

Como automatizar y almacenar en un fichero xml queets de una cuenta en GNU social y convertirlos en una página web con la terminal/console

Autor: BibliotecaCopyleft
Junio 2015

Esta obra está bajo licencia Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
<https://creativecommons.org/licenses/by-sa/4.0/>

<https://gnusocial.no/bibliotecacopyleft/>
<http://androide400.webcindario.com/>
<http://media.espora.org/u/anonima/m/biblioteca-copyleft/>
<https://www.dropbox.com/s/5pzv1xqigpss1ti/bibliotecacopyleft.pdf>
<https://www.dropbox.com/s/nwv1arsk90fca64/xmlqueets.tar.xz>

GNU social es la alternativa libre a twitter, es software libre, descentralizado (puedes montarte tu propio nodo) y respeta tu privacidad, no como twitter que te la invade, he de decir esto porque habrá gente que llegue a éste manual con el buscador sin saber que es. El siguiente script va almacenando los queets de GNU social en un fichero xml y los transforma en una página web, funciona con RSS 2.0 no con la versión 1.0, puede servir para cualquier otro RSS 2.0 haciendo algún pequeño cambio, sólo hay que escribir en la terminal la palabra actualizaBiblio y lo hace. El RSS 2.0 almacena cierta cantidad de entradas y cuando se supera ese límite las va eliminando, este script es útil si queremos almacenar toda la información desde el principio de los RSS 2.0 en xml y/o transformarla en otro formato fácilmente, lo cual hace la información portable, evitando su pérdida y pudiendo usarla después.

actualizaBiblio.sh

```
#!/bin/bash
```

```
<<COMMENTBLOCK
```

Esta obra está bajo licencia Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
<https://creativecommons.org/licenses/by-sa/4.0/>

Autor: BibliotecaCopyleft
Junio 2015

<https://gnusocial.no/bibliotecacopyleft/>
<http://androide400.webcindario.com/>
<http://media.espora.org/u/anonima/m/biblioteca-copyleft/>
<https://www.dropbox.com/s/5pzv1xqigpss1ti/bibliotecacopyleft.pdf>

----- INSTRUCCIONES -----

-Editar y poner en .bashrc: alias actualizaBiblio='cd /home/toor/scripts;./actualizaBiblio.sh'
-Crear carpeta scripts en esa ruta y copiar este script a la carpeta, darle permisos de ejecución, con chmod 777 actualizaBiblio.sh

-Copiar tambien en la misma carpeta los ficheros bibliotecaXSLT.xslt, bibliotecaFormatoHtml.xslt

Ejecutar estos 2 comandos en la terminal para crear el fichero inicial biblioteca.xml:

```
cd /home/toor/scripts;
```

```
curl https://gnusocial.no/api/statuses/user_timeline/3517.rss > biblio.xml;xsltproc bibliotecaXSLT.xslt biblio.xml | sed 's/^\r//g' | sed -e '1,2d' > biblioteca.xml;rm biblio.xml;
```

biblioteca.xml es la base de datos en xml inicial a la que se iran sumando mas datos

Despues de haber hecho todo y ejecutado este script sin fallos tienen que existir en la carpeta scripts 5 ficheros: actualizaBiblio.sh, bibExtension.html, biblioteca.xml, bibliotecaFormatoHtml.xslt, bibliotecaXSLT.xslt, haga una copia de biblioteca.xml cada cierto tiempo por si existiese algun fallo

----- EXPLICACION DEL SCRIPT -----

Transforma rss descargado en xml
xsltproc bibliotecaXSLT.xslt descTemp.xml

Borra al principio de todas las lineas \r
sed 's/^\r//g'

Borra las 2 primeras lineas
sed -e '1,2d'

Mostrara los items nuevos
diff -c desc.xml biblioteca.xml

Muestra lineas entre *cadena**** y --cadena---**
sed -n -e '/^*\|*\|*\|*\|*\|*\|*\|*\|*\|*\|*\/,---.*---/p'

Borra lineas que contengan 3 * o - seguidos de un espacio, seguido por 0 o varios caracteres entre 0 y 9 seguido por una , seguido por 0 o varios caracteres entre 0 y 9 seguido por 1 espacio, seguido por 3 * o -
sed '/[*-][*-*][*-] [0-9]*\|[0-9]* [*-][*-*][*-]/d'

Borra al principio de todas las lineas los 2 primeros caracteres, incluyendo espacios
sed 's/^\|./g'

Borra las lineas que empiecen por 13 *
sed '/^\|*\|*\|*\|*\|*\|*\|*\|*\|*\|*/d'

Borra las ultimas 5 lineas
sed -e '\$d' | sed -e '\$d' | sed -e '\$d' | sed -e '\$d' | sed -e '\$d'

Concatena los items nuevos al principio del fichero y elimina los ficheros no necesarios
cat nuevos biblioteca.xml > bibliotecaTemp;
cp bibliotecaTemp biblioteca.xml;
rm descTemp.xml;rm desc.xml;rm nuevos;rm bibliotecaTemp;

Añade al principio del fichero en la linea 1 la cadena <rss>
sed -i '1i <rss>' bibTemp;

Añade al final del fichero en la ultima linea la cadena </rss>
sed -i '\$a </rss>' bibTemp;

Finalmente transforma en una pagina web la base de datos en xml y elimina una cadena de texto con el comando sed
xsltproc bibliotecaFormatoHtml.xslt bibTemp | sed 's/bibliotecacopyleft: //g'

COMMENTBLOCK

```
curl https://gnusocial.no/api/statuses/user_timeline/3517.rss > desc.xml;
cp desc.xml descTemp.xml;
xsltproc bibliotecaXSLT.xslt descTemp.xml | sed 's/^\\r//g' | sed -e '1,2d' > desc.xml;
diff -c desc.xml biblioteca.xml | sed -n -e '\^\*\*\*.\*\*\*\*/,/--.*---/p' | sed '/[*-][*-][*-] [0-9]*\,
[0-9]* [*-][*-][*-]/d' | sed 's/^..//g' | sed '/^\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/d' | sed -e '$d' | sed -e '$d' |
sed -e '$d' | sed -e '$d' | sed -e '$d' > nuevos;
cat nuevos biblioteca.xml > bibliotecaTemp;
cp bibliotecaTemp biblioteca.xml;
rm descTemp.xml;rm desc.xml;rm nuevos;rm bibliotecaTemp;
```

```
cp biblioteca.xml bibTemp;
sed -i '1i <rss>' bibTemp;
sed -i '$a </rss>' bibTemp;
xsltproc bibliotecaFormatoHtml.xslt bibTemp | sed 's/bibliotecacopyleft: //g' >
bibExtension.html;rm bibTemp;
```

bibliotecaXSLT.xslt

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
<!-- Esta obra está bajo licencia Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
https://creativecommons.org/licenses/by-sa/4.0/
Autor: BibliotecaCopyleft
Junio 2015
```

```
https://gnusocial.no/bibliotecacopyleft/
http://androide400.webcindario.com/
http://media.espora.org/u/anonima/m/biblioteca-copyleft/
https://www.dropbox.com/s/5pzv1xqigpss1ti/bibliotecacopyleft.pdf
```

```
-->
<xsl:template match=""/>
<xsl:apply-templates select="/rss/channel/item"/>
</xsl:template>
```

```
<xsl:template match="//item">
\r<item>
```

```

\r<id>
\r<xsl:value-of select="./link"/>
\r</id>
\r<title>
\r<xsl:value-of select="./title"/>
\r</title>
\r</item>

</xsl:template>
</xsl:stylesheet>

```

bibliotecaFormatoHtml.xslt

```

<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
<!-- Esta obra está bajo licencia Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
https://creativecommons.org/licenses/by-sa/4.0/
Autor: BibliotecaCopleft
Junio 2015

```

```

https://gnusocial.no/bibliotecacopleft/
http://androide400.webcindario.com/
http://media.espora.org/u/anonima/m/biblioteca-copleft/
https://www.dropbox.com/s/5pzv1xqigpss1ti/bibliotecacopleft.pdf
-->

```

```

<xsl:template match="/">
<meta charset="UTF-8"/>
<html>
<head></head>
<title>Biblioteca Extension</title>
<body bgcolor="black">
<center>
Esta obra está bajo licencia
<a href="https://creativecommons.org/licenses/by-sa/4.0/"
target="_blank">Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)</a>
<p><a href="https://gnusocial.no/bibliotecacopleft/"
target="_blank">https://gnusocial.no/bibliotecacopleft/</a></p>
<p><a href="http://androide400.webcindario.com/"
target="_blank">http://androide400.webcindario.com/</a></p>
<p><a href="http://media.espora.org/u/anonima/m/biblioteca-copleft/"
target="_blank">http://media.espora.org/u/anonima/m/biblioteca-copleft/</a></p>
<p><a href="https://www.dropbox.com/s/5pzv1xqigpss1ti/bibliotecacopleft.pdf"
target="_blank">https://www.dropbox.com/s/5pzv1xqigpss1ti/bibliotecacopleft.pdf</a></p>
</center>
<table border="1">
<th>Noticia</th>
<th>Contenido</th>
<xsl:apply-templates select="/rss"/>
</table>
</body>
</html>

```

```
</xsl:template>

<xsl:template match="//item">
  <tr>
  <td>
  <a>
  <xsl:attribute name="href">
  <xsl:value-of select="./id"/>
  </xsl:attribute>
  <xsl:attribute name="target">
  _blank
  </xsl:attribute>
  <xsl:value-of select="./id"/>
  </a>
  </td>

  <td><xsl:value-of select="./title"/></td>
  </tr>
</xsl:template>
</xsl:stylesheet>
```