

Les moulinantes

Outils pour l'exemplarisation automatique

Description

Formation à l'utilisation des "moulinantes", une suite de petits utilitaires pour Windows destinés à traiter des fichiers texte en vue d'une exemplarisation automatique.

Public

Coordinateur Sudoc chargés de l'intégration d'un fonds et désirant effectuer une exemplarisation automatique.

Intervenants

Formateur et développeur : Michel Lemer

Expert : Sophie Goron

Modérateur : Agnès Manneheut

Plan

- Qu'est-ce qu'une exemplarisation automatique ?
- Utiliser les moulinantes.
- Encore des moulinantes après l'exemplarisation !

PARTIE 1

Qu'est-ce qu'une exemplarisation automatique ?

Qu'est-ce qu'une exemplarisation automatique ?

Lorsqu'une bibliothèque absorbe le fonds d'une autre bibliothèque, on peut envisager de procéder à des tâches automatisées lorsque le contenu du fonds peut être exporté sous forme numérique (fichier tableur). Chacune des lignes exportées correspondra alors à un exemplaire du fonds et contiendra à la fois des données d'exemplaire (code-barres, cote) et des données destinées à repérer le document (au mieux, l'ISBN, au pire, le trio date-auteur-titre).

Pour l'ABES, exemplariser automatiquement consiste à tenter d'identifier l'ouvrage grâce aux données de repérage puis de lui attacher les données d'exemplaires.

Et pour le bibliothécaire, exemplariser automatiquement consistera à effectuer (selon un protocole précis) plusieurs traitements sur ce fichier tableur puis de l'envoyer à l'ABES.

Tous ces traitements pourront être effectués par des petites applications pour Windows rassemblées dans une suite que nous avons nommée *les Moulinantes*.

Au final :

- les exemplaires seront intégrés au SUDOC,
- un lot de notices MARC sera produit par l'ABES.



Ce lot de notices pourra alors – après traitement d'une Moulinante – être importé dans le SIGB.

Conception d'un fichier d'exemplaires

Un fichier d'exemplaires est un fichier texte dont chaque ligne est un ensemble de données séparées par une tabulation (format CSV).

Ces données sont de deux ordres :

- données pour repérer le document (soit ISBN, soit date, auteur et titre)
- données d'exemplaire (cote et code-barres)

Exemples de lignes

ISBN + cote et code-barres

284239030X↔MC23-35↔589770931

286628187X↔M 6782↔589700246

Date, auteur et titre + cote et code-barres.

1994↔Develay, Michel↔Peut-on former les enseignants ?↔M 6285↔589730207

DL 1979↔Poë, Edgar↔Le scarabée d'or↔MR 298↔589702688

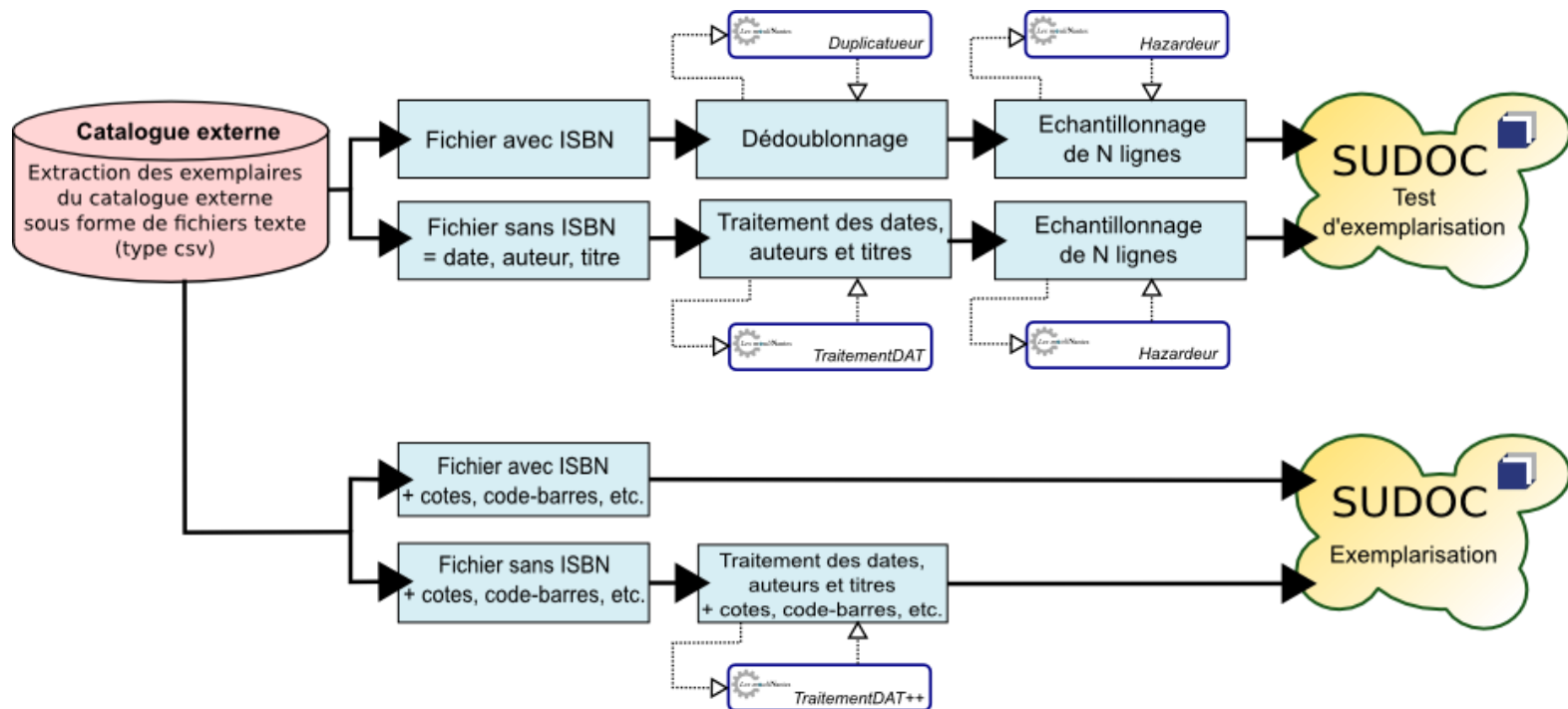
NB Le symbole ↔ représente le caractère *Tabulation*. C'est lui qui fera office de séparateur de champ pendant les phases de traitement. Dans les documents finaux, c'est le *point-virgule* qui sera utilisé.

On aura compris que les fichiers d'exemplaires seront donc à subdiviser en deux types de fichiers en fonction d'un critère simple :

- soit les données d'exemplaire contiennent l'ISBN
- soit les données d'exemplaire **ne contiennent pas** l'ISBN

Page suivante, un schéma résume ces deux modes.

