

# Le fonctionnement de CoCo-SAlma2

## l'outil de contrôle des correspondances des cotes Sudoc-Alma

*Réécrire le fonctionnement est un exercice plus compliqué que prévu mais qui, je pense, peut apporter des précisions utiles à l'utilisateur.*

### Préparation du fichier

#### Remise à zéro du fichier

La remise à zéro des données permet de supprimer toutes les entrées dans la feuille "Résultats" puis celles de la feuille "Données", sauf les formules en B2 et J2, ainsi que rendre aux cellules G2 et J2 leur valeur par défaut.

En effectuant cette remise à zéro, on évite la présence de résidus d'un autre traitement qui ne seraient pas réécrits par la suite (si vous traitez une liste de 100 PPN après une de 500, 400 PPN sans rapport se glisserait dans le processus).

#### Le nom du fonds et le dossier créé

La création d'un dossier avec le nom du fonds permet de rassembler tous les documents relatifs à ce traitement au même endroit, et permet de raccourcir le nom des fichiers puisqu'ils sont compris dans le dossier avec le nom du fonds. (Et cela me simplifie la vie dans le script.)

Ainsi, enregistrer dans un premier temps le fichier dans un emplacement spécifique permet de générer le dossier à l'emplacement voulu sans devoir le déplacer par la suite.

#### Le choix de la bibliothèque

On entre le RCR de la bibliothèque et pas son nom car le nom de celle-ci doit avoir une correspondance parfaite avec ce qui est renseigné dans Alma, plus difficile donc de le connaître par cœur que le RCR. Toutefois, avec l'aide des listes de données validées dans Excel, l'option de choisir le nom dans une liste et auto-compléter le RCR est envisageable.

### Import et préparation de la liste dans CoCo-SAlma

Lorsque l'on colle la liste de PPN originale, une certaine attention doit être portée à la forme des PPN car si la formule de la colonne "035 FIELD" prend en compte différents type d'écriture, cette correction n'est pas poussée, mais elle est nécessaire pour le bon déroulement de l'import dans Alma. La correction en question est assez basique : si le PPN original contient "(PPN)", alors il est repris tel quel, sinon, "(PPN)" est écrit devant les neuf derniers caractères du PPN original.

Puisqu'un tri est effectué juste avant de générer la liste, il importe peu que des lignes soient laissées vides entre les PPN.

## L'import depuis Alma

En ce qui concerne l'import d'une requête depuis Alma, CoCo-SAlma ouvre et lit directement la colonne "Numéro de notice" dans le fichier nommé *export\_alma.xlsx* et colle ces informations dans la colonne appropriée du fichier principal. Elle ferme ensuite l'export d'Alma qui a accompli son but.

CoCo-SAlma détecte après, pour chaque PPN, l'emplacement de la chaîne de caractères "(PPN)" dans la cellule du numéro de notice importé, puis colle les 14 caractères suivant cet emplacement (5 pour "(PPN)" et les 9 du PPN). Cette analyse vise à supprimer les identifiants comme les NNT qui pourraient venir brouiller son analyse.

## Génération de la liste et du dossier

Une fois la liste de PPN originale importée, il faut générer la liste des PPN pour l'export. Cette action créera dans un premier temps le dossier, avec le nom du fonds choisi (certains caractères inutilisables dans des noms de fichiers Windows sont remplacés par des underscores), puis sauvegardera le fichier dans le dossier.

Ensuite, CoCo-SAlma tri la liste originale pour laisser de coter les cellules vides, détermine le nombre de PPN, puis colle la formule générant la bonne forme pour le PPN à côté de chacun des PPN. Elle enregistre après une copie de la feuille actuelle ("Données") dans un nouveau fichier *import\_alma.xlsx*.

## L'analyse

Avant toute chose, CoCo-SAlma va garder en mémoire l'heure de début de l'analyse. Elle va ensuite supprimer l'intégralité des données de la page de résultats puis écrire et formater les en-têtes.

## Côté Sudoc

Pour importer les données du Sudoc, CoCo-SAlma s'appuie sur le [Webservice UNIMARC/MARCXML](#) du Sudoc. Elle compte dans un premier lieu le nombre de cellules non-vides dans la colonne "035 FIELD", puis commence une boucle.

