

Pour procéder à l'analyse d'une plate-forme éditeur et des liens (URLs) grâce au logiciel libre Ezpaarse, il vous faudra utiliser différents outils comme :


- [Trello](#) (pour le suivi des plates-formes),
- la plate-forme [Analogist](#) qui héberge Ezpaarse

Le principe d'une analyse de plate-forme - et de ses urls - est de récupérer le plus d'informations documentaires possibles pouvant décrire une url d'accès à une plate-forme, à ses ressources (revues, books) et ses "types de documents" (articles pdf, html, abstract, sommaire, protocol...).

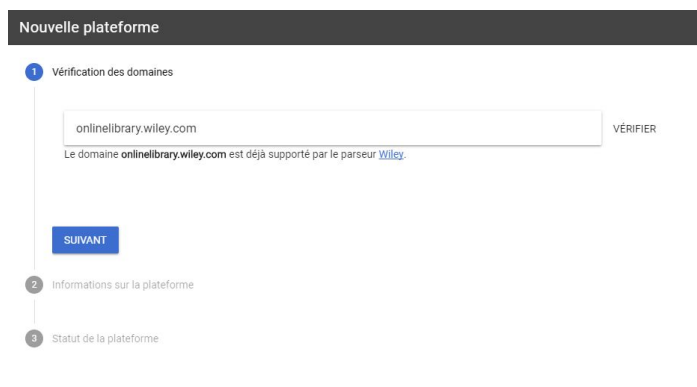
L'objectif d'une analyse est ensuite de permettre l'écriture de parseurs qui reconnaîtront les accès par les lecteurs à ces plates-formes, ressources et documents analysés.

- 1 Analogist

La première opération à faire quand on analyse une nouvelle plate-forme est de vérifier qu'elle n'est pas déjà présente dans [analogIST](#) (faire une recherche du nom de la plate-forme dans la barre de recherche).

Si celle-ci n'existe pas, vous pouvez créer la plate-forme en cliquant sur .

Indiquez le domaine (ex: [onlinelibrary.wiley.com](#)) pour vérifier s'il est déjà pris en charge par un parseur.



Nouvelle plateforme

1 Vérification des domaines

onlinelibrary.wiley.com VÉRIFIER

Le domaine [onlinelibrary.wiley.com](#) est déjà supporté par le parseur [Wiley](#).

SUIVANT

2 Informations sur la plateforme

3 Statut de la plateforme

Si le domaine n'existe pas, renseignez les informations de la plateforme (nom, abréviation, page d'accueil)

Nouvelle plateforme

1 - Vérification des domaines

2 - Informations sur la plateforme

Nom* nouvelle plateforme

Abréviation* nv

Page d'accueil http://nouvelleplateforme.com

SUIVANT PRÉCÉDENT

3 - Statut de la plateforme

Choisissez le statut de la plateforme (pour une création, on retient le statut 1 “plateformes à faire ou à réviser”) dans le menu déroulant :

Nouvelle plateforme

1 - Vérification des domaines

2 - Informations sur la plateforme

3 - Statut de la plateforme

1 - Plateformes à faire ou à réviser

2 - URLs en cours d'analyse (penser à l'annoncer dans ezpaarse-contribute)

3 - Analyse des URLs terminée

4 - Parseur d'URLs en dev

5 - Parseur fait, à tester avec de vraies logs

6 - PKB / scraper en cours d'analyse

Terminez la création en cliquant sur “créer”.

1 - Plateformes à faire ou à réviser

CRÉER PRÉCÉDENT

Vous pouvez aussi retourner sur le tableau de suivi Trello, votre identifiant sera alors automatiquement renseigné sur la carte concernant votre plateforme.

<https://trello.com/b/wEaLnz8d/ezpaarse-analogist-suivi-des-plateformes>

Quand la fiche **plate-forme** est créée,

commencer l'**analyse des url** en cliquant sur le bouton

ANALYSES (17) →

puis sur “+”



Une nouvelle fenêtre s'ouvre et vous pouvez commencer l'analyse

Wiley [wiley]

Titre

URL

Type Format UnitID

Remarques

Champs reconnus

Éléments de la route

Paramètres de la query

Complétez le formulaire en indiquant :

- le titre de votre analyse (ex: consultation d'un article en PDF)
- l'url de consultation
- le type document (ARTICLE, TOC, ABS, BOOK,...)
- le format (HTML, PDF,...)
- le UnitID (l'identifiant de l'article)
- Une remarque (si vous le souhaitez)

Cliquez sur

Champs reconnus



pour ajouter un champs reconnu dans l'url (doi, pii, print_identifier, online_identifier...)

← RETOUR

Wiley [wiley]

Mise à jour il y a un mois par Pascale Pauplin

Article PDF
<http://onlinelibrary.wiley.com/doi/10.1002/cbf.3219/epdf>

ARTICLE PDF cbf.3219

Champs reconnus

Type	Valeur
doi	10.1002/cbf.3219

Éléments de la route

Valeur	Remarques
doi	
10.1002	
cbf.3219	
epdf	

L'enregistrement du formulaire se fait automatiquement.

❖ *Consignes/recommandations :*

Bien décrire tous les éléments identifiables dans une url.

Renseigner absolument l'identifiant de ressources (title_id) et l'identifiant de l'article ou d'un chapitre (UnitID) .

l'identifiant title_id de la ressource (revue, ebook) est important car il va permettre par la suite d'être associé à une base de connaissance (Listes Kbart) et de créer une PKB pour identifier précisément chaque ressource (publication_title, ISSN, éditeur,...) dans les EC

Ne pas hésiter à faire des remarques qui pourront être utiles à l'écriture par la suite des parseurs.

