaHit.asp?dir=userfiles/archivos/RESOL. (351-400) 2012.zip&nom=RESOL. (351-400) 2012.zip
http://www.osce.gob.pe/DescargaHit.asp?dir=userfiles/archivos/RESOL. (301-350) 2012.zip&nom=RESOL. (301-350) 2012.zip
http://www.osce.gob.pe/DescargaHit.asp?dir=userfiles/archivos/RESOL. (251-300) 2012.zip&nom=RESOL. (251-300) 2012.zip
http://www.osce.gob.pe/DescargaHit.asp?dir=userfiles/archivos/RESOL. (201-250) 2012.zip&nom=RESOL. (201-250) 2012.zip

8. Ahora convertimos cualquier espacio en blanco por su respectivo equivalente y hacemos uso de **wget** para descargar la lista de ficheros; borramos los archivos usados anteriormente:

Código:
awk '{gsub(" ","%20",$0); print $0;}' enlacesrevisados > enlacesrevisados2
wget –i enlacesrevisados2
rm –f pagina enlaces enlacesrevisados

9. Podemos automatizar esto poniendo todos los pasos en un archivo bash:

Código:

```
#!/bin/bash
awk 'BEGIN {FS="\""} /DescargaHit*/ {print $4}' pagina > enlaces
awk '{print "http://www.osce.gob.pe/" $0}' enlaces > enlacesrevisados
awk '{gsub(" ","%20",$0); print $0;}' enlacesrevisados > enlacesrevisados2
wget –i enlacesrevisados2
rm –f pagina enlaces enlacesrevisados
```