Schéma d'une exemplarisation automatique avec ISBN et sans ISBN



Avec ISBN

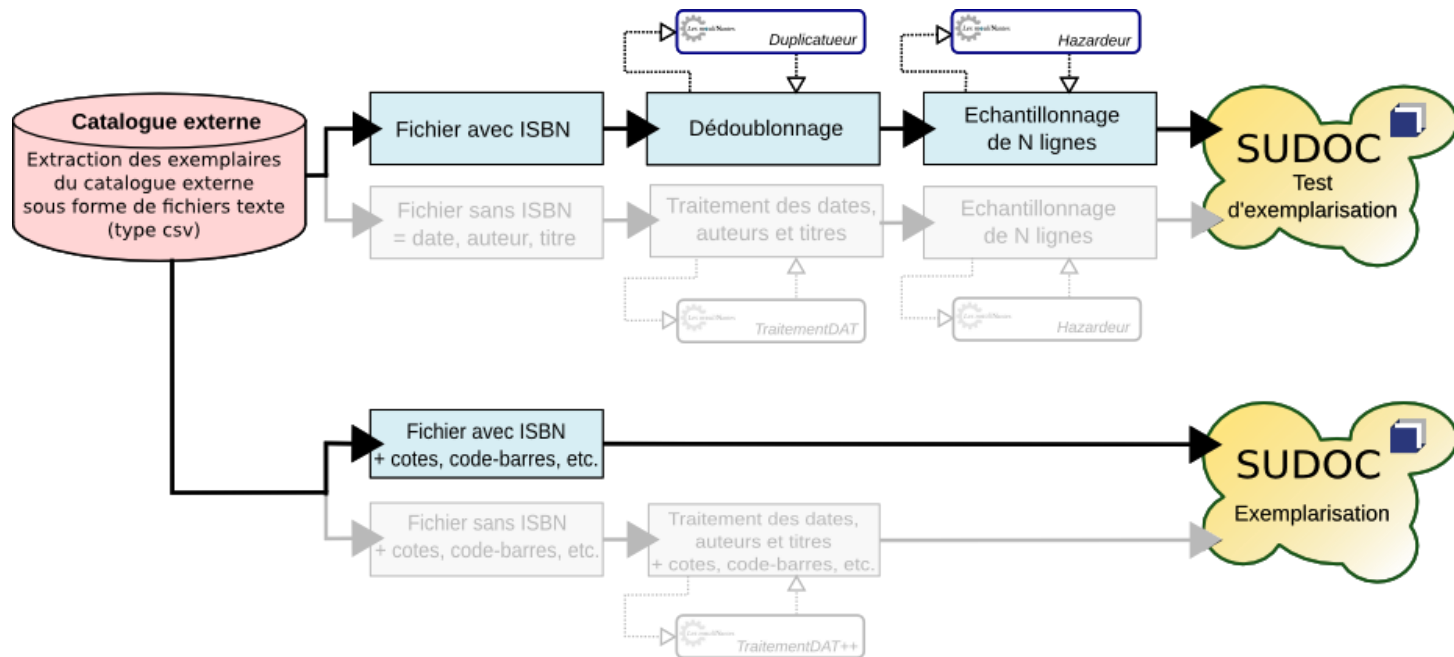
- Chaque ligne du fichier est un ensemble de données séparées par un point-virgule. C'est le fichier idéal !
- C'est ce type d'exemplarisation qui donne les meilleurs résultats.
- Les données habituelles sont l'ISBN, le code-barres et la cote.

Exemple :

```
123456789X;00800112234;681.3 PAR
2342345678;00800118866;681.3 ZAP
4564567890;00800115489;681.3 POU
689689774X;00800111421;681.3 QUI
```

- Avant envoi final au SUDOC, il faudra effectuer un test d'exemplarisation. Pour ce test :
 - on ne garde que le champ ISBN
 - puis les données sont dédoublonnées (avec la moulinante **Duplicatueur**)
 - puis on extrait 1000 lignes au hasard (avec la moulinante **Hazardeur**).
 - C'est ce fichier de 1000 ISBN qui sera envoyé pour test.

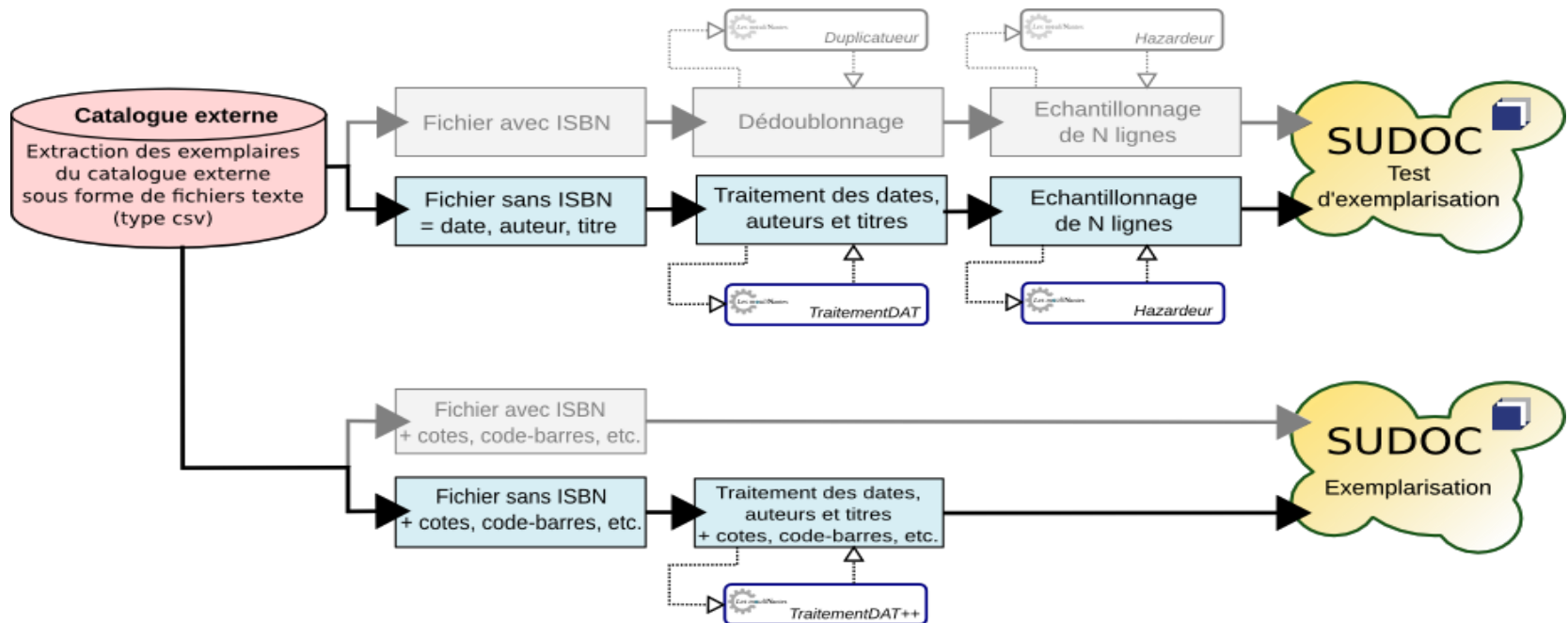
Schéma d'une exemplarisation automatique avec ISBN



Sans ISBN

- Le protocole *Sans ISBN* est sensiblement différent de celui *Avec ISBN* car il va falloir essayer d'identifier l'ouvrage dans le Sudoc au vu de données plus subjectives :
 - la date de publication
 - l'auteur
 - le titre
- Ces données devront être formatées conformément aux exigences de l'ABES selon des modalités précises parmi lesquelles :
 - traitement du titre (suppression des booléens, mots vides et mnémoniques d'index)
 - uniformisation de la date (ne garder que les 4 chiffres de l'année)
 - uniformisation du champ auteur (ne garder que le nom du 1^{er} auteur)
 - uniformisation de la casse et encodage du fichier
- Les deux moulinantes **TraitementDAT** et **TraitementDAT++** ont été développées pour effectuer tout ce formatage. Le traitement ne dure que quelques secondes.

Schéma d'une exemplarisation automatique sans ISBN (avec date auteur et titre)



Exemples de transformation effectuées par TraitementDAT

Cop. 1998;Hugo V.;La légende des siècles

devient

1998;HUGO;LEGENDE SIECLES

2001;DUBOIS F, DUPONT T.;Le truc et la chose

devient

2001;DUBOIS;TRUC CHOSE

2002;La Fontaine Jean de;Fables

devient

2002;LA FONTAINE;FABLES

DL 1999;DURAND JEAN;Les chercheurs d'or

devient

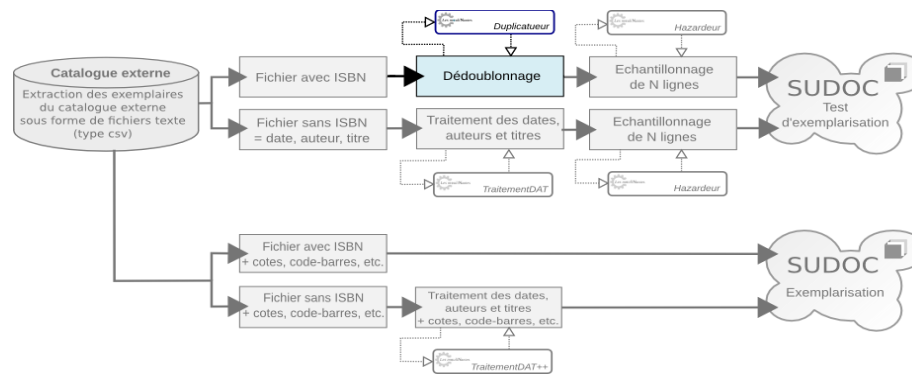
1999;DURAND;CHERCHEURS

Devinette : que deviendront ces titres ?

