

Projet Analyse des fichiers logs des sites Web de l'INRIA

Description des tables de la BD

Sergiu CHELCEA
Doru TANASA
Brigitte TROUSSE

1. Table **LOG**

LOG
IDRequest
IDNavigation
ISession
IDIP
IDUserAgent
IDDateTime
Duration
FileSize
IDStatus
IDURL
IDReferer

IDRequest - identifiant pour les requêtes (lignes du fichier csv), clé primaire

IDNavigation - identifiant des navigations, clé étrangère (table NAVIGATION)

ISession - identifiant des sessions, clé étrangère (table SESSION)

IDIP - identifiant des IPs, clé étrangère (table IP)

IDUserAgent - identifiant des agents, clé étrangère (table USER_AGENT)

IDDateTime - identifiant des instants, clé étrangère (table DATE_TIME)

Duration - durée de la requête (8^{ème} champs dans une ligne du fichier csv)

FileSize - dimension du fichier demandé (9^{ème} champs dans une ligne du fichier csv)

IDStatus - identifiant des statuts, clé étrangère (table STATUS)

IDURL - identifiant des URLs, clé étrangère (table URL)

IDReferer - identifiant des URL "referer", clé étrangère (table REFERER)

2. Table **NAVIGATION**

NAVIGATION
IDNavigation
OldIDNavigation
NbRequests
NbRequestsDif
NbInstance
DureeTotale
Repetition

IDNavigation - identifiant des navigations, clé primaire

OldIDNavigation - ancien identifiant des navigations, 1^{er} champ dans le fichier CSV

NbRequests - nombre des requêtes, pour la navigation

NbRequestsDif - nombre des requêtes différentes (URL distincts), pour la navigation

NbInstants - nombre des instances (couple Date_req, Time_req) distinctes

DureeTotale - durée de la navigation, en seconds (somme des durées de toutes les requêtes de la navigation)

Repetition (taux de répétition) = $1 - \frac{Nb\ Re\ quests}{Nb\ Re\ questsDif}$

3. Table **SESSION**

SESSION
IDSession
OldIDSession
NbNavigation
DureeTotale

IDSession - identifiant des sessions, clé primaire

OldIDSession - ancien identifiant des sessions, 2^{ème} champ dans le fichier csv

NbNavigations - nombre des navigations distinctes, pour la session

DureeTotale - durée totale de la session, en seconds (somme des durées de toutes les navigations de la session)

4. Table **IP** voir « Codage du nom de machine »

IP
IDIP
IP
Country
IDDomaine
UniteRecherche
IDProjet
IDService

IDIP - identifiant des IPs, clé primaire

IP - l'IP, 3^{ème} champ dans le fichier csv

Country - pays pour l'IP, voir « Codage du nom de machine »

IDDomaine - l'identifiant de l'IP anonyme

UniteRecherche - l'unité de recherche (« sophia », « rocquencourt », « inrialpes », « loria » ou « irisa »)

IDProjet - l'identifiant du projet auquel l'IP appartient

IDService - l'identifiant du service auquel l'IP appartient

5. Table **REAL_IP** - *non-utilisée*

REAL_IP
IDIP
REAL_IP

IDIP - identifiant des IPs, clé étrangère (table IP)

REAL_IP - l'IP réel, avant « anonymisation »

6. Table **USER_AGENT**

USER_AGENT
IDUserAgent
UserAgent
TypeNavigateur
Version
System

IDUserAgent - identifiant des agents, clé primaire

UserAgent - user agent de la requête, 5^{ème} champ du fichier csv

TypeNavigateur - nom du navigateur dans le User Agent (voir ex.)

Version - version du navigateur (voir ex.)

System - système d'exploitation de l'utilisateur

Ex. UserAgent : MS Internet Explorer v5 (Windows)

TypeNavigateur : MS Internet Explorer

Version : v5

System : Windows

7. Table **DATE_TIME**

DATE_TIME
IDDateTime
ReqDate
ReqTime
ReqHour
ReqMFM
ReqSFM

IDDateTime - identifiant des instants, clé primaire
ReqDate - date de la requête, 6^{ème} champ dans le fichier csv
ReqTime - temps de la requête, 7^{ème} champ dans le fichier csv
ReqHour - heure de la requête
ReqHour - minute (depuis minuit) de la requête
ReqSFM - seconds (depuis minuit) de la requête

8. Table **STATUS**

STATUS
IDStatus
Status

IDStatus - identifiant des statuts, clé primaire
Status - le statut, 10^{ème} champ dans le fichier csv

9. Table **URL**

URL
IDURL
URL
Site
FileExtension

IDURL - identifiant des URLs, clé primaire
URL - URL demandé, 11^{ème} champ dans le fichier csv
Site - nom du site de l'URL (voir ex.)
FileExtension - extension du fichier demandé de l'URL (voir ex.)

Ex. URL : <http://www-sop.inria.fr/axis/teaching/std.html>
 Rubrique syntactique 1 : axis (à mettre dans RSY)
 Rubrique syntactique 2 : teaching (à mettre dans RSY)
 Site : www-sop.inria.fr
 FileExtension : html

10. Table **RSY**

RSY
IDRsy Name Count

IDRsy - identifiant des rubriques syntactiques, clé primaire

Name - le nom de la rubrique syntactique

Count - le nombre des URLs contenant cette rubrique syntactique

11. Table **RSY_2_URL**

RSY_2_URL
IDURL IDRsy Level

Mets en correspondance les URLs (IDURL) et les rubriques syntactiques (IDRsy) :

IDURL - identifiant des URLs

IDRsy - identifiant d'une rubrique syntactique contenue dans l'URL (IDURL) au niveau « Level »

Level - le *niveau* de la rubrique syntactique (IDRsy) dans l'URL (IDURL)

12. Table **REFERER**

REFERER
IDReferer Referer

IDReferer - identifiant des referers, clé primaire

Referer - le referer, 12^{ème} champ dans le fichier csv