

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 1 sur 7

Date	Version N°	Nature des modifications	Rédigé par	Vérifié par	Validé par
29/08/2016	00	Création	Annie	Mariolla/Bebi	Mariolla/Bebi

Outil :

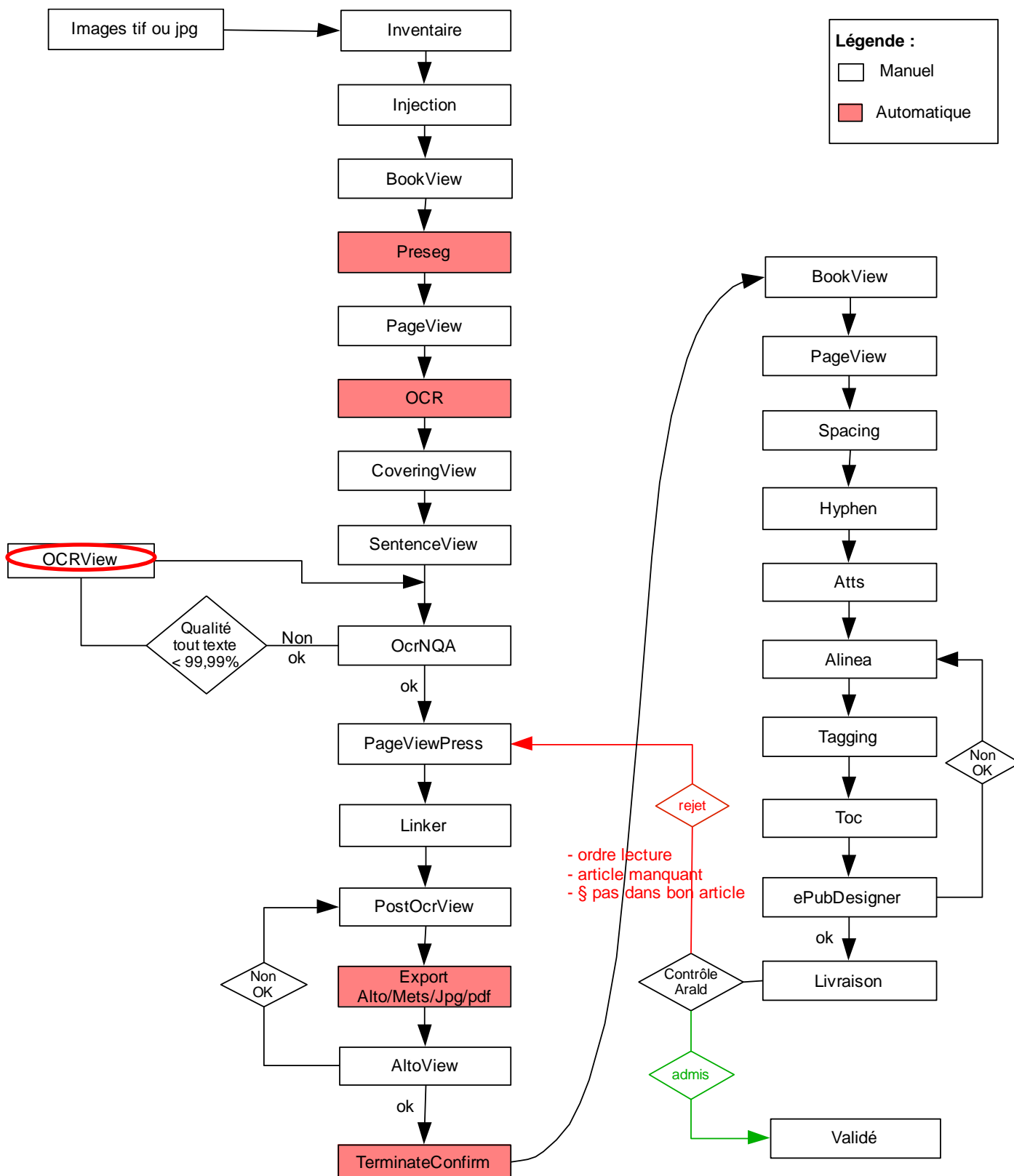
- OcrView for IsakoStudio

Tâche :

