

Production de statistiques locales d'accès aux ressources numériques : pré-requis

Document à destination des DSI des SCD utilisateurs justifiant la nécessité d'usage des log

Présentation

ezPAARSE permet de mesurer les accès aux ressources électroniques proposées par les plateformes web des éditeurs. Pour cela, il exploite les fichiers de journalisation (ou fichiers logs) d'accès collectés localement.

La « matière première » utilisée par ezPAARSE est donc constituée par des fichiers de log, générés par le serveur mandataire (généralement EZproxy) qui gère aussi l'authentification des utilisateurs.

Le serveur mandataire se situe entre le navigateur Web de l'utilisateur et la plateforme de l'éditeur. Il faut donc avoir accès aux fichiers de log des serveurs mandataires pour qu'ezPAARSE puisse les traiter.

ezPAARSE est développé en collaboration entre l'INIST-CNRS, le consortium Couperin.org et l'université de Lorraine. C'est un logiciel libre dont le code est [disponible](#).

Pré-requis

Pré-requis Techniques

En tant qu'organisme intéressé par l'utilisation régulière d'ezPAARSE, vos personnels utilisateurs du logiciel devront avoir la possibilité d'accéder aux résultats des traitements d'analyse des fichiers de log.

Il existe plusieurs cas de figure pour l'accès à ces résultats, ils sont précisés dans la partie *détail du fonctionnement* en annexe.

Le but d'ezPAARSE est de réaliser des statistiques d'usage locales. Une installation locale (à l'organisme) du logiciel est souhaitable, et peut facilement être effectuée en téléchargeant le logiciel sur la page : <http://analogist.couperin.org/ezpaarse/download>.

Lorsque les pré-requis sont remplis, l'installation de ce logiciel libre prend moins de 10 minutes dans un environnement Linux classique ou sur un poste windows.

Une instance nationale est aussi disponible à l'url : <http://ezpaarse.couperin.org/> pour les organismes qui ne disposent pas des ressources informatiques suffisantes.

Pour des tests, une instance nationale spécifique est aussi disponible : <http://ezpaarse-preprod.couperin.org/>

Pré-requis Juridiques

Les fichiers de log sont la matière première d'ezPAARSE. Par défaut, ezPAARSE reconnaît sans paramétrage particulier les formats de log des principaux serveurs mandataires utilisés pour la mise en ligne de ressources numériques dans leur configuration standard (EZproxy, apache, squid).

La législation sur les fichiers de log

Le rôle des DSI dans la politique de type gestion des journaux informatiques du service réseau de l'établissement est assimilable à celui d'un opérateur de communication électronique. Ce rôle consiste à conserver des traces dans un contexte de proxies web. Pour répondre à cette politique de conservation des traces, une déclaration CNIL existe déjà dans la plupart des établissements. Pour sa part, le CNRS dispose d'une politique de gestion de trace déclarée à la CNIL en 2004 pour l'ensemble de ses structures.

Les informations exploitées dans les fichiers de log par ezPAARSE sont celles de n'importe quel outil de mesure d'usage de site web. La mesure de l'usage des ressources mise à disposition des usagers peut entrer dans le cadre déjà fixé par la signature de la charte informatique lors de l'entrée dans l'organisme, celui-ci opérant pour ses propres personnels. Les utilisateurs des statistiques n'ont accès qu'aux résultats, de la même façon que s'ils consultent les données mesurant l'audience d'un site web.

La législation sur les données personnelles

Les données d'identification contenues dans les fichiers de log exploités par ezPAARSE concernent l'adresse IP de la machine et le ticket d'authentification de l'utilisateur connecté. Ces deux informations sont utilisées à des fins de dédoublement des accès de façon à pouvoir produire des rapports conforme aux recommandations COUNTER.

Le ticket d'authentification est propre à chaque organisme et n'est pas forcément en lien avec des données personnelles. Dans la plupart des cas, cette problématique est déjà traitée dans le cadre de la charte informatique.

Recommandation aux DSI des organismes

- 1. mettre à disposition des responsables des statistiques d'usage un accès aux résultats des traitements effectués par le logiciel ezPAARSE sur les fichiers de log des serveurs mandataires donnant l'accès aux ressources numériques**
- 2. automatiser cette production de résultats**

Annexe technique

L'accès aux fichiers de log

L'accès aux fichiers de log est un aspect crucial de l'usage d'ezPAARSE. Cet aspect est du ressort des DSI, et l'objectif de ce document est d'explicitier le rôle de chacun dans l'usage du logiciel.