Pour for my bro !
Sur l'Oder sans le sou
Pay or Die, c'est selon.
Il mit notre nom à ceux du col.

PARTIE 2

Utiliser les moulinantes.



Etape 1-1 : ISBN - Dédoublonnage

Dédoublonner un fichier d'ISBN avec le programme Duplicatueur

- Exécutez le programme *Duplicatueur*
- Ouvrez le fichier d'exemple *isbn_5000.txt*
- Effectuez la fonction *Dédoublonner*.
- Enregistrez les 3833 lignes dédoublonnées dans le fichier résultat. Nommez-le *isbn_3883.txt*

Copie d'écran page suivante

Dédoublonnage avec le duplicatueur

Nom du fichier ouvert

Nombre de lignes après dédoublement

Nombre de lignes avant dédoublement

Duplicatueur - Le dédoubleur

Duplicatueur

A propos...

Ouvrir isbn_5000.txt

Dédoubler

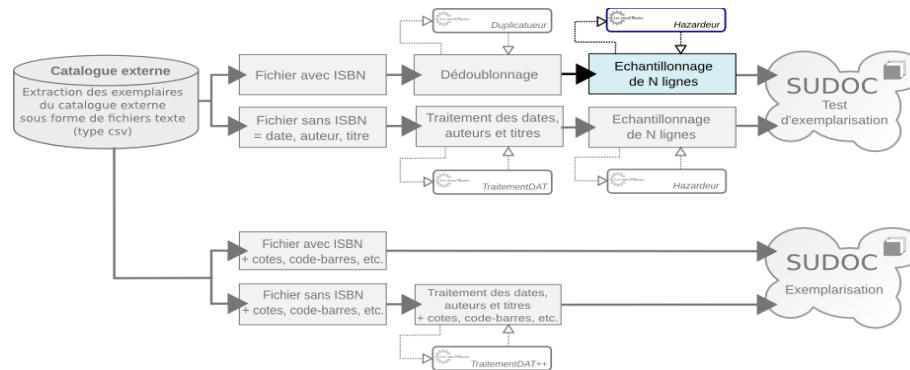
000370257X
000370288X
019432091X
019434648X
019459226X
019459372X
052131108X
052131240X
058242951X
140810145
140813594
175556954
175568782
1853462349
1874016402
1874016410
1874016496
1874016739
1874016836
1900702312
1900702401
1902031016
1902031040
1902031091
1902031318
194317889
194319040
194319229

5000 lignes

000370257X
000370288X
019432091X
019434648X
019459226X
019459372X
052131108X
052131240X
058242951X
140810145
140813594
175556954
175568782
1853462349
1874016402
1874016410
1874016496
1874016739
1874016836
1900702312
1900702401
1902031016
1902031040
1902031091
1902031318
194317889
194319040
194319229

3833 lignes

Enregistrer sous ...



Étape 1-2 : ISBN - Échantillonnage

**Extraire au hasard 1000 lignes d'un fichier d'ISBN
avec le programme *Hazardeur***

- Exécutez le programme *Hazardeur*.
- Ouvrez le fichier d'exemple *isbn_3883.txt* créé à l'étape 1.
- Entrez le nombre de lignes à garder. *1000* est un bon nombre.
- Cliquez sur le bouton *Action*.
- Enregistrez le fichier résultat. Nommez-le *isbn_1000.txt*

C'est un tel fichier qui sera envoyé pour un test d'exemplarisation ISBN.

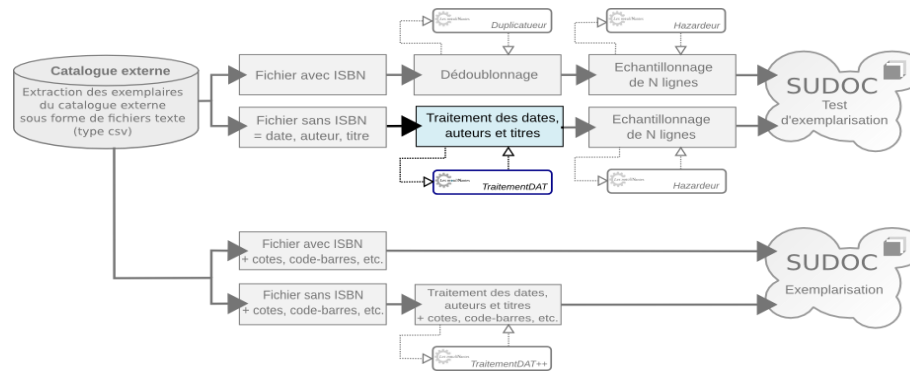
Échantillonnage avec le hazardeur

The screenshot shows a software window titled "Echantillonneur de lignes". The interface is divided into a control panel on the left and a list of sampled lines on the right. The control panel includes a title "Hazardeur", a description of the program's function, a file selection area with an "Ouvrir" button, a text input field for the number of lines to keep (set to "1000"), and two buttons: "Action !" and "Enregistrer les résultats". A yellow die icon is at the bottom. The list on the right shows line numbers and their corresponding text from a file named "isbn_3833.txt".

Annotations:

- Nombre de lignes à tirer au hasard:** Points to the "1000" input field.
- Nom du fichier ouvert:** Points to the "isbn_3833.txt" text.
- Le bouton **Action** ne s'allume que si le nombre de lignes a été renseigné.** Points to the "Action !" button.
- Le bouton **Enregistrer** ne s'allume que si le bouton **Action** a été pressé.** Points to the "Enregistrer les résultats" button.

Line	Content
1 : 2	019432091X
2 : 5	019459372X
3 : 7	052131240X
4 : 21	1902031016
5 : 25	194317889
6 : 26	194319040
7 : 40	194372022
8 : 47	2010157303
9 : 50	2010163982
10 : 57	2010179633
11 : 69	2010191560
12 : 72	2010194438
13 : 73	2010194551
14 : 84	2010206010
15 : 85	2010206029
16 : 86	2010206037
17 : 88	201020607X
18 : 93	2010212649
19 : 96	2011254523
20 : 101	2011452139
21 : 102	2011452465
22 : 105	201145297X
23 : 106	2011453496
24 : 107	2011453704
25 : 112	201155098X
26 : 113	2011551390
27 : 117	2011553849
28 : 120	2011703484
29 : 127	2011703840
30 : 128	2011703921
31 : 130	2011703980
32 : 135	2011704116



