

ezPaarse visite les logs

Issu d'une collaboration étroite depuis novembre 2012 entre l'Inist-CNRS, Couperin.org et l'université de Lorraine, ezPaarse est un logiciel libre¹ et multiplateforme qui, à partir de l'exploitation des logs, permet aux bibliothèques de disposer de statistiques d'usage (qu'est-ce qui est consulté et par quelle catégorie d'utilisateurs), de les comparer aux données fournies par les éditeurs et de proposer des outils d'aide à la décision. Décryptage.

ezPaarse se présente sous la forme d'une application web² disposant d'un formulaire et d'une interface de programmation (API) pour l'ingestion manuelle ou automatique de fichiers de logs d'accès aux ressources numériques proposées par les plateformes des éditeurs de littérature scientifique. Cette application est complétée par un portail wiki dénommé AnalogIST³ où, en sus d'informations sur le projet, est publié le travail d'analyse des plateformes⁴ des fournisseurs qui sert de base à la création des parseurs pour enrichir les capacités de reconnaissance d'ezPaarse. Un parseur est un petit programme, spécifique à chaque plateforme, qui se charge de découper une ligne de log pour en extraire des informations (par exemple : l'ISSN de la revue consultée, le type de ressource, etc.). Les contributions extérieures sont encouragées pour augmenter la couverture d'ezPaarse : toutes les personnes intéressées peuvent facilement analyser une nouvelle plateforme en s'appuyant sur les outils⁵ et procédures⁶ mis à disposition. Les plateformes en ligne sont uniques, même s'il peut exister entre elles des régularités ou des ressemblances. Chacune d'elles a donc besoin d'être spécifiquement analysée par un collègue, bon connaisseur de la ressource, qui saura dresser un catalogue complet des actions qu'un usager peut y accomplir dans le cadre de l'abonnement souscrit et faire correspondre des éléments des URL à des informations sur la ressource consultée.

ENJEUX ET BASES DE FONCTIONNEMENT

ezPaarse tente de répondre au besoin exprimé par les bibliothèques et leurs tutelles de disposer d'outils pour justifier et évaluer la bonne utilisation des sommes importantes consacrées aux abonnements académiques et scientifiques. Les données fournies par les éditeurs ne suffisent souvent pas à atteindre ces buts parce qu'elles :

- ne sont tout simplement pas disponibles ;
 - sont disponibles, mais non certifiées Counter⁷, ce qui les rend peu fiables et difficiles à comparer d'une ressource à l'autre ;
 - sont disponibles et certifiées Counter, mais doivent être nettoyées et enrichies pour en tirer des informations exploitables, par exemple sur les profils des utilisateurs.
- Pour fonctionner, ezPaarse s'appuie sur un maté-

riau courant, mais encore peu systématiquement exploité : les fichiers logs des serveurs proxy. On y trouve, ligne par ligne, le détail des accès réalisés par les utilisateurs accrédités aux ressources numériques. C'est l'analyse de ces fichiers qui permet de mesurer l'activité de consultation des ressources numériques et de disposer d'informations précises sur les événements de consultation identifiés.

Une fois correctement réglé, ezPaarse peut être automatisé et délivrer des rapports complets sur une base régulière. Cette mise en production est d'ores et déjà effectuée à l'Inist-CNRS et à l'université de Lorraine et en cours dans une vingtaine d'autres établissements.

UN FONCTIONNEMENT MUTUALISÉ ET COLLABORATIF

Un fonctionnement agile pour l'implémentation de nouvelles fonctionnalités

Les méthodes agiles sont des groupes de pratiques de développement en informatique (conception de logiciel) pouvant s'appliquer à divers types de projets. La méthode agile Scrum⁸ utilisée pour développer et faire évoluer ezPaarse permet de livrer régulièrement et rapidement, toutes les six semaines en moyenne, des versions fonctionnelles. Les retours des utilisateurs sont immédiatement pris en compte et orientent sans cesse le projet vers les besoins constatés. On peut dresser un parallèle entre les itérations agiles et une série télévisée : les épisodes se succèdent et sont regroupés en « saisons ». L'intérêt des utilisateurs est maintenu et leur implication grandit au fil des développements. Dans le cas d'ezPaarse, la « saison 1 » a duré un an : la version 1.0 du logiciel a été livrée en 10 itérations (versions 0.1 à 1.0) thématiques. La « saison 2 » se terminera en décembre 2014 et sera suivie en 2015 d'une troisième et dernière saison sur la thématique des statistiques de l'*open access* et des archives institutionnelles.

Une mutualisation des efforts et les moyens d'y parvenir

Des outils collaboratifs ouverts sont présents à tous les endroits pertinents, depuis l'analyse des plateformes, à travers :

- un portail wiki : AnalogIST (qui permet la participation sur simple inscription) ;

[1] C'est la licence CeCILL (www.cecill.info/licences/Licence_CeCILL_V2-fr.html), compatible avec la GPL (General Public Licence) en droit français, qui a été choisie. Le code source est disponible en ligne sur : <https://github.com/ezpaarse-project/ezpaarse>

[2] Une instance publique du logiciel est utilisable sur <http://ezpaarse.couperin.org> sur simple inscription.

