

Date: Vend. 24 Mai 2019 - 10h45

Présents: Soizick L., Magali N., Grégory M.

État actuel

- 3 types d'articles:
 - article normal + letters (pour le MNRAS et l'APJ)
 - Circulaires (e.g. GCN, ATel, AstBu, IBVS)
(*sorte de nouvelle/info à partager rapidement ou article très court*)
 - Les journaux entrés manuellement ne sont entrés dans Simbad que si ils concernent un objet astro.
 - erratum (*correction(s) d'un article déjà publié*)
 - jamais listé dans Simbad (même si bibcode automatiquement créé par les éditeurs)
 - jamais traité par Djin
 - seule une note est ajoutée à la ref dans Simbad (afin de notifier les documentalistes d'éventuelles infos erronées dans l'article)
- Récupération des articles:
 - depuis FTP (articles complets): A&A, MNRAS
 - depuis stacks IOP (*mdp nécessaire ; cf Gilles et Soizick*):
IOP (i.e. AJ, ApJ, ApJS, PASP)
 - depuis mails: GCN, AstBu, IBVS
 - depuis site web: ATel (*article HTML...n'ayant jamais changé*)
 - ...
- Liste des journaux donnés par Magali:
 - liste des journaux actuellement analysés par les DJINistes, avec le lien vers le site web, le nombre de volumes par an et le nombre d'issues par volume et par mois
 - <https://docs.google.com/document/d/1tyy-1A1gyxYOzDouGe3Jp3N81B2OOXqL5cqrT6kqWbY/edit?usp=sharing>
- Pour l'instant, les articles ne sont traités que si leur volume/fascicule est complet
 - *pourquoi ?*
 - pour mieux identifier les articles traités ou pas
 - afin de ne pas en oublier
 - traitement plus facile dans DJIN avec la fonction "Bibcode suivant" (i.e. incrément du bibcode pour traiter l'article suivant dans le volume/fascicule)
 - Soizick et Magali ne sont pas contre traiter les articles au fur et à mesure qu'on les récupère, et ce, même si le volume/fascicule n'est pas complet, **MAIS...**
 - il faut en rediscuter ; cela changera l'organisation des DJINistes
 - il faut être capable de ne pas oublier d'articles (souvent, les articles d'un volume/fascicule sont disponibles dans un ordre imprévisible)

- => *il faut un moyen de suivre quels articles sont déjà traités et ceux qui ne le sont pas encore et qu'il faut faire (tout en sachant, qu'il arrive aux DJINistes de revenir un peu plus tard sur un article déjà traité)*
 - risque qu'un article soit disponible mais pas à jour (des corrections sont encore susceptibles d'arriver)
 - => **bref:** *pour l'instant, on fait en sorte de garder le traitement par volume/fascicule ; à rediscuter plus tard si organisation par article est faisable*
- Traitement actuel d'un article:
 1. script de Thomas pour en extraire le TOC
 - extraction des titres, auteurs, affiliation, mots-clefs, abstract, DOI, ISSN, nom du fichier original de l'article, %O (objets)...
 - titre (casse ; prend aussi en compte acronymes présents dans l'abstract) et auteurs (nom de famille en maj. et abréviations du/des prénoms) retouchés automatiquement par le script
 - le Bibcode est créé automatiquement par ce script (il n'est retouché manuellement que si erreur dans le nom du premier auteur ; cf ci-dessous)
 - => **bref:** *revoir ce qui peut être récupéré du code de Thomas*
 2. les TOCs de tous les articles d'un volume/fascicule sont agrégés au sein d'un seul fichier Parfile
 3. vérification manuelle (*cf ci-dessous*)
 4. conversion en script de Maj de Simbad
- En fait, le Parfile est utilisé comme format pivot (d'une liste de TOCs) entre la liste des articles et le script de Maj de Simbad
 - le but est de pouvoir vérifier les articles à ajouter
 - et effectuer quelques modifs si nécessaire (*cf ci-dessous*)
- Vérification humaine, et si nécessaire, quelques traitements manuels:
 - titre
 - noms auteurs (i.e. correctement abrégés....surtout les noms espagnols)
- Il arrive souvent que MNRAS corrige a posteriori (mais pas en mode Erratum) les articles disponibles (mais généralement ils préviennent Soizick et les corrections ne sont souvent pas si importantes pour DJIN)

Semi-automatisation

- Automatisation possible:
 1. récupération des articles
 2. conversion Errata en notes (à réfléchir car quelques problèmes soulevés)
 3. conversion en XCDS
 4. extraction des TOCs depuis XCDS
 5. agglomération des TOCs de tous les articles d'un volume/fascicule en 1 seul Parfile (pour ensuite Maj plus simple dans Simbad)
- Ce qui ne peut pas être automatisé pour l'instant:
 - vérification des titres et auteurs des articles dans les TOCs agrégés
 - une expertise humaine est nécessaire
 - correction des Parfiles manuellement au cas par cas

- => *il devrait toutefois être possible d'améliorer les traitements sur les titres et auteurs faits par le script de Thomas => à voir*
 - Principale difficulté:
 - détection de complétude d'un volume/fascicule
 - Soizick et Magali vont sur le site web du journal visé pour voir si le volume/fascicule cible est complet ou pas
 - pour A&A, il est possible que Patricia soit (automatiquement ?) notifiée
 - => *à confirmer avec Patricia*
 - pour les IOP, au sujet des notifications + récupération:
 - => *à voir avec Gilles*
 - il a apparemment déjà qq scripts pour la récupération automatique des articles complets
 - Prévoir un outil de supervision (Web) de la complétude des volumes/fascicules pour pouvoir donner le "Go!" aux DJINistes
-

Cas de quelques petits journaux

- GCN:
 - format text/plain
 - parsing déjà pris en compte par le code de Thomas
 - format extrêmement simple et articles très courts
 - => *la conversion en XCDS devrait être très simple*
 - ATel:
 - format HTML (*format/structure resté inchangé depuis plusieurs années*)
 - articles très courts aussi
 - => *la conversion en XCDS devrait être relativement simple*
-

ViZieR VS Simbad : génération des Bibcodes

- Il peut arriver à Patricia de traiter les tables d'un article qui n'a pas encore été traité par Soizick/Magali => pas de Bibcode
 - elle crée alors un Bibcode dans ViZieR avant qu'il n'existe dans Simbad
 - **Suggestion (GM):** *possible récupération/vérification du Bibcode en utilisant le DOI comme clef commune ; si existant dans ViZieR, on l'utilise tel quel, sinon on le crée*
 - => *à voir avec Gilles*