

03 - Traitement des indicateurs globaux

Indicateurs globaux d'utilisation d'esup-portail

**

Auteur : Vincent MATHIEU ([Université Nancy 2](#))

- [Indicateurs globaux d'utilisation d'esup-portail](#)
- [Généralités](#)
- [Description de l'utilitaire](#)
 - [Arguments de ce programme](#)
 - [Fichiers manipulés](#)
 - [ficstat](#)
 - [ficmatypes](#)
 - [ficrej](#)
 - [ficxml](#)
 - [ficsess](#)
 - [ficastlogins](#)

Généralités

Voir au préalable le [document de présentation générale](#) du mécanisme de statistiques esup-portail.

Les statistiques globales utilisent les enregistrements d'ouverture et de fermeture de sessions des utilisateurs (codes d'événement SSTART et SSTOP).

Voici un extrait d'enregistrement de ces événements :

```
2004-10-31 11:57:21,000 esup1 SSTART student 17 vmath999 20
2004-10-31 11:57:52,812 esup1 SSTART UNKNOWN 7 null 21
2004-10-31 11:58:10,859 esup1 SSTOP UNKNOWN 7 null 21
2004-10-31 11:59:26,046 esup1 SSTART employee 18 vmathieu 22
2004-10-31 11:59:34,156 esup1 SSTOP student 17 vmath999 20
2004-10-31 11:60:28,375 esup1 SSTOP employee 18 vmathieu 22
```

Dans cet exemple, les logs marquées UNKNOWN concernent un login avec le compte admin, qui n'est pas connu dans le LDAP de l'établissement (pas d'attribut eduPersonPrimaryAffiliation valué).

Description de l'utilitaire

C'est un programme perl, `traiteStatsSessionsEsup.pl`. Il utilise les bibliothèques `common.pm` et `common_sessions.pm`.

D'une manière générale, il exploite en entrée un fichier `stats.log` mensuel issu d'esup-portail (voir cet exemple de fichier [stats.log](#)), et produit en sortie un fichier xml qui synthétise l'activité du mois (voir cet exemple de résultat [stats.sess.xml](#)).

Il est exécuté depuis un script shell `traiteStatsEsup.sh`, qui réalise des pré-traitements (tri et 'nettoyage' du fichier en entrée, calcul du mois à traiter, ...).

Le fichier principal en entrée doit être trié par date. S'il contient des lignes avec des dates en dehors du mois choisi, ces lignes sont automatiquement rejetées. Les différentes statistiques sont fournies en fonction de catégories d'utilisateurs d'esup-portail.

Le code retour est égal à 0 s'il n'y a pas d'erreur fatale, différent de 0 sinon.

Ce chapitre décrit le fonctionnement général de cet utilitaire.

Arguments de ce programme

La syntaxe de lancement est la suivante :

```
traiteStatsSessionsEsup.pl [-mois <mois>] [-etablissement <etablissement>] [-nbsessactif <nbre>] [-ficstat <ficstat>] [-ficrej <ficrej>] [-ficxml <ficxml>] [-ficsess <ficsess>] [-ficmatypes <ficmatypes>] [-ficastlogin <ficastlogin>] [-noprintdays] [-noprinttrait]
```

Les arguments sont précédés du caractère "-"; aucun argument n'est obligatoire.

mois

Il détermine le mois concerné pour les statistiques. Le format est "AAAA-MM".
ex : -mois 2004-11
Par défaut, la première date rencontrée dans le fichier à traiter indique le mois de traitement

etablissement

facultatif. L'argument suivant permet d'indiquer l'établissement concerné dans le rapport xml de sortie.
Par défaut : univ.fr

nbsessactif

facultatif. donne le nombre de sessions ouvertes dans le mois pour qu'un utilisateur soit considéré comme "actif".
Par défaut : 5.

ficstat

facultatif. Indique le nom du fichier de stats issu d'esup-portail.
Par défaut : ./stats.log ou ./stats.AAAA-MM.log si l'argument -mois est passé.

ficrej

facultatif. Indique le nom de fichier contenant les lignes de ficstat rejetées.
Par défaut : ./stats.sess.rej ou ./stats.AAAA-MM.rej si l'argument -mois est passé.

ficxml

facultatif. Indique le nom de fichier de résultat principal.
Par défaut : ./stats.sess.AAAA-MM.xml

ficsess

facultatif. Donne le nom d'un fichier sous-produit du traitement, fourni pour d'éventuels retraitements.
Il contient un ligne par session utilisateur complète.
Par défaut : ./stats.session ou ./stats.AAAA-MM.session si l'argument -mois est passé

ficmatypes

facultatif. Indique le nom du fichier permettant la correspondance entre des catégories de personnes issues du fichier de stats esup-portail et les catégories du rapport final.
Par défaut : ./types.map

ficlastlogin

facultatif. Donne le nom d'un fichier sous-produit du traitement, utilisé en entrée-sortie, ou en sortie uniquement s'il n'existe pas lors de l'exécution de l'utilitaire.
C'est un fichier texte, contenant une ligne par utilisateur déjà logué dans l'ent ; le login et la date de dernier accès à l'ENT est écrite.

noprintdays

facultatif. Si présent, évite la sortie de l'élément <days> dans l'état xml principal, et de ses sous-éléments

noprinttrait

facultatif. Si présent, évite la sortie de l'élément <traitement> dans l'état xml principal, et de ses sous-éléments

Fichiers manipulés

Les fichiers manipulés par cet utilitaire sont les suivants :

ficstat

C'est le fichier de logs de sessions issu d'esup-portal. Voir cet exemple de [stats.log](#). Il est utilisé en entrée.