[3] <http://analogist.couperin.org>

[4] Voir la liste complète des plateformes en cours d'analyse ou analysées sur : <http://analogist.couperin.org/platforms/start>

[5] <https://trello.com/b/wEaLnz8d/ezpaarse-analogist-suivi-des-plateformes> et <http://ezpaarse.couperin.org/doc/tools.html#platform-init>

[6] <http://analogist.couperin.org/platforms/contribute/start>

[7] Voir www.projectcounter.org et <http://couperin.org/groupe-de-travail-et-projets-deap/statistiques-dusage/counter>

[8] [https://fr.wikipedia.org/wiki/Scrum_\(methode\)](https://fr.wikipedia.org/wiki/Scrum_(methode))

- un outil de collaboration en ligne en temps réel : un *board* Trello⁹ pour le suivi fin de l'état d'avancement des analyses ;
- une liste de diffusion publique¹⁰ qui intègre la participation à la création de parseurs, ou même la contribution au code source, grâce à un répertoire sur Github¹¹ qui permet la récupération libre de la totalité du code du logiciel, mais aussi la participation à de nouvelles fonctionnalités¹². En plus des outils décrits ci-dessus, nous tentons d'assurer une présence active sur Twitter¹³, Google+¹⁴ et Facebook¹⁵ pour diffuser rapidement et simplement des informations d'actualité, techniques ou non, sur les avancées du travail et informer sur les événements auxquels nous participons. Nous accordons beaucoup d'importance aux différents événements, formations et ateliers pour faire découvrir ce projet et y impliquer les collègues. L'atelier organisé lors du BibCamp ADBU (juin 2014) aura été une bonne occasion pour faire connaître ezParse et l'environnement AnalogIST, présenter l'équipe et la structure mutualisée du travail et, surtout, entamer et parfois finaliser des analyses de plateformes dont les parseurs ont depuis été développés. Sur le même modèle, nous souhaitons organiser des Hackathon/Parsathon/ezCamp pour sensibiliser les professionnels à l'outil et au processus d'analyse des plateformes. Deux journées, dont les détails seront communiqués bientôt, devraient avoir lieu en mars puis en juin 2015.

Des ouvertures internationales

Les préalables techniques pour utiliser ezParse sont très basiques et notre logiciel peut être adopté par une large communauté. Nous l'avons donc présenté lors de conférences et de salons et notamment à Liber (Riga), où un « *Liber Award for Library Innovation* » nous a été décerné, à l'Ifla (Lyon), à l'Icolc (Lisbonne) et prochainement lors de la conférence UKSG, en mars 2015.

UNE DYNAMIQUE NATIONALE ET INTERNATIONALE

Il existe des articulations possibles entre ezParse, la base de connaissance nationale Bacon et le hub de métadonnées de l'Abes, qui ont été discutées et délimitées lors d'une réunion commune des équipes concernées, début septembre 2014. ezParse collecte actuellement des données sur les plateformes éditeurs et met en forme des bases de connaissance qui pourraient être reversées dans Bacon, après nettoyage et enrichissement par le hub. Comme cela apparaît sur le schéma ci-contre, ezParse peut progressivement passer d'une position de producteur à celle de consommateur/contributeur en s'abreuvant à Bacon, mais aussi à GoKB (*Global Open Knowledgebase*), rejoignant ainsi une initiative internationale très prometteuse.



phot. Stéphane Gully

➔ L'atelier ezParse lors du BibCamp ADBU en juin 2014.

Successeurs d'un projet interne à l'Inist-CNRS, le logiciel ezParse et l'environnement AnalogIST sont ouverts à la communauté, que ce soit pour une simple installation ou une collaboration plus avancée, depuis maintenant deux ans. ezParse est utilisé en production dans plusieurs établissements dont l'université de Lorraine est le pilote. L'équipe du projet se charge d'un soutien technique très réactif et se fixe l'objectif de fédérer une communauté d'utilisateurs et de contributeurs pour que le logiciel puisse poursuivre sa route à l'issue de la phase 3, fin 2015.

THOMAS PORQUET
Chargé de mission, Couperin
thomas.porquet@couperin.org
et L'ÉQUIPE EZPARSE
ezparse-team@googlegroups.com

[9] cf. note 5.

[10] <https://groupes.renater.fr/sympa/info/ezparse-contribute>

[11] <https://github.com/ezparse-project>

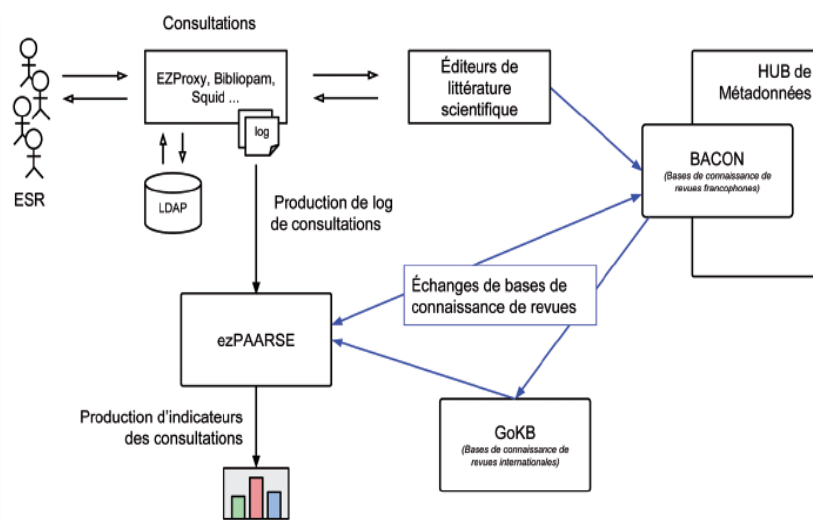
[12] Et cela a été récemment le cas avec une contribution inattendue des bibliothèques de l'université du Minnesota :

<https://github.com/UMNLibraries>

[13] <https://twitter.com/ezparse>

[14] <https://plus.google.com/+ezPARSE>

[15] www.facebook.com/Ezparse



➔ Articulations entre ezParse, Bacon et le hub de métadonnées de l'Abes.