- Vérification/Correction des textes douteux en rouge et les textes non coloriés

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 2 sur 7

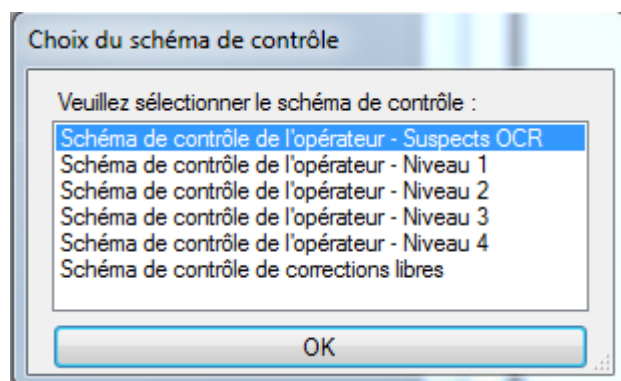
Logigramme



	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 3 sur 7

OcrView (1 p/h)

L'application OcrView permettra plusieurs niveaux d'échantillonnage (profils d'ouverture)



Chaque niveau permet d'échantillonner "un peu plus" jusqu'au dernier niveau qui échantillonne tous les mots qui n'ont pas déjà été contrôlés aux passes précédents.


L'application OcrView affiche, en pourcentage, la progression du travail de correction par rapport au niveau en cours mais ne dit pas quel est le taux réel.

Quand l'opérateur arrive à 100% des mots échantillonnés, il passe à la page suivante.

Si l'opérateur veut forcer le passage à la page suivante sans finir de corriger les mots échantillonnés dans la page en cours, OcrView lui affiche un message d'avertissement avant passer à la page suivante.

A chaque niveau (i.e., à chaque échantillonnage) les mots déjà contrôlés par l'opérateur ne sont pas échantillonnés à nouveau

L'échantillonnage est effectué chaque fois que la page est ouverte (ou ré-ouverte) dans OcrView. Cela veut dire que si l'opérateur revient en arrière à une page déjà vue, des mots supplémentaires peuvent être échantillonnés à nouveau.

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 4 sur 7

Les niveaux existants

Mots pour lesquels le moteur OCR a détecté ou marqué :	... un ou plusieurs caractères suspects	Hors dictionnaire ou Noms propres (selon le moteur OCR)	Mots commençant par un Majuscule	Mots avec hyphen/césure en fin de ligne	Mots qui correspondent à certaines RegEx	Pourcentage des mots restants, i.e., mots qui n'ont pas été échantillonnés	Commentaires
Niveau 1	100%	100%	100%	100%			
Niveau 2	x	x	x	x	100% de [RegEx-A]		
Niveau 3	x	x	x	x	100% de [RegEx-A] et de [RegEx-B]	30%	
Niveau 4						100%	

Les RegEx référencées sont :

RegEx-A

Mots ayant les caractères ou les combinaisons des caractères suivants :

[IL], [II], [il], [ii], [11], [1I], [1i], [Y], [1I], [m], [nn], [ri], [cl], [<char-alpha>']

RegEx-B

[e], [c]

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 5 sur 7

Description des couleurs

Caption
Espaces blancs suspects
Espaces échantillonnés en tant que mot et pas encore vus
Mots échantillonnés, pas encore vus et suspects
Echantillonné à moins de 100% : mots échantillonnés, pas encore vus et non suspects
Mots échantillonnés, pas encore vus et non suspects
Caractères suspects
Noms propres suspects
Mots suspects
Noms propres
Mots césurés
Mots hors dictionnaires
Taille de police <= 7 (au moins un caractère)
Mots ayant un ou plusieurs caractères en italique, sup ou sub
Mot ayant mélange de MAJ-MIN
Mot ayant mélange de chiffres-Alpha
Mot ayant mélange de AlphaNUM et ponctuation
Mot contenant une combinaison de [nmui][nmui]
Mot ne contenant qu'un seul caractère
Mots échantillonnés déjà vus par l'opérateur et pas encore vus par le contrôleur
Mots échantillonnés jamais vus et pas encore vus par le contrôleur
Mots échantillonnés déjà vus par l'opérateur et vus/corrigés par le contrôleur
Mots échantillonnés jamais vus et vus/corrigés par le contrôleur
Caractères en intersection avec les coordonnées du paramètre "boxCoords"
Caractères vus par l'opérateur
Caractères modifiés par l'opérateur
Caractères échantillonnés pour l'opérateur
Caractères vus par le contrôleur
Caractères modifiés par le contrôleur
Caractères échantillonnés pour le contrôleur
Caractères suspects
Caractères suspects OCR

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 6 sur 7

- ✓ Lire ligne par ligne sur les passages non coloriés et ceux en rouge



Recommencer | 1 4 OK

AVIS AUX GENS D'AFFAIRES.

AVIS AUX GENS D'AFFAIRES.