Dans le cas fort probable où plusieurs instances d'esup-portal fonctionnent en 'load-balancing', il faut s'assurer que le nom logique de chaque instance est bien renseigné (esup.host.logicalName du fichier esup.properties) afin d'éviter des risques d'anomalies.

Ce fichier doit impérativement être trié par date ascendante.

En cas de multiples serveurs, c'est la concaténation des fichiers des différents serveurs, triée ensuite par date.

ficmatypes

C'est un fichier qui fait la correspondance entre des catégories d'usagers issus du portail, et les catégories qu'on désire voir afficher en sortie. Il est utilisé en entrée. En voici un [exemple](#).

C'est un fichier texte, dont chaque ligne utile est composée de 3 champs obligatoires et un champ optionnel, séparés par un ou des caractères espace ou tabulation :

- categorie (ou type) en entree, donc tel qu'il est dans le fichier stats.log d'entree
- categorie (ou type) en sortie, donc tel qu'il apparaîtra dans le fichier resultant
- nombre d'utilisateurs : les utilisateurs potentiels de la categorie d'entrée
- identifiant de la categorie. Cette information n'est pas nécessaire pour l'utilitaire traiteStatsSessionsEsup.pl, mais servira de clé pour l'enrichissement d'une base de données des indicateurs esup.

En voici les règles :

- Une ligne commençant par "#" est ignorée
- Les 2 premiers champs sont balisés par des doubles quotes (")
- Une catégorie d'entrée ne peut apparaître qu'une seule fois
- Une catégorie de sortie peut apparaître plusieurs fois
- La catégorie d'entree peut être identique à la catégorie de sortie
- Ne pas nommer une catégorie de sortie avec la valeur " UNKNOWN "
- Si une categorie d'entree n'est pas listée, elle sera mise dans la catégorie " UNKNOWN "
- Si 2 catégories d'entrée ont la même catégorie de sortie, le nombre potentiel d'utilisateurs en sortie est la somme des 2 lignes
- L'identifiant est un champ optionnel. S'il n'est pas valué il prendra automatiquement la valeur du premier champ
- S'il y a regroupement de 2 populations (si la même catégorie de sortie est utilisée 2 fois), il faut alors que l'identifiant soit identique pour ces 2 lignes.

ficrej

C'est un fichier de rejet. Il contient les lignes de type ouverture/fermeture de sessions issues de ficstat, et rejetées par le traitement.

Voir cet [exemple](#). La cause de rejet est indiquée en début de ligne. Ca peut être :

BAD_MONTH

la date de la ligne ne correspond au mois désiré

END_SESSION_NOSTART

*fin de session sans début. Ne devrait arriver que pour le premier jour du mois, pour les sessions ayant débutées la veille et terminées le premier du mois.
Donc, normalement marginal.*

START_SESSION_NOEND

début de session sans fin de session. Peut se produire en cas d'arrêt d'un portail.

SESSION_DUP

*une ouverture de session de même rang, pour le même utilisateur, la journée et pour laquelle une fermeture de session n'a pas eu lieu précédemment s'est produite.
Ne devrait pas arriver dans une vie normale...*

ficxml

C'est le fichier de résultat désiré, au format xml. Voir cet [exemple](#).

Il donne d'abord une information globale pour le mois, puis pour chaque jours du mois, puis des informations concernant le déroulement du traitement.

La signification des différentes information est la suivante :

Pour les infos de stats d'utilisation :

- nbMaxUsers : le nombre maximum d'utilisateurs possibles pour la catégorie listée
 - nbUsers : le nombre d'utilisateurs ayant ouvert au moins une fois une session
 - nbActiveUsers : le nombre d'utilisateurs ayant ouvert au moins x sessions dans le mois
 - nbSess : nombre de sessions ouvertes dans la période
 - averageTimeSess : temps moyen d'une session pour la période, en minutes
 - nbSessActif : c'est le nombre de sessions mensuelles pour qu'un utilisateur soit considéré comme actif
- Pour les infos relatives au traitement de l'utilitaire :

- firstDay : premier jour de traitement dans le mois
- lastDay : dernier jour de traitement dans le mois
- nbLines : nombre de lignes de type ouverture/fermeture de sessions traitées
- nbServ : nombre d'occurrences logiques esup-portail traitées
- badMonth : nombre de rejets pour des ouvertures / fermetures de session en dehors du mois d'observation
- endWithNoStart : nombre de rejets pour des fin de sessions sans ouverture
- startWithNoEnd : nombre de débuts de session sans fin
- sessionDup : nombre de sessions dupliquées

ficssess

Défaut : stats.AAAA-MM.session. Voir cet [exemple](#).

C'est un sous-produit du traitement, fourni pour d'autres traitements éventuels.

C'est un fichier texte CSV, composé d'une ligne par session complète : les informations de sessions sont fournies, avec la durée en minutes.

Ce résultat n'est pas trié par l'utilitaire (en fait, il est écrit au fur et à mesure des fins de sessions).

ficlastlogins

Facultatif. Voir cet [exemple](#).

Ce paramètre donne le nom d'un fichier utilisé pour mémoriser les personnes qui se sont loguées dans l'ENT, avec la date de dernière connexion.

C'est également un sous-produit du traitement.

Si le fichier existe déjà lors du lancement de l'utilitaire, il est utilisé en entrée, puis mis à jour par le traitement.