Etape 2-1 : Traitement Date-Auteur-Titre

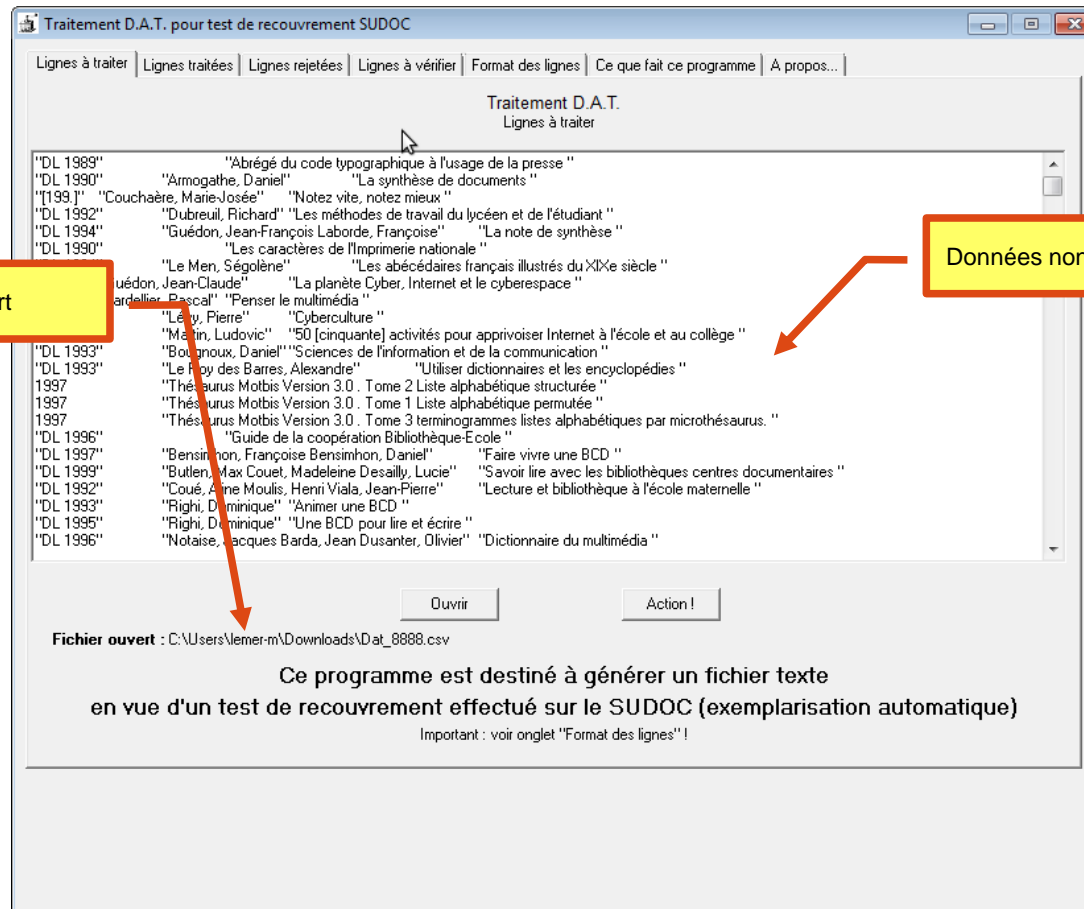
Traiter les données date-auteur-titre avec le programme TraitementDAT

- Exécutez le programme *TraitementDAT*
- Ouvrez le fichier d'exemple *dat_8888.csv*
- Lancez le traitement en cliquant sur le bouton *Action*.
- Consultez les onglets *Lignes traitées*, *Lignes rejetées* et *Lignes à vérifier*.
- Enregistrez les 7928 lignes traitées dans le fichier résultat. Nommez-le *dat_7928.txt*

Attention : TraitementDAT trie et dédoublonne !

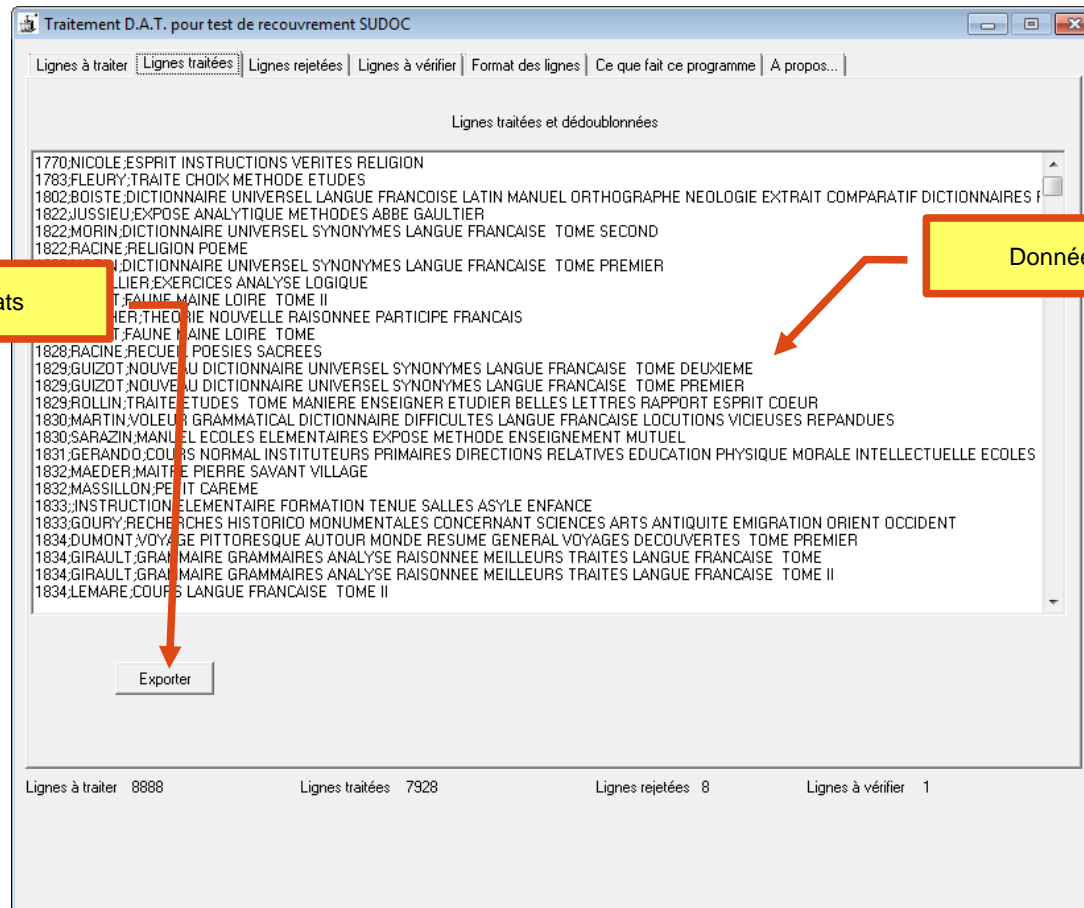
Traitement date-auteur-titre avec TraitementDAT

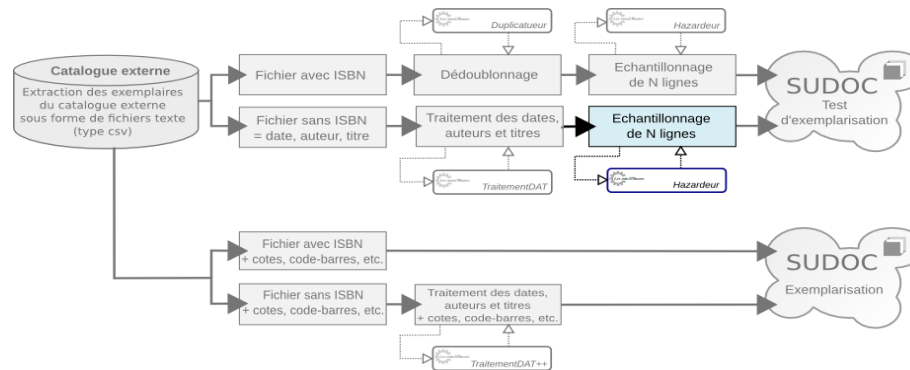
Ouverture du fichier



Traitement date-auteur-titre avec TraitementDAT

Après avoir cliqué sur *Action*





Etape 2-2 : Date-Auteur-Titre - Échantillonnage

**Extraire au hasard 1000 lignes d'un fichier D.A.T.
avec le programme Hazardeur**

- Exécutez le programme *Hazardeur*.
- Ouvrez le fichier *dat_7928.txt* créé à l'étape 2-1
- Entrez le nombre de lignes à garder. *1000* est un bon nombre (bis).
- Cliquez sur le bouton *Action*.
- Enregistrez le fichier résultat. Nommez-le *dat_1000.txt*

C'est un tel fichier qui sera envoyé pour le test d'exemplarisation D.A.T.

Copie d'écran page suivante

Échantillonnage avec le hazardeur

Echantillonneur de lignes

Hazardeur

Ce programme extrait N lignes d'un fichier texte.
Ces lignes sont tirées au hasard.
Une ligne ne peut pas être choisie plus d'une fois.

