

```
#!/bin/bash

# CONDITIONS DE DIFFUSION ET LICENCE
# =====
# Copyright Sellig Zed, octobre 2016.
# sz(ad_ANTISPAM)myriapyle.eu
# ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
# Ce logiciel chklinxnn.sh est un programme testant la validité des liens internet dans le fichier fourni en argument.
# Ce logiciel est régi par la licence CeCILL soumise au droit français et respectant les principes de diffusion des logiciels libres. Vous pouvez
# utiliser, modifier et/ou redistribuer ce programme sous les conditions de la licence CeCILL telle que diffusée par le CEA, le CNRS et l'INRIA
# sur le site "http://www.cecill.info".
#
# En contrepartie de l'accessibilité au code source et des droits de copie, de modification et de redistribution accordés par cette licence,
# il n'est offert aux utilisateurs qu'une garantie limitée. Pour les mêmes raisons, seule une responsabilité restreinte pèse sur l'auteur du
# programme, le titulaire des droits patrimoniaux et les concédants successifs.
#
# A cet égard l'attention de l'utilisateur est attirée sur les risques associés au chargement, à l'utilisation, à la modification et/ou au
# développement et à la reproduction du logiciel par l'utilisateur étant donné sa spécificité de logiciel libre, qui peut le rendre complexe à
# manipuler et qui le réserve donc à des développeurs et des professionnels avertis possédant des connaissances informatiques approfondies. Les
# utilisateurs sont donc invités à charger et tester l'adéquation du logiciel à leurs besoins dans des conditions permettant d'assurer la
# sécurité de leurs systèmes et ou de leurs données et, plus généralement, à l'utiliser et l'exploiter dans les mêmes conditions de sécurité.
# Le fait que vous puissiez accéder à cet en-tête signifie que vous avez pris connaissance de la licence CeCILL-C, et que vous en avez accepté les
# termes.
# =====
# This software chklinxnn.sh is a computer program that tests the validity of web links included on the file given as command line argument.
# This software is governed by the CeCILL license under French law and abiding by the rules of distribution of free software. You can use,
# modify and/ or redistribute the software under the terms of the CeCILL license as circulated by CEA, CNRS and INRIA at the following URL
# "http://www.cecill.info".
#
# As a counterpart to the access to the source code and rights to copy, modify and redistribute granted by the license, users are provided only
# with a limited warranty and the software's author, the holder of the economic rights, and the successive licensors have only limited
# liability.
#
# In this respect, the user's attention is drawn to the risks associated with loading, using, modifying and/or developing or reproducing the
# software by the user in light of its specific status of free software, that may mean that it is complicated to manipulate, and that also
# therefore means that it is reserved for developers and experienced professionals having in-depth computer knowledge. Users are therefore
# encouraged to load and test the software's suitability as regards their requirements in conditions enabling the security of their systems and/or
# data to be ensured and, more generally, to use and operate it in the same conditions as regards security.
#
# The fact that you are presently reading this heading means that you have had knowledge of the CeCILL license and that you accept its terms.
# =====
# Historique des versions
# Les versions 1 et 2 étaient des versions de mise au point.
# La version 3 fut la première pleinement opérationnelle.
# La version 4 introduisit un test minimal de l'existence d'un argument donné à l'appel du script.
# Et des commentaires dans la liste d'instructions pour en faciliter la lecture. Ce fut la première mise en ligne.
# La version 5 fait la même chose que la 4 mais simplifie les manipulations de fichiers finaux et supprime les fichiers temporaires inutiles
# en fin de traitement.
# Version 6: adaptation aux changements des sorties des nouvelles versions de wget. Passage sous licence CeCILL. Plus d'information utilisateur.