Pendant la boucle, CoCo-SAlma lit les informations de la page "https://www.sudoc.fr/" + le PPN + ".xml" (la notice bibliographique d'un document fourni au format MarcXML). Une fois la page chargée, CoCo-SAlma va récupérer la valeur du tag 001, le PPN. Si cette opération échoue, parce que le PPN ne correspond pas à une notice bibliographique, parce qu'une page d'erreur s'est affichée, etc., CoCo-SAlma désigne ce PPN comme incorrect et passe à la valeur suivante. Si l'opération est un succès, pour chaque 930, elle regarde si la valeur du \$b correspond à notre RCR.

Si c'est le cas, elle regarde d'abord si elle a déjà une cote en mémoire : si ce n'est pas le cas, elle garde alors en mémoire la valeur du \$a de cette même 930 ; si c'est le cas, elle

conserve les informations qu'elle a déjà, ajoute le séparateur “;\_”, et garde la nouvelle cote à la suite. Si au terme de sa boucle CoCo-SAlma ne détecte aucune correspondance avec le RCR, elle garde en mémoire qu'elle n'a pas trouvé de cote.

Elle inscrit dans la feuille “Résultats” le PPN stocké et la valeur stockée en tant que cote.

Une fois tous les PPN traités, elle supprime les doublons présents dans la feuille “Résultats” en se basant sur la colonne des PPN.

### Côté Alma

CoCo-SAlma va ouvrir l'*export\_alma.xlsx* et compter le nombre d'entrées dans la colonne “Numéro de notice”.

Ensuite, pour chaque ligne, elle récupère le PPN dans la colonne du PPN en veillant à retirer les autres numéros de système, puis elle divise les disponibilités (colonne “Disponibilités” en fonction du nom de la bibliothèque, en ignorant de force la première partie, qui pourrait générer des erreurs dans son analyse.

CoCo-SAlma cherche après le premier point-virgule qu'elle va rencontrer (qui se situe forcément devant la bonne cote, sauf si la localisation du document en contient un), puis cherche :

- soit si le prochain point-virgule se trouve avant la prochaine “ (” ;
- soit si elle ne trouve pas de prochain point-virgule.

Si l'une de ces deux conditions est remplie, elle récupère comme cote tout ce qui se trouve entre ce point de virgule (moins un espace) et le premier double espace qu'elle va rencontrer. Si aucune de ces deux conditions n'est remplie, CoCo-SAlma récupère tout ce qui se trouve entre les deux points-virgules comme cote. Cette opération relativement compliqué sur le papier sert à ne conserver que la cote, sans l'état des collections ou le numéro d'entrée. Comme pour le Sudoc, elle sépare les cotes si elle en détecte plusieurs, ou indique qu'elle n'en trouve aucune.

### Tri et alignement des données

CoCo-SAlma trie ensuite les colonnes de chaque logiciel par ordre via leur PPN. Elle compare après laquelle des deux listes est la plus grande pour générer la limite de la boucle qui va suivre.

Elle compare ensuite pour chaque ligne les neuf derniers caractères du PPN de chaque logiciel. Si les PPN ne correspondent pas, CoCo-SAlma détermine lequel des deux est le plus grand, et ajoute une ligne vide dans la colonne de celui-ci en poussant le reste de la liste vers le bas, conservant un compteur du nombre de fois où cette action est effectuée et diminuant le compteur de la boucle originale. Cette opération sert à aligner chaque PPN. Une fois la liste complète effectuée, elle additionne la limite de la boucle au nombre de fois où elle a poussé pour transmettre cette nouvelle valeur au reste du script.

Cette partie du script a toujours été la plus sensible aux problèmes inattendus, il ne serait pas étonnant qu'elle soit fautive de nouveaux problèmes.

### Comparaison des données

Pour chaque ligne du tableau (=chaque PPN), CoCo-SAlma compare les cotes qu'elle a extraites et génère une réponse booléenne (Oui / NON). Elle indique en complément s'il y a 2 cotes dans le Sudoc ou Alma (il peut y avoir correspondance avec 2 cotes dans chaque, auquel cas elle répond "Oui" et indique qu'il y a deux cotes pour chaque logiciel), ou s'il manque une cote dans un logiciel. Enfin, si la colonne du PPN contient "PPN INCORRECT", elle réécrit sa réponse en "ERREUR PPN".

### Finalisation

Une fois la comparaison terminée, CoCo-Salam va appliquer quelques modifications de style et supprimer les lignes superflues (théoriquement). Elle enregistre ensuite le fichier et aura terminé son travail, indiquant la durée de l'opération en soustrayant à l'heure actuelle l'heure qu'elle a conservé en début de script.