

Chantier nettoyage du CFP

Journée des adhérents au PCMath
Mardi 8 mars 2022, IHP, Paris

Dominique Barrère, Francesca Leinardi
Réseau National des Bibliothèques de Mathématiques

CFP Catalogue fusionné des périodiques

Débute en 1995 : collaboration Mathdoc et RNBM

Mathdoc assure le développement et la maintenance technique

Alimenté :

- ~ par imports des catalogues des bibliothèques (périmètre disciplinaire dépasse les mathématiques)
- ~ puis en fonction des besoins pour les projets dont il est le support

Diagnostic posé en 2014 par Mathdoc et les GT RNBM travaillant à partir du CFP sur le contenu du CFP :

- ~ clarifier et désherber !

Pourquoi un chantier de nettoyage ?

Nécessité de s'appuyer sur un corpus clairement défini, centré sur les mathématiques et contrôlé pour pouvoir assurer les services identifiés :

- la sélection du corpus du PCMath ;
- la recherche et la localisation des périodiques ;
- la gestion des périodiques électroniques acquis par le biais d'une licence nationale ;
- la gestion des collections imprimées et électroniques dans les bibliothèques du réseau ;
- l'identification des ressources à numériser.

Premiers travaux

- 2014-2015 : état des lieux, Sophie Min-Picault (Mathdoc)
 - ~ 19.380 notices :
 - 5.206 en maths – 3.694 avec au moins une localisation – 1.745 unicas
 - ~ Groupe de travail définit des règles sur :
 - Les bibliothèques participantes
 - Le besoin de recentrer le corpus
 - Les bases d'appui
 - Le besoin de connecter le sudoc avec le CFP
 - La nécessité de nettoyer la base et définir les critères de contenu du CFP
- 2015-2020 : ré-écriture informatique par Mathdoc
 - ~ Ajout de nouvelles fonctionnalités pour la gestion du PCMath et des ressources électroniques

Chantier à partir de 2020

- mai 2020, le groupe de travail est ré-activé , accompagné par 2 mathématiciens.
- 1^{ère} étape préalable au nettoyage :
 - identifier et formaliser les critères de tri ;
 - définir des critères de contenu du CFP.
- Mise en œuvre
 - Tri automatique (sans localisation) : 12 104 notices retirées
 - Tri manuel (4 138 titres étudiés) : 1 984 notices retirées, 531 avec localisation en instance

a – Critères du tri automatique

- Titres sous lesquels seuls sont localisés :
 - La BNF, Inria, l'INIST
- Sont à garder dans le CFP pour cette étape :
 - le corpus du PCMath sélectionné à cette date, soit 1005 titres ;
 - les titres qui ont au moins une bibliothèque localisée ;
 - les titres du DOAJ Math ;
 - les titres de la base zbMath ;
 - les titres liés (historique des titres à garder).
- Bilan à l'issue de cette première phase :
 - 7 251 titres dans le corpus du CFP
 - 9 017 titres hors du corpus CFP

b - Critères de contenu du CFP

Font partie du corpus du CFP pour cette phase

- titres localisés dans une bibliothèque de mathématiques ou ayant une composante mathématique ;
- titres du corpus PCMath ;
- titres liés par une relation historique à des titres présents dans le corpus dans la mesure où ils sont centrés sur les mathématiques ;
- titres ayant une période d'accès électronique, dont les titres électroniques DOAJ ;
- titres indexés dans la base zbMath et/ou dans la base Mathscinet ;
- titres électroniques sans localisation papier mais avec une localisation « électronique » = titre avec une licence active dans au moins une bibliothèque, soit un titre en libre accès sélectionné à partir de l'AMS Digital Mathematics Registry 1 et le portail Mathnet.ru ;
- publications des IREMs et de l'APMEP ;
- titres mathématiques localisés dans le Sudoc mais absents du CFP.

Sont à exclure du CFP

- revues généralistes en science, même si elles contiennent des articles en mathématiques (exemples Nature, Science, etc)

Résultats chiffrés

	Total notices CFP	Modifications apportées	Répartition notices CFP			
			Hors corpus CFP		Corpus CFP	
			Total hors corpus CFP	Dont « En sursis », hors corpus avec au moins 1 localisation	Total corpus CFP	Dont « Corpus PCMath »
2015 - Etat des lieux	19 380					
exclusion titres BNF/Inria/INIST sans localisation		-3 112				
Etat eu 15/06/2020	16 268		9 017		7 251	1 005
hors corpus : tri automatique des titres hors corpus non localisés + réajustement titres MathNetRu						
Etat au 10/11/2020	16 268		8 992		7 276	
purge d'automne 2020		-8 992				
Etat au 17/11/2020	7 276				7 276	1 004
déc 2020-fév2022 : Analyse « manuelle » des titres restants dans le corpus + ajustements		43	2 515	531		
Résultat de la purge d'hiver 2021-2022		-1 984				
Etat au 25//02/2022	5 335			531	4 804	1 004
retrait du CFP (BNF Inria INIST)	3 112					
Retrait sur tri automatique	8 992					
Retrait sur sélection manuelle	1 971					
Nbre notices supprimées	14 075					

Dernière étape : à vous !

Porte sur les titres à retirer du CFP avec des bibliothèques localisées.

- chaque bibliothèque peut voir dans son tableau de bord (après authentification) les titres qu'elle détient parmi une liste de 531 titres concernés ;
- pour les titres qu'elle souhaite conserver dans le CFP, chaque bibliothèque envoie au groupe de travail un avis argumenté ;
- **après le 30 juin 2022**, les titres non signalés seront retirés du CFP.

Et pour la suite...

- Conclusions du GT rendues au Comité stratégique du RNBM le 17/02/2022 ;
- Souhait principal : mise en place d'un « comité éditorial » du CFP, composé des représentants des différentes parties prenantes du CFP et dont le rôle serait de garantir la qualité et la cohérence de la base ;
 - définir les critères de contenu du CFP et les règles de mise à jour ;
 - procéder à l'alimentation et au maintien du corpus du CFP ;
 - accompagner les groupes dans leurs projets basés sur le CFP.