	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01
		Page 1 sur 7


Date	Version N°	Nature des modifications	Rédigé par	Vérifié par	Validé par
11/11/2014	00	Création	Annie	Bebi/Mariolla	Bebi/Mariolla
27/07/2016	01	Mise à jour	Annie	Bebi/Mariolla	Bebi/Mariolla

- Outil :