Fichier à ouvrir

Ouvrir

7928 lignes

dat_7928.csv

Nombre de lignes à garder

1000

Action !

Enregistrer les résultats

Michel.lemer@univ-nantes.fr

1 : 19

2 : 30

3 : 35

4 : 81

5 : 83

6 : 85

7 : 98

8 : 100

9 : 116

10 : 124

11 : 125

12 : 127

13 : 134

14 : 148

15 : 153

16 : 168

17 : 171

18 : 188

19 : 207

20 : 211

21 : 216

22 : 220

23 : 235

24 : 236

25 : 242

26 : 245

27 : 251

28 : 255

29 : 258

30 : 267

31 : 275

32 : 277

1832;MASSILLON;PETIT CAREME

1837;JEAN;CONDUITE ECOLES CHRETIENNES

1843;ALLARD;RECUEIL METHODIQUE LOIS ORDONNANCES REC

1859;ROUSSEAU;OEUVRES LYRIQUES

1860;JANET;FAMILLE

1861;DUPANLOUP;HAUTE EDUCATION INTELLECTUELLE TOME

1865;PONSARD;OEUVRES COMPLETES TOME DEUXIEME

1865;SIMON;ECOLE

1870;MICHELET;FILS

1874;ELEMENTS GEOMETRIE COURS MATH ELEMENTAIRE

1874;BRACHET;NOUVELLE GRAMMAIRE FRANCAISE FONDEE HI

1874;CHARBONNEAU;COURS THEORIQUE PRATIQUE PEDAGOG

1876;DEZOBRY;DICTIONNAIRE GENERAL BIOGRAPHIE HISTOIR

1878;JANET;PHILOSOPHIE FRANCAISE CONTEMPORAINE

1879;BERSOT;QUESTIONS ENSEIGNEMENT

1879;MICHELET;HISTOIRE FRANCE TOME XIII HENRI IV RICHEL

1879;MICHELET;HISTOIRE FRANCE TOME XV LOUIS XIV REVOC

1880;PAROZ;HISTOIRE UNIVERSELLE PEDAGOGIE

1882;AYER;GRAMMAIRE COMPAREE LANGUE FRANCAISE

1882;MUTEAU;ECOLES COLLEGES PROVINCE TEMPS RECULES

1883;DUMESNIL;PEDAGOGIE REVOLUTIONNAIRE

1883;ROUCHE;TRAITE GEOMETRIE

1885;PEREZ;ENFANT TROIS SEPT ANS

1885;STEEG;COURS MORALE USAGE INSTITUTEURS ELEVES E

1886;CATALOGUE BIBLIOTHEQUES ECOLES NORMALES

1886;BIZOS;FENELON EDUCATEUR

1886;JANET;HISTOIRE SCIENCE POLITIQUE RAPPORTS MORALE

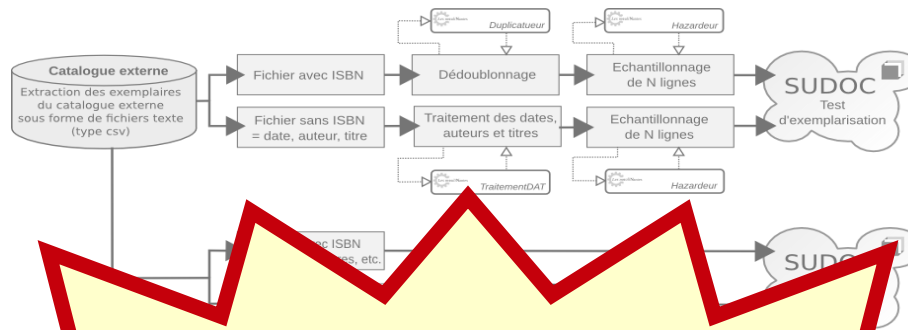
1887;CERTIFICAT ETUDES PRIMAIRES SUPERIEURES

1887;ECOLE SCIENCE JUSQU A RENAISSANCE

1887;GREARD;EDUCATION INSTRUCTION ENSEIGNEMENT SEC

1887;TOCQUEVILLE;DEMOCRATIE AMERIQUE TOME

1887;TOCQUEVILLE;DEMOCRATIE AMERIQUE TOME III



Attention !

TraitementDAT++ ne dédouble pas !

Il est donc normal de ne pas obtenir
le même nombre de notices
qu'avec TraitementDAT !

- Ouvrez le fichier
- Lancez l'application
- Consultez les notices à vérifier.
- Enregistrez le résultat dans un fichier nommé `datCCb_7928.txt`

C'est un tel fichier qui sera envoyé pour l'exemplarisation D.A.T.

Traitement date auteur titre cote code-barres avec TraitementDAT++

Traitement D.A.T.++ pour exemplarisation SUDOC

Lignes à traiter | Lignes traitées | Lignes rejetées | Lignes à vérifier | Format des lignes | Ce que fait ce programme | A propos...

Traitement D.A.T. ++

Lignes à traiter : **8888**

"DL 1989"	"Abrégé du code typographique à l'usage de la presse "	"001.4 ABR"	7002498
"DL 1990"	"Armogathe, Daniel" "La synthèse de documents "	"001.4 ARM"	7004119
"[199.]"	"Couchaère, Marie-Josée" "Notez vite, notez mieux "	"001.4 COU"	7003392
"DL 1992"	"Dubreuil, Richard" "Les méthodes de travail du lycéen et de l'étudiant "	"001.4 DUB"	7003389
"DL 1994"	"Guédon, Jean-François Laborde, Françoise" "La note de synthèse "	"001.4 GUE"	7003444
"DL 1990"	"Les caractères de l'Imprimerie nationale "	"002 CAR"	7002279
"DL 1984"	"Le Men, Ségolène" "Les abécédaires français illustrés du XIXe siècle ""	"002 LEM"	7002503
1997	"Guédon, Jean-Claude" "La planète Cyber, Internet et le cyberspace "	"004 GUE"	7003309
1997	"Lardellier, Pascal" "Penser le multimédia "	"004 LAR"	4019859
"DL 1997"	"Lévy, Pierre" "Cyberculture "	"004 LEV"	7003306
"DL 1998"	"Martin, Ludovic" "50 [cinquante] activités pour apprivoiser Internet à l'école et au collège "	"004 MAR"	7003377
"DL 1993"	"Bougnoux, Daniel" "Sciences de l'information et de la communication "	"020 BOU"	7002383
"DL 1993"	"Le Roy des Barres, Alexandre" "Utiliser dictionnaires et les encyclopédies "	"020.7 LER"	7003441
1997	"Thésaurus Motbis Version 3.0 . Tome 2 Liste alphabétique structurée "	"025.4 THE"	7039491
1997	"Thésaurus Motbis Version 3.0 . Tome 1 Liste alphabétique permutée ""	"025.4 THE"	7039492
1997	"Thésaurus Motbis Version 3.0 . Tome 3 terminogrammes listes alphabétiques par microthésaurus. "	"025.4 THE"	7039493
"DL 1996"	"Guide de la coopération Bibliothèque-Ecole "	"027 GUI"	7018378
"DL 1997"	"Bensimhon, Françoise Bensimhon, Daniel" "Faire vivre une BCD "	"027.007 BEN"	7018379
"DL 1999"	"Butlen, Max Couet, Madeleine Desailly, Lucie" "Savoir lire avec les bibliothèques centres documentaires "	"027.007 BUT"	7018562
"DL 1992"	"Coué, Aline Moulis, Henri Viala, Jean-Pierre" "Lecture et bibliothèque à l'école maternelle "	"027.007 COU"	7018533
"DL 1993"	"Righi, Dominique" "Animer une BCD "	"027.007 RIG"	7018452
"DL 1995"	"Righi, Dominique" "Une BCD pour lire et écrire "	"027.007 RIG"	7018453
"DL 1996"	"Notaise, Jacques Barda, Jean Dusanter, Olivier" "Dictionnaire du multimédia "	"030 DIC"	7019760

Ouvrir Action !