# Version 7: ajout d'un compteur d'anomalies et utilisation dans le compte rendu d'exécution.
# =====
```

```

clear ;
echo "===== " ;
echo "Script bash chklinx de test de liens hypertextuels." ;
echo "Logiciel libre sous licence CeCILL (voir le texte source du script)." ;
echo "===== " ;

if [ $# = 1 ] # Si le nombre d'arguments de lancement du script est exactement 1
then # alors le traitement est poursuivi
{
echo "Lancement du script." ;
# La premiere partie comporte deux instructions
# La premiere extrait les liens du fichier fourni et les place temporairement dans messages.txt
sed 's/\(<\)/\n\1/g' $1 | grep '<[aA] ' |
    sed -n "s/^\(.*<[aA] [hH][rR][eE][fF].*\([hH][tT][tT][pP]s*:.[][a-zA-Z0-9._~:/?#@!$&'()*+;,=%-]*\).*>.*\)$/\2/gp" > messages.txt ;
# Expression rationnelle pour l'ensemble "caractères d'url" : [][a-zA-Z0-9._~:/?#@!$&'()*+;,=%-]
sed '/^$/d' messages.txt > liens.txt ; # Envoie le fichier extrait dans liens.txt en supprimant les lignes vides

# La seconde partie fait le test des liens extraits
ce_jour=$(date +%d-%B-%Y_%H:%M:%S) ; # Note la date dans la variable ce_jour
echo $ce_jour > messages.txt ; # et la place en haut du fichier messages.txt en reinitialisant ce dernier
echo $ce_jour > erreurs.txt ; # et la place en haut du fichier erreur.txt en initialisant ce dernier
taille=$(wc -l liens.txt | sed 's/^\([0-9]*\).*\1/'); # Fait compter et extrait le nombre de lignes dans le fichier de liens
echo "Nombre de liens dans le fichier : $taille ; # et en informe l'utilisateur au terminal
i=0 ; # initialise le compteur de tests de liens
nbAnomalies=0 ; # initialise le compteur d'anomalies

for ligne in `cat liens.txt` ; do # lit chaque ligne contenant un URL dans liens.txt
{ # et boucle : pour chacune de ces lignes :
((i++)) ; # Augmente le compteur de tests de liens
echo $i / "$taille ; # Affiche la progression du travail dans le terminal pour l'utilisateur
# l'instruction suivante fait le travail important
# wget en mode spider fait 2 tentatives de consultation de l'URL $ligne et met son commentaire dans erreur.tmp
wget -t 2 --spider $ligne --connect-timeout=60 --dns-timeout=60 -T 60 -o erreur.tmp ;
cat erreur.tmp >> messages.txt ; # On garde tous les messages dans messages.txt

sortietmp=$(cat erreur.tmp | grep 'HTTP.*[0123456789][0123456789][0123456789]') ; # Extraction de la ligne de requete
nosortietmp=$(echo $sortietmp | sed 's/\([^\0-9]*\([ ]\0-9*\)\)\2/g' | sed 's/^[^\0-9]*$//') ; # et de tous ses codes de reponse
sortietmp=$sortietmp$(cat erreur.tmp | grep '[Ii][nN][cC][oO][nN][nN][uU]') ; # Ces 4 lignes ajoutent a la variable sortietmp
sortietmp=$sortietmp$(cat erreur.tmp | grep '[Uu][nN][Kk][Nn][Oo][Ww][Nn]') ; # les possibles messages d'erreur
sortietmp=$sortietmp$(cat erreur.tmp | grep '[Ee][Rr][Rr][Ee][Uu][Rr]') ; # en french ou en anglais
sortietmp=$sortietmp$(cat erreur.tmp | grep '[Ee][Rr][Rr][Oo][Rr]') ; # en mijuscules ou en manuscules

if [ "$nosortietmp" != " 200" ] # Si le code de la réponse n'est pas celui de "[...] 200 OK", present et seul
then
sortietmp="■ Lien "$i": "$ligne" - "$sortietmp ; # alors on cree un message d'erreur ...
((nbAnomalies++)) ; # ... on incremente le compteur d'anomalies ...
echo " => Erreur n° $nbAnomalies !" ; # ... et on en informe l'utilisateur au terminal
else
sortietmp="" ; # sinon (si la sortie dit seulement ok) on vide le message d'erreur

```

```

fi
    echo $sortietmp >> erreurs.txt ;
} done

echo ;
cat erreurs.txt | sed '/^$/d' > erreur.tmp ;
rm erreurs.txt ;
mv erreur.tmp erreurs.txt ;
echo ;
echo "Les fichiers dans ce dossier sont les suivants:" ;
ls ;
echo "Les messages entiers sont dans le fichier message.txt" ;
echo "=====" ;
toutesLesErreurs=$(cat erreurs.txt) ;
if [ "$toutesLesErreurs" != "$ce_jour" ]
    then
        echo "$nbAnomalies anomalies survenues:" ;
        cat erreurs.txt ;
    else
        cat erreurs.txt ;
        echo "Pas de retours anormaux" ;
fi
echo "=====" ;
echo "Fin du script." ;
echo "=====" ;
}

else
echo "Il faut donner un nom de fichier." ;
echo "Syntaxe : $0 nom_de_fichier"
echo "=====" ;
echo "Fin du script." ;
echo "=====" ;
fi

# on place le message d'erreur en fin de erreurs.txt
# Fin de la boucle

# Saute une ligne au terminal
# Supprime les lignes vides de messages.txt pour remplir erreur.tmp
# supprime erreurs.txt
# renomme le fichier temporaire en erreurs.txt
# Saute une ligne au terminal avant affichage final du resultat
# Legende
# Montre notamment les fichiers crees par le script
# Signale ce fichier a toutes fins utiles

# Commodite pour le test suivant
# S'il existe un message signalant des erreurs...
#     ... alors, on le dit au terminal et on...
#     ... affiche le resultat
# sinon, ...
#     ... on affiche le contenu du fichier, c'est-a-dire la date, ...
#     ... et on signale l'absence d'anomalie.

# That's all folks.

# Si le nombre d'arguments de lancement du script n'est pas exactement 1
# alors un message invite l'utilisateur a revoir la syntaxe
# et le script s'interrompt.

```