```
#/bin/bash
#
# mksitemap.sh - create sitemaps by crawling a site using wget
# Copyright (C) 2010 Benjamin M.
#
# This program is free software: you can redistribute it and/or modify
# it under the terms of the GNU General Public License as published by
# the Free Software Foundation, either version 3 of the License, or
# (at your option) any later version.
#
# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
# GNU General Public License for more details.
#
# You should have received a copy of the GNU General Public License
# along with this program.  If not, see <http://www.gnu.org/licenses/>.

############################
##   F U N C T I O N S    ##
############################

# downloads most important files
# doesn't work with no filenames
function geturl
{

    wget --spider -m -nv -A "*.html,*.htm,*.php,*.txt" -o wget.log "$1" # -l 1

    if [ 0 -ne $? ]; then
        echo "Konnte Adresse $1 nicht auflösen." 2>/dev/stderr
        exit 2
    fi
}

# grep and parse URLs, then put them into an array
function makelist
{
    local  -a urllist

    if [ -z "$1" ]; then
        echo "Leere URL-Liste. Beende." 2>/dev/stderr
        exit 4
    fi

    # Aufbau:
    #
    urllist=(`  grep "http" wget.log  | \
                sed 's/^.*URL:\(http:.*\)\( \[.*\[.*$\)/\1/g' | \
                uniq | \
                sort | \
                xargs -0 `)
    echo ${urllist[*]}
}
```

```
 55
 56 # XML-File-Definition, bla
 57 write_header()
 58 {
 59     (
 60     cat <<AOF
 61 <?xml version='1.0' encoding='UTF-8'?>
 62 <urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
 63     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 64     xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
 65     http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
 66 AOF
 67     ) >> "$1"
 68 }
 69
 70 # End of XML file
 71 write_footer()
 72 {
 73     (
 74     cat <<EOF
 75 </urlset>
 76 EOF
 77     ) >> "$1"
 78 }
 79
 80 # Make some important sites more important.
 81 # Should be changed to fit your page's URLs.
 82 rate_url()
 83 {
 84     if [[ "$1" =~ .*News*0\.html ]] ||
 85         [[ "$1" =~ .*News\.html ]]; then
 86         echo "0.9"
 87     elif [[ "$1" =~ .*News.* ]]; then
 88         echo "0.5"
 89     elif [[ "$1" =~ .*Termine.* ]]; then
 90         echo "0.8"
 91     elif [[ "$1" =~ .*Kontakt.* ]]; then
 92         echo "0.2"
 93     elif [[ "$1" =~ .*Fotoalbum.* ]]; then
 94         echo "0.6"
 95     else
 96         echo "0.4"
 97     fi
 98 }
 99
100 show_version()
101 {
102 (
103     cat <<-EOV
104     mksitemaps Copyright (C) 2010 Benjamin M.
105     This program comes with ABSOLUTELY NO WARRANTY;
106     This is free software, and you are welcome to redistribute it
107     under certain conditions;
108
109 EOV
```

```
110 ) > /dev/stderr
111
112
113 }
114
115
116 ################
117 #  M  A  I  N  #
118 ################
119
120 # Variablen
121 declare -a urllist
122 priority=0
123 wgetlogfile=wget.log
124 sitemapfile=sitemap.xml
125
126 show_version
127
128 # Parameter
129 if [ -z $1 ]; then
130     echo "Keine URL angegeben - beende." 2>/dev/stderr
131     exit 1
132 fi
133
134 if [ -f "$wgetlogfile" ]; then
135     rm -f "$wgetlogfile"
136 fi
137 if [ -f "$sitemapfile" ]; then
138     rm -f "$sitemapfile"
139 fi
140
141 geturl "$1"
142 urllist=(`makelist "$wgetlogfile"`)
143
144 write_header "$sitemapfile"
145
146 i=0
147 while [ $i -lt ${#urllist[@]} ]; do
148     priority=`rate_url "${urllist[$i]}"`
149     echo -ne "Parsing link #"`echo -n $((i+1))`" von ${#urllist[@]}: [$priority]
${urllist[$i]}\n" > /dev/stderr
150     echo -ne "\t<url>\n" >> "$sitemapfile"
151     echo -ne "\t\t<loc>${urllist[$i]}</loc>\n" >> "$sitemapfile"
152     echo -ne "\t\t<priority>$priority</priority>\n" >> "$sitemapfile"
153     echo -ne "\t</url>\n" >> "$sitemapfile"
154     let i++
155 done
156
157 write_footer "$sitemapfile"
158
159 rm -f wget.log
160
161 exit 0
162
163 #
```