Fichier ouvert : C:\Users\lemer-m\Downloads\DatCCb_8888.csv

Ce programme est destiné à générer un fichier texte en vue d'un test de recouvrement effectué sur le SUDOC

Important : voir onglet "Format des lignes" !

Lignes à traiter 8888 Lignes traitées 8880 Lignes rejetées 8 Lignes à vérifier 0

PARTIE 3

Encore des moulinantes après l'exemplarisation !

Cette partie ne concerne plus directement l'ABES.

En effet, il s'agira ici de l'importation dans le SIGB local des notices MARC renvoyées par le Sudoc après exemplarisation automatique.

On traitera principalement de l'outil **MarcRCR** destiné à modifier un lot de notices MARC en vue d'une importation **avec données d'exemplaires**.

Ajouter des données locales à des notices MARC avec la moulinante MarcRCR

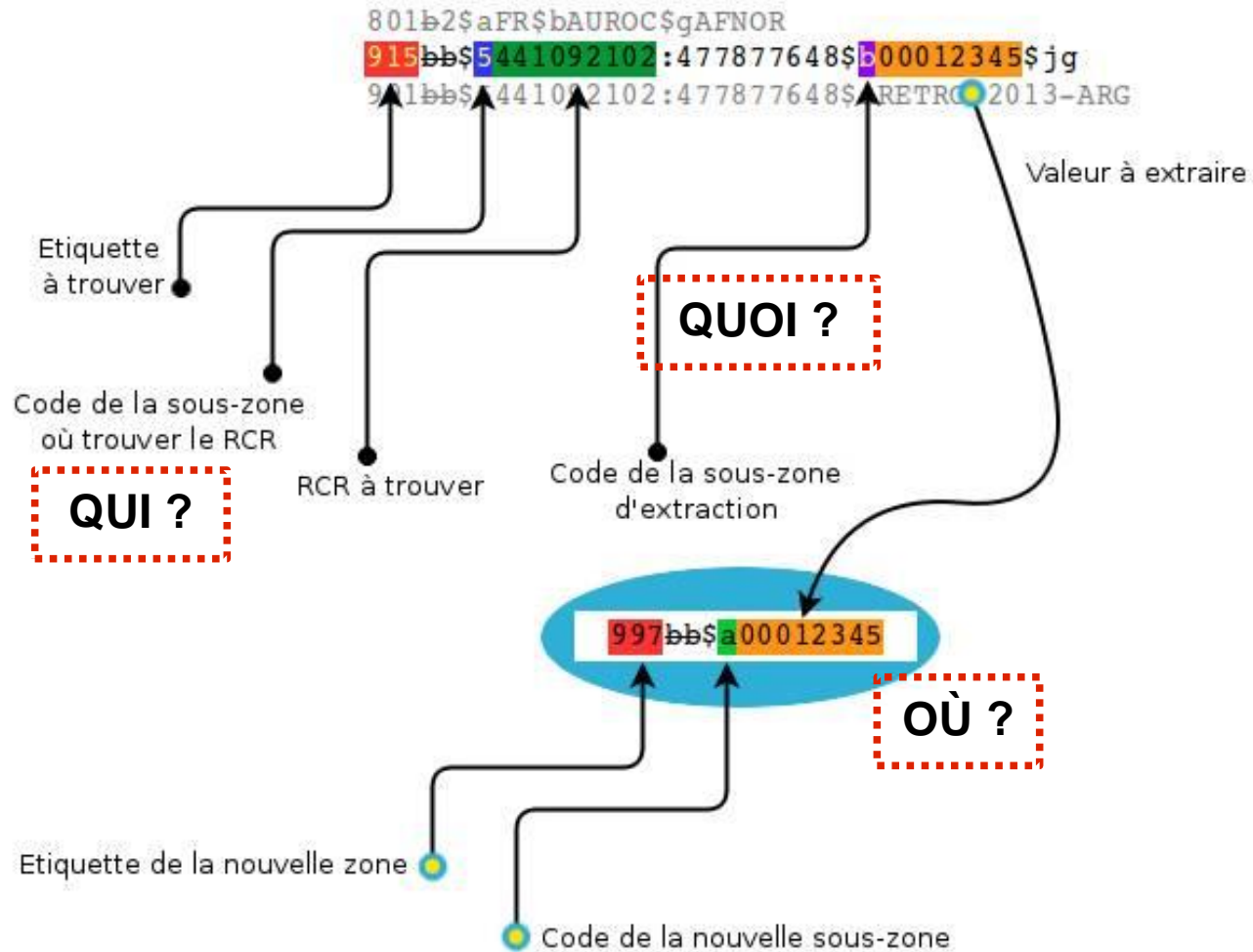
Selon le SIGB et son paramétrage, les données locales des notices MARC renvoyées après l'exemplarisation automatique ne peuvent pas toujours être importées telles quelles.

La moulinante **MarcRCR** a été conçue pour ça : lorsque la notice répond à certaines conditions (c'est-à-dire lorsque tel RCR est trouvé dans une sous-zone précise), **MarcRCR** repère les données d'exemplaires et les ajoute dans une zone prévue (*) par le SIGB pour l'importation.

Schéma du principe MarcRCR page suivante

(*) Dans ce document, on a retenu la valeur **997** pour l'étiquette de zone d'exemplaire.

Schéma du principe MarcRCR



Pré-requis (Kik Waooo)

Vous devez être en possession de quelques éléments avant le traitement MarcRCR.

Quel est le RCR des notices concernées ? (*Qui ?*)

Quelles zones MARC désirez-vous extraire de la notice ? (*Quoi?*)

Dans quelles zones MARC de votre SIGB iront les données supplémentaires ? (*Où?*)

Utilisation de la moulinante MarcRCR

Exécutez le programme MarcRCR.

A la première utilisation, aucun paramètre n'est encore défini : c'est à vous de le faire.

Tant que le paramétrage n'a pas été effectué, il est impossible d'utiliser le bouton *Action*.

Pour cela, cliquez sur le bouton *Modifier les paramètres*.

Par défaut, l'écran des définitions est verrouillé. Cliquez sur le bouton *Passer en mode "Déverrouillé"* afin d'entrer, de modifier ou de supprimer des définitions.

Les définitions

Vous devrez définir trois types de paramètres :

- les valeurs globales (RCR, zone d'import du SIGB)
- les valeurs statiques (Quels ajouts systématiques ?)
- les valeurs dynamiques (Qu'extraire et qu'en faire ?)

Valeurs globales

Les valeurs globales que vous devez entrer sont :

- le RCR concerné,
- l'étiquette MARC de la zone à ajouter,
- les descripteurs.

Dans la copie d'écran ci-dessous :

- le RCR vaut **441092103**,
- pour ce SIGB, les données d'exemplaires des notices MARC importées doivent se situer en zone **997**,
- les descripteurs doivent être **deux blancs**.

Définissez les paramètres globaux

Définition des paramètres globaux

RCR à trouver :	<input type="text" value="441092103"/>	Exemple : 441099901	Zone à ajouter à chaque notice MARC (si concordance !)
			Etiquette (852 ou 9xx) <input type="text" value="997"/> Exemple : 997
<input checked="" type="checkbox"/> Ne garder que les notices où il y a au moins une concordance.			Descripteurs <input type="text" value=""/> Par défaut : blanc blanc

En mode verrouillé, le caractère "blanc" est représenté par le caractère β

Si la case "Ne garder que les notices où il y a au moins une concordance" est cochée, les notices non concordantes ne seront pas écrites. Sinon, toutes les notices sont écrites, celles modifiées comme celles non-modifiées.

Valeurs statiques

Définissez les valeurs statiques et leur code de sous-zone.






Ces ensembles (code de sous-zone et sa valeur) seront ajoutés **systematiquement** à chaque notice.

Exemple.