Dans le cas de l'usage d'ezPAARSE, le rôle des DSI est double et doit permettre :

- la mise à disposition d'une instance du logiciel (**facultatif**, voir la partie *détail du fonctionnement*, la mise à jour du logiciel est automatisée)
- l'accès aux résultats du traitement des fichiers de log (**obligatoire**, voir la partie *détail du fonctionnement*)

La mise en place de l'automatisation des traitements permet de respecter la politique de gestion des journaux informatiques. Les utilisateurs n'ont accès qu'aux résultats, de la même façon que s'ils consultent les données mesurant l'audience d'un site web.

La phase de mise en place du processus de production des résultats peut demander des échanges d'échantillons de fichiers de log à des fins de vérification et de réglage des paramètres de traitement. L'écosystème d'ezPAARSE contient les [outils nécessaires](#) pour réaliser l'anonymisation automatique des données en cas de besoin.

Il peut apparaître utile aux DSI de déléguer provisoirement l'accès aux fichiers de log aux utilisateurs le temps qu'ils affinent les réglages leur permettant de déterminer les informations qu'ils souhaitent retrouver dans les fichiers résultats issus des traitements. Ces réglages pourront être enregistrés sous forme de paramètres prédéfinis et servir dans les traitements automatisés.

Le contenu des fichiers de log

EzPAARSE n'a pas besoin de données particulières dans les fichiers de log autres que celles qui se trouvent classiquement dans les fichiers journaux destinés à l'analyse d'audience d'un site web. Si la configuration standard du serveur mandataire a été conservée, les fichiers peuvent être traités directement (hormis squid, logiciel assez agé, qui demande un ajustement).

Les fichiers de log peuvent être traités "tel que" sans ajout ou retrait d'aucune sorte, sous forme brute ou compressée. L'ajout d'information peut se justifier dans le cas d'organismes souhaitant disposer d'éléments supplémentaires comme de données de composantes qui sont alors disponibles directement dans les fichiers résultats d'ezPAARSE, mais cet ajout peut aussi se faire *a posteriori*.

Cependant, en cas de paramétrage particulier du format des log du serveur mandataire, la reconnaissance des lignes de log par ezPAARSE peut être configurée en conséquence. Même si cette configuration est [entièrement documentée](#), il est recommandé de prendre contact avec l'équipe ezPAARSE (ezpaarse@couperin.org) pour faciliter cette étape importante qui n'est faite qu'une seule fois.

Détail du fonctionnement

Traitements lancés par l'utilisateur

Les traitements peuvent être lancés manuellement par un utilisateur via un formulaire web. L'utilisateur choisit les fichiers de log qu'il veut traiter et les soumet. Il obtient un résultat téléchargeable sur son poste, exploitable via un tableur (Excel, OpenOffice). Les fichiers de log peuvent être dans un format compressé.

Dans la plupart des cas, avec une configuration standard du serveur mandataire, le format de log est reconnu par ezPAARSE. Sinon, l'équipe ezPAARSE se charge de créer et rendre disponibles dans le formulaire des paramètres adaptés aux organismes qui en font la demande.

Sur son poste

Il est possible d'installer une version windows locale d'ezPAARSE qui s'utilise via un navigateur comme tout autre instance. Une installation MacOS est aussi possible et se rapproche de celle de l'environnement Linux.

Sur une instance locale

Si l'organisme dispose des ressources informatiques sous linux, une instance locale peut y être installée et mise à disposition d'un ensemble d'utilisateurs locaux.

Sur l'instance nationale

L'instance nationale est disponible après création d'un compte utilisateur à la première connexion. Aucun historique ou fichier de log n'est conservé plus d'une journée (durée jugée suffisante pour que l'utilisateur télécharge ses résultats).

Traitements lancés automatiquement

ezPAARSE a été conçu pour pouvoir être utilisé de façon automatique grâce à une API. L'idée est qu'à partir du moment où des fichiers de log sont générés de façon périodique, ils peuvent être envoyés régulièrement à une instance ezPAARSE. L'écosystème d'ezPAARSE contient des outils permettant de faciliter ce travail (traitement en masse de fichiers, génération de rapports statiques d'exécution). L'équipe ezpaarse (ezpaarse@couperin.org) peut assister les DSI souhaitant automatiser le traitement de leurs données.

Sur une instance locale

N'importe quelle instance locale peut-être utilisée via l'API.

Sur l'instance nationale

L'instance nationale peut recevoir des traitements automatiques via l'API après création d'un compte utilisateur par l'interface web.

Mise à jour du logiciel

La connexion à ezPAARSE se fait par un compte utilisateur avec des droits associés.
L'administrateur peut mettre à jour le logiciel par un simple clic.