Kako promijeniti kodnu stranicu na web sadržajima

uto, 2019-07-09 13:10 - Damir Mrkonjić

Na jednom od poslužitelja imamo web stranice koje su izrađene u kombinaciji html i php datoteka, korištenjem Dreamweavera. Stranice se izrađuju u tom obliku više godina i sadrže velik broj direktorija i datoteka. Ove stranice napravljene su u kodnoj stranici ISO-8859-2, što je i definirano odgovarajućom META oznakom u zaglavlju svake datoteke:

```
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-2">
```

Nakon nadogradnje na Debian jessie, pojavio se problem prikaza slova hrvatske abecede, što je jednostavno riješeno dodavanjem linije `php_value default_charset ISO-8859-2` u konfiguracijsku datoteku web poslužitelja (`/etc/apache2/sites-available/www.conf`). Godinama su stranice dobro radile, dok se nije pokazala potreba da se na stranci prikaže RSS sadržaj s jednog Wordpress poslužitelja. Sadržaj na tom poslužitelju je u kodnoj stranici UTF-8, pa se sadržaj nije dobro prikazivao. Kako smo već ranije naučili da i Apache očekuje stranice u takvom obliku, zaključili smo da je došlo vrijeme da se cjelokupni sadržaj prekodira po UTF-8 standardu.

Da se ne bi nešto poremetilo na web sadržaju, a i kako bi izbjegli da posjetitelji naših stranica gledaju promjene na stranicama (uz moguće pogreške) napravili smo privremeni web na adresi `www1.domena.hr` (to je zahtijevalo odgovarajući zapis u DNS konfiguraciji, kao i konfiguraciju posebnog web poslužitelja). Sadržaj weba ostao je nepromijenjen u direktoriju `/srv/www`, za probni web smo predvidjeli direktorij `www1`, i u njega prebacili sadržaj postojećeg:

```
cd /srv

cp -pr www www1
```

koristili smo opciju `-p` da sačuvamo vrijeme kreiranja datoteka. Ovo nam omogućava da u slučaju nekih pogrešaka jednostavno obrišemo direktorij `www1` i na gore opisan način prebacimo nepromijenjeni sadržaj.

Nakon ove pripreme trebalo je razmišljati kako napraviti promjenu kodne stranice u svim datotekama. Jedna od opcija je korištenje perla

```
perl -pi -e 's/charset=iso-8859-2/charset=utf-8/gi' *
```


Copyright ©2003.-2017. CARNet. Sva prava pridržana.
Mail to portal-team(at)CARNet.hr

Kako su stranice raspoređene u više direktorija, onda bi ovaj postupak trebalo ponoviti u svakom od direktorija. Ili raditi nešto ovakvo:

```
perl -pi -e 's/charset=iso-8859-2/charset=utf-8/gi' */*

perl -pi -e 's/charset=iso-8859-2/charset=utf-8/gi' */*/*

perl -pi -e 's/charset=iso-8859-2/charset=utf-8/gi' */*/*/*
```

Naravno, ovo bi trebalo ponoviti i za svih 5 slova hrvatske abecede (i velika i mala).

Očito je da treba pronaći jednostavnije rješenje. Uz pomoć Googlea saznali smo da postoji naredba iconv, i ne samo to - taj paket je već instaliran na našim poslužiteljima:

```
> dpkg -l | grep iconv

ii  libtext-iconv-perl 1.7-5+b2 i386 converts between character sets in Perl
```

S ovom naredbom sve je jednostavnije, konverzija kodne stranice u jednoj datoteci napravi se ovako:

```
iconv -f ISO-8859-2 -t UTF-8 index.html > index.html.novi
```

Sigurno je da nećemo raditi konverziju jedne po jedne datoteke, pa treba napraviti jednostavnu shell skriptu:

```
#! /bin/bash

for f in $(find . -name "*.html");

do

  filename="${f%.*}"

  echo -n "$f"

    iconv -f ISO-8859-2 -t UTF-8 $f > "${filename}_utf8.tex"

    mv "${filename}_utf8.tex" $f

    echo ": CONVERTED TO UTF-8."

done
```


Copyright ©2003.-2017. CARNet. Sva prava pridržana.
Mail to portal-team(at)CARNet.hr

Skripta vrlo jednostavno napravi konverziju kodne stranice u svim datotekama.

Ne smijemo zaboraviti ni META oznake na početku svake datoteke, koje izgledaju ovako

```
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-2">
```

Iskoristit ćemo ranije spomenutu mogućnost zamjene niza znakova s perlom i to ubaciti u našu skriptu:

```
#! /bin/bash

for f in $(find . -name "*.html");

do

  filename="${f%.*}"

  echo -n "$f"

  #

  perl -pi -e 's/charset=iso-8859-2/charset=utf-8/gi' $f

  #

  iconv -f ISO-8859-2 -t UTF-8 $f > "${filename}_utf8.tex"

  mv "${filename}_utf8.tex" $f

  echo ": CONVERTED TO UTF-8."

done
```

I s ovim smo obavili konverziju svih html datoteka. Pojavila se još jedna sitnica - poželjeli smo zadržati originalni datum kreiranja svake od datoteka.

Jedan od načina za saznati datum kreiranja datoteke je upotrebom naredbe date:

```
> /bin/date -R -r www/greska.html

Wed, 09 Aug 2006 13:36:23 +0200
```

S naredbom `touch -d` možemo svakoj datoteci promijeniti datum kreiranja. Preostalo je i ovo dodati u skriptu:

Copyright ©2003.-2017. CARNet. Sva prava pridržana.
Mail to portal-team(at)CARNet.hr

```
#! /bin/bash

for f in $(find . -name "*.html");

do

  filename="${f%.*}"

  echo -n "$f"

  #

  last_modif_ts=$(/bin/date -R -r $f)

  #

  perl -pi -e 's/charset=iso-8859-2/charset=utf-8/gi' $f

  iconv -f ISO-8859-2 -t UTF-8 $f > "${filename}_utf8.tex"

  mv "${filename}_utf8.tex" $f

  #

  touch -d "$last_modif_ts" $f

  #

  echo ": CONVERTED TO UTF-8."

done
```

Skriptu spremimo pod nekim imenom, npr. utf-conv.sh, i spremimo je negdje po želji, npr. u direktorij /srv

Nakon ovoga sve je spremno za konačnu konverziju, iz originalnog direktorija vraćamo originalne stranice u direktorij www1, obavimo konverziju, stari direktorij preimenujemo, kao i www1 u www.

ovako bi to izgledalo:

```
cd /srv

rm -r www1

cp -pr www www1

cd  www1
```


Copyright ©2003.-2017. CARNet. Sva prava pridržana.
Mail to portal-team(at)CARNet.hr

/srv/utf-conv.sh

cd /srv

mv www www-stari-iso

mv www1 www

I posao je završen... :)

Vijesti: Linux [1]
Kuharice: Linux [2]
Kategorije: sys.kuharica [3]
Vote: 5

Vaša ocjena: Nema Average: 5 (1 vote)

story_tag: web kodna stranica [4]

Source URL: https://sysportal.carnet.hr/node/1851

Links
[1] https://sysportal.carnet.hr/taxonomy/term/11
[2] https://sysportal.carnet.hr/taxonomy/term/17
[3] https://sysportal.carnet.hr/taxonomy/term/69
[4] https://sysportal.carnet.hr/taxonomy/term/330

Copyright ©2003.-2017. CARNet. Sva prava pridržana.
Mail to portal-team(at)CARNet.hr