- Selon le protocole d'importation de mon SIGB, le type de prêt doit être renseigné en 997\$**e**
- Tous les exemplaires traités sont empruntables, type **EM** dans mon SIGB.
- En conséquence, on ajoutera un code de sous-zone **e** auquel on attribuera la valeur **EM**

Définissez les sous-zones à créer dans

Définitions des sous-zones à ajouter - Valeurs statiques

Code de sous-zone	<input type="text" value="a"/>	Valeur	<input type="text" value="NUM"/>	
Code de sous-zone	<input type="text" value="d"/>	Valeur	<input type="text" value="NUM20"/>	
Code de sous-zone	<input type="text" value="e"/>	Valeur	<input type="text" value="EM"/>	
Code de sous-zone	<input type="text" value="s"/>	Valeur	<input type="text" value="d"/>	
				

Un exemple où on ajoutera les sous-zones **\$a** (Division=NUM), **\$d** (Collection=NUM20), **\$e** (Type de prêt=EM) et **\$s** (Statut=d)

Valeurs dynamiques

Pour chaque notice du fichier ouvert, MarcRCR ira chercher le RCR dans une ou plusieurs sous-zones précises, définies par l'utilisateur. En cas de succès, le programme extraira la valeur contenue dans telle sous-zone de cette zone. Cette valeur et son code de sous-zone (lui-aussi à définir par l'utilisateur) seront ajoutés à la zone d'export.

Ces ensembles (code de sous-zone et sa valeur)
ne seront ajoutés à chaque notice
que s'il y a concordance.

Exemple de données dynamiques

Dans une notice renvoyée par le Sudoc, la cote se situe dans la zone **930**\$**a** et le RCR correspondant se situe dans la même zone mais en \$**5**

Par ailleurs, dans mon SIGB, la cote importée est à écrire en zone 997\$c




Pour la cote, il faudra donc créer une définition de valeurs dynamiques ainsi :

- Etiquette : **930**
- Sous-zone du RCR : **5**
- Sous-zone à extraire : **a**
- Sous-zone à insérer : **c**

Dans cette copie d'écran, on trouve deux définitions dynamiques.

Pour la première (le code-barres), le programme cherchera le RCR en **915\$5**.
S'il est trouvé, la valeur est à extraire de **915\$b** et sera insérée en zone **997\$b**

Pour la seconde (la cote), le programme cherchera le RCR en **930\$5**.
S'il est trouvé, la valeur est à extraire de **930\$a** et sera insérée en zone **997\$c**

Etiquette	Sous-zone du RCR	Sous-zone à extraire	Sous-zone insérée	
915	5	b	b	
930	5	a	c	
				

Exemple de zones à valeurs concordantes pour le RCR N° **441092103**

915__\$**5441092103**:123456789\$b0081111112

▶ **441092103** est bien en **915\$5**. On extrait donc la valeur de **915\$b** à savoir 0081111112

930__\$**5441092103**:123456789\$aM 123456

▶ **441092103** est bien en **930\$5**. On extrait donc la valeur de **930\$a** à savoir M 123456

Exemple de zone générée avec nos exemples de définitions.

997__ \$aNUM\$b00811111112\$cM 123456\$dNUM20\$eEM\$s d

- Etiquette : **997**
- Descripteurs : **blanc blanc**
- \$a Division = **NUM**
- \$b Code-barres = **00811111112**
- \$c Cote = **M 123456**
- \$d Collection = **NUM20**
- \$e Type de prêt = **EM**
- \$s Statut = **d** (pour disponible)

Les données globales sont **en vert**, les données statiques **en bleu**
et les données dynamiques **en rouge**

Notes sur MarcRCR

- Lors du paramétrage des définitions statiques et dynamiques, des contrôles empêchent l'utilisation d'un code de sous-zone plus d'une fois.
- Dans la réalité, la valeur de la sous-zone **\$5** n'est pas le RCR tout seul mais le couple RCR et numéro (caractère de séparation : deux-points). Ce numéro de l'exemplaire est utilisé par MarcRCR pour pointer sur les bons enregistrements.

Exemple :

915__\$5441092103:123456789\$b00811111112

915__\$5441092103:987654321\$b00822222221

930__\$5441092103:123456789\$aM 123456

930__\$5441092103:987654321\$aBE 44 (2)

- Pour les puristes, MarcRCR se prononce *mar creu creu*.

Notes sur le format CSV (1/3)

Historiquement, le format CSV est un format de fichier texte destiné à un tableur.

- Chaque ligne est séparée de la précédente par un retour-chariot.
- Chaque ligne contient un enregistrement.
- Chaque enregistrement contient une ou plusieurs valeurs, séparées par une virgule d'où le nom CSV : *Comma Separated Values*

Désormais, l'utilisateur peut choisir le caractère de séparation : *virgule* mais aussi *point-virgule*, *tabulation*, etc.

Il peut aussi choisir le caractère qui entoure les valeurs de type Chaînes de caractère qui contiennent un espace : *apostrophes* ou *guillemets*.

Comme pour tous les fichiers texte, il faudra aussi lui choisir son encodage (UTF-8, ANSI, etc.).

Caractère de séparation : *point-virgule* - Séparateur de texte : *guillemets*

```
2010;"LA FONTAINE";FABLES
2000;PAGNOL;"GLOIRE PERE"
```

Caractère de séparation : *point-virgule* - Séparateur de texte : *apostrophe*

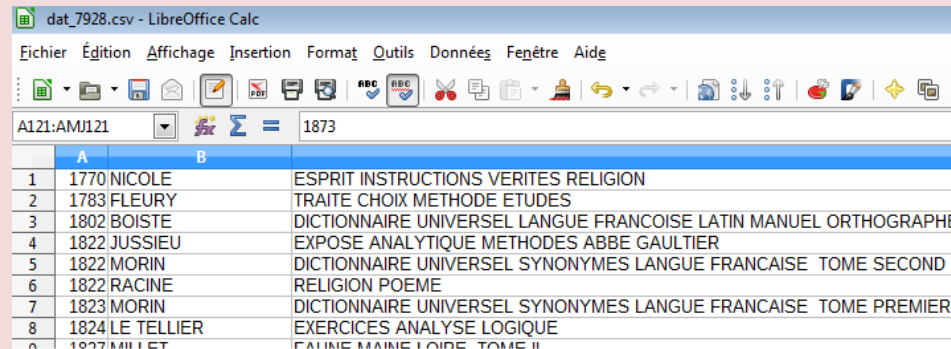
```
2010;'LA FONTAINE';FABLES
2000;PAGNOL;'GLOIRE PERE'
```

Caractère de séparation : *tabulation* - Séparateur de texte : *guillemets*

```
2010↔"LA FONTAINE"↔FABLES
2000↔PAGNOL↔"GLOIRE PERE"
```

Notes sur le format CSV (2/3)

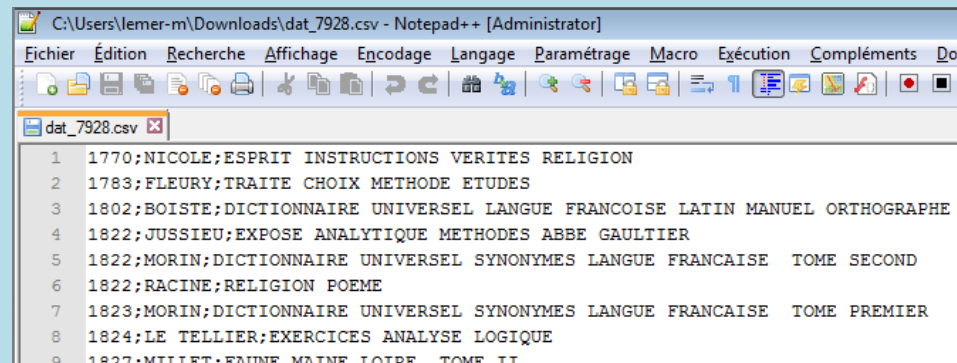
Lorsqu'on l'ouvre avec un tableur, le fichier CSV est **interprété** et chaque valeur est alors intégrée dans une cellule.