B I M X² X₂

AVIS • AUX • GENS • D'AFFAIRES.

n décret du Gouvernement provisoire

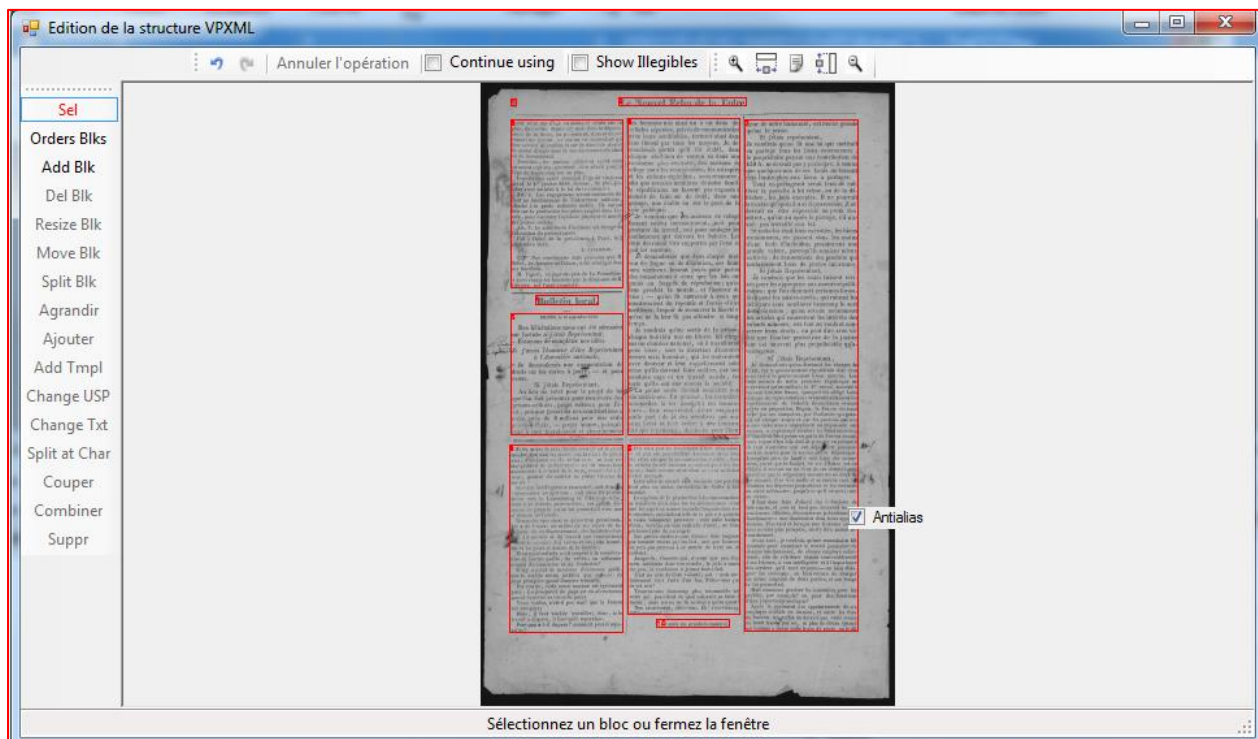
Correspondance, rue
N. Dame-des-Vict., 46
ANNONCES JUDICIAIRES ET AVIS DIVERS.
Cr. inséré gratuitement les articles qui ont un but d'utilité publique.
ABONNEMENT :
un année, 6 fr.
six mois, 3 50
trois mois, 2
PRIX des Insertions :
20 cent la ligne.
Annonces judiciaires :
10 centimes.
AVIS AUX GENS D'AFFAIRES.
Un décret du Gouvernement provisoire ayant

- En cas de saut de lignes, on peut les rajouter à partir du menu ci-dessous

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5554
		Date : 29/08/2016
	Arald - EPUB Etape : OcrView	Ver : 00
		Page 7 sur 7

- Erreur segmentation, possibilité de correction dans :

Menu : Tools/Graphic Editor



Transcription ancien français : Les deux numeros concernés "69GAZETTELYO-17470708" et "69NOUVELLESO-16781203".

De manière générale et comme il s'agit d'ancien français, il faut :

- remplacer tous les signes esperluette (« & ») par « et »
- remplacer les lettres 'e tréma' (« ë ») par des 'e sans tréma' (« e »). Par exemple « reçue » au lieu de « reçuë »
- remplacer les lettres 'i tréma' (« ï ») par des 'i sans tréma' (« i »). Par exemple « pays » au lieu de « païs »
- les "s" en "f" à changer en "s"
- les "oi" devraient être interprétés en "ai", ainsi que les "uy" en "ui", etc.