Analyser aussi bien les numéros récents que les archives, les suppléments... les écritures d'urls pouvant différer au sein d'une même revue. Cette remarque vaut encore plus au sein d'une plate-forme donnant accès à n revues. Vous pouvez ainsi, pour une même plate-forme, avoir une multitude de descriptions d'url pour un même type d'articles en format pdf par exemple.

Les abstracts, sommaires...sont souvent dans un format autres que pdf ou html. Le plus souvent , il faut le renseigner comme étant au format misc.

Il est difficile d'être exhaustif dès le premier passage dans l'analyse d'une plate-forme ; il sera alors nécessaire de revenir et mettre à jour des analyses de plates-formes après les phases d'écriture/tests de parseurs, voire après la phase finale (list 10 / plate-forme mise à jour).

- Dans une analyse de plate-forme et de ses urls il est possible de trouver et décrire une multitude de documents (à prendre au sens générique du terme) : revue, sommaire

(TOC), abstract, article PDF, article HTML, BOOK, chapitre de book, protocole etc. Liste exhaustive disponible sous :

<http://ezparse.readthedocs.io/en/master/essential/ec-attributes.html#resources-types-rt>
[ype](#)

- 2 Trello

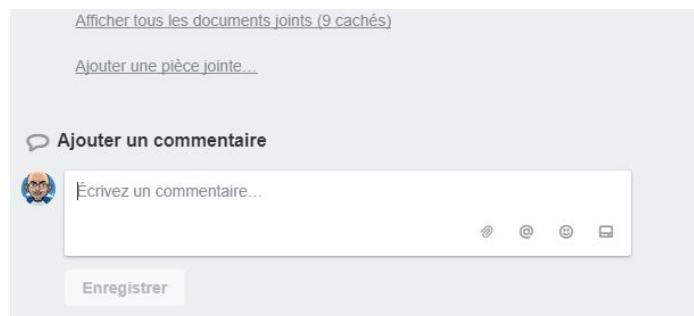
Trello : <https://trello.com/b/wEaLnz8d/ezparse-analogist-suivi-des-plateformes>

Outil collaboratif pour le suivi en 10 étapes des analyses de plates-formes depuis l'étape de création/révision de plates-formes à l'étape finale où la plate-forme est à jour au niveau d'Ezparse en passant par les étapes d'analyse d'url, création de parseurs, tests de parseurs ...

Quand une tâche est terminée, ne pas oublier de glisser cette étiquette/card vers la liste suivante.

Consignes/recommandations :

Possibilité d'ajouter des commentaires sur des "cards" (même sur celles que vous n'avez pas créées vous-même). A ce titre **ne pas oublier de cliquer sur "Comment"** pour enregistrer son commentaire (peut paraître un détail mais ça m'est arrivé plusieurs fois de ne pas valider mon commentaire).



Vous pouvez également poser des questions via le tableau Trello pour vous faire aider.

- 3 Suite des opérations (parseurs, pkb, scrapeur)

Les informations recueillies lors de l'[étape précédente](#) d'analyse d'urls sont le point de départ du travail sur Ezparse et sont déterminantes pour la suite des opérations qui ont un caractère plus informatique que documentaire et nécessitent des compétences informatiques avérées.

- Les informations d'analyse permettent de passer à l'étape d'écriture du **parseur** (écriture entre autres d'expressions régulières à partir des informations données dans l'analyse documentaire

voir la rubrique "L'écriture du parseur (codeur)" sous

<http://ezparse.readthedocs.io/en/master/>

En résumé, ça ressemble à ça :

```
if ((match = /S{4}([0-9]{3}[0-9Xx])([0-9A-Za-z]*)/.exec(param._pii)) !==
null) { // example :
http://pdn.sciencedirect.com.gate1.inist.fr/science?_ob=MiamiImageURL&_cid=2821
79&_user=4046392 //
&_pii=S221267161200100X&_check=y&_origin=browseVolIssue&_zone=rslt_list_item&_c
overDate=2012-12-31 //
&wchp=dGLbVlB-zSkWz&md5=79a307d3c9bdbea6d6a6092d73c25545&pid=1-s2.0-S2212671612
00100X-main.pdf result.unitid = 'S' + match[1] + match[2] + match[3];
result.title_id = match[1] + match[2]; result.print_identifiier = match[1] + '-'
+ match[2]; result.rtype = 'ARTICLE'; result.mime = 'PDF';
```

Ce parseur sera alors à tester avec de vrais logs

<http://ezparse.readthedocs.io/en/master/>

- Après cette étape parseur (et tests de parseur) , il est possible de constituer un fichier PKB via une base de connaissance éditeur (données éditeur tels que publication title, title ID...).

voir comment créer une PKB sous <http://ezparse.readthedocs.io/en/master/>

- ainsi qu'à la création d'un scraper liée à celle de la PKB car une PKB peut être automatiquement générée par un scraper.

Tous les parseurs, pkbs, scrapeurs déjà créés sont visibles sur Github* à l'adresse

<https://github.com/ezparse-project/ezparse-platforms>

En conclusion, l'étape d'analyse de la plate-forme et de ses urls est l'étape déterminante et cruciale dans le dispositif Ezparse

*Github (source Wikipédia <http://fr.wikipedia.org/wiki/GitHub>) = **GitHub** (exploité sous le nom de *GitHub, Inc.*) est un service web d'hébergement et de gestion de développement de logiciels, utilisant le logiciel de gestion de versions Git. Ce site est développé en Ruby on Rails et Erlang par Chris Wanstrath, PJ Hyett et Tom Preston-Werner. GitHub propose des comptes professionnels payants, ainsi que des comptes gratuits pour les projets de logiciels libres.