The screenshot shows the LibreOffice Calc interface with the file 'dat_7928.csv' open. The data is displayed in a table with columns A and B. The first column (A) contains an ID number, and the second column (B) contains a list of books. The data is as follows:

	A	B
1	1770 NICOLE	ESPRIT INSTRUCTIONS VERITES RELIGION
2	1783 FLEURY	TRAITE CHOIX METHODE ETUDES
3	1802 BOISTE	DICTIONNAIRE UNIVERSEL LANGUE FRANCOISE LATIN MANUEL ORTHOGRAPHE
4	1822 JUSSIEU	EXPOSE ANALYTIQUE METHODES ABBE GAULTIER
5	1822 MORIN	DICTIONNAIRE UNIVERSEL SYNONYMES LANGUE FRANCAISE TOME SECOND
6	1822 RACINE	RELIGION POEME
7	1823 MORIN	DICTIONNAIRE UNIVERSEL SYNONYMES LANGUE FRANCAISE TOME PREMIER
8	1824 LE TELLIER	EXERCICES ANALYSE LOGIQUE
9	1827 MILLET	FAUNE MAINE LOIRE TOME II

Lorsqu'on l'ouvre avec un éditeur de texte brut, le fichier CSV se présente sous la forme d'une **suite de lignes**, pas forcément faciles à lire par un être humain...



The screenshot shows the Notepad++ interface with the file 'dat_7928.csv' open. The data is displayed as a plain text file with each line representing a row of data, separated by semicolons. The data is as follows:

```

1 1770;NICOLE;ESPRIT INSTRUCTIONS VERITES RELIGION
2 1783;FLEURY;TRAITE CHOIX METHODE ETUDES
3 1802;BOISTE;DICTIONNAIRE UNIVERSEL LANGUE FRANCOISE LATIN MANUEL ORTHOGRAPHE
4 1822;JUSSIEU;EXPOSE ANALYTIQUE METHODES ABBE GAULTIER
5 1822;MORIN;DICTIONNAIRE UNIVERSEL SYNONYMES LANGUE FRANCAISE TOME SECOND
6 1822;RACINE;RELIGION POEME
7 1823;MORIN;DICTIONNAIRE UNIVERSEL SYNONYMES LANGUE FRANCAISE TOME PREMIER
8 1824;LE TELLIER;EXERCICES ANALYSE LOGIQUE
9 1827;MILLET;FAUNE MAINE LOIRE TOME II
  
```

N'importe quel fichier texte peut être qualifié de CSV si toutes ses lignes sont conçues à l'identique.

Notes sur le format CSV (3/3)

Lors d'une exemplarisation automatique, il est indispensable de disposer d'un outil :

- capable d'ouvrir rapidement un fichier texte de type CSV sans pour autant l'interpréter,
- capable d'accepter de modifier des données
- capable de gérer l'encodage

Ce type d'outil se nomme *éditeur de texte*, à ne pas confondre avec *traitement de texte*.

Avec un éditeur de texte, pas de formatage (gras, italiques, ...), pas de tableau, pas d'images, rien de tout ça : on est dans le texte brut.

Il offre néanmoins des fonctions rechercher-remplacer, et aussi les fonctions d'encodage, ce qui est indispensable lors de communication avec le Sudoc.

On trouve des éditeurs de texte non seulement gratuits mais très efficaces, faciles d'emploi et capables d'ouvrir de très gros fichiers.

- Notepad++ : <https://notepad-plus-plus.org/fr>
- Geany : <http://www.geany.org>
- PSPad : <http://www.pspad.com/fr>

Notes sur les moulinantes en général

Les moulinantes ne sont pas toutes décrites dans ce document.

Le site de dépôt officiel des Moulinantes se situe sur le gitlab de l'université de Nantes.

<https://gitlab.univ-nantes.fr/michel.lemer/moulinantes>

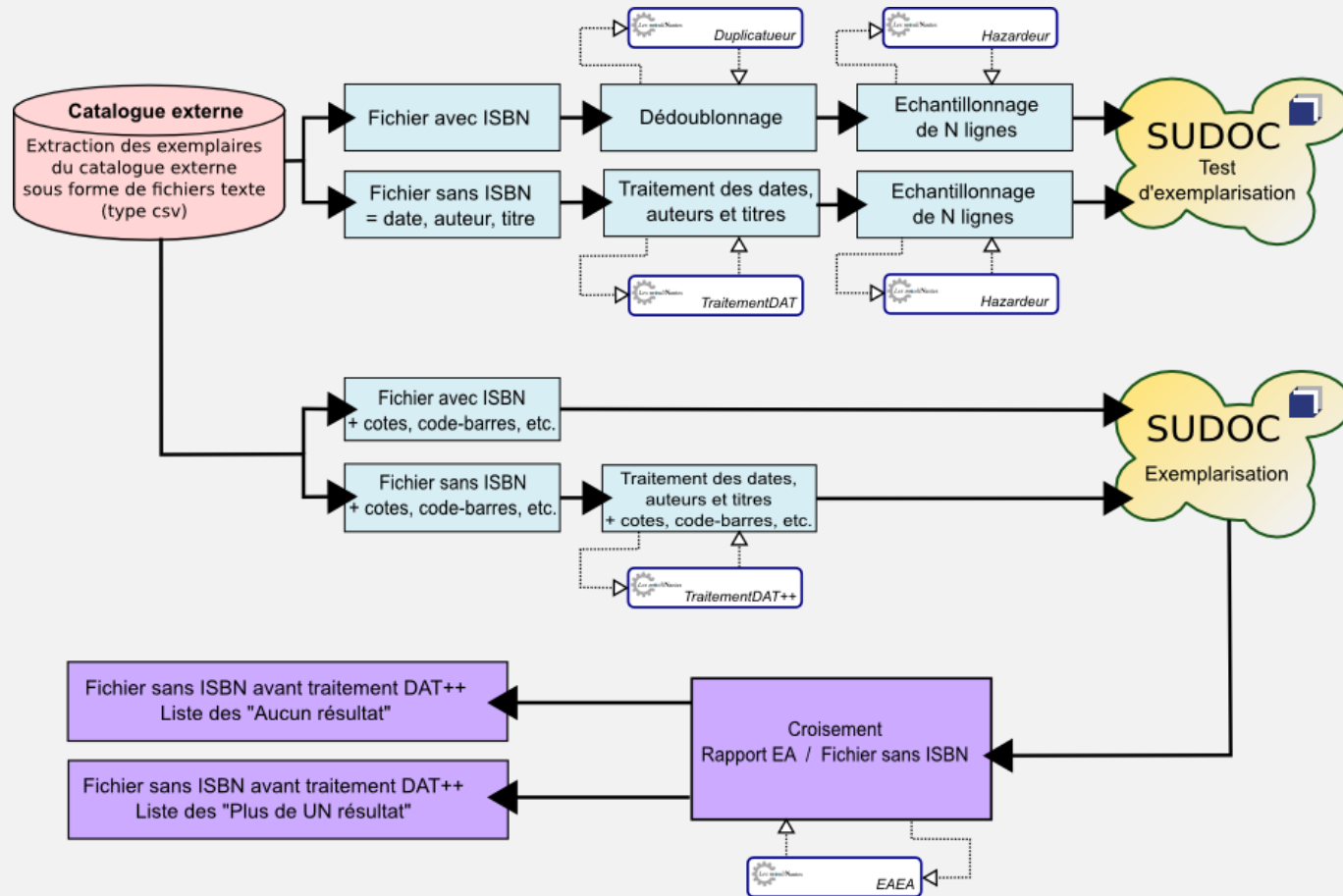
Cependant, il existe une version plus intuitive sur le site personnel du développeur.

<http://michel.lemer.free.fr/moulinantes/>

Au moment où ce document a été créé, si l'on tape le mot *moulinantes* dans Google, on obtient les liens ci-dessus dès les premières réponses.

Novembre 2015 La moulinantes EAEA est quasi terminée.

Avec EAEA, la boucle est bouclée : croisez le fichier de rapport d'EA fourni par l'ABES avec votre fichier tableur original (avant traitementDAT) afin de générer deux beaux fichiers tableur avec données non traitées : un pour les « zéro réponses », l'autre pour les « plus d'une réponse ».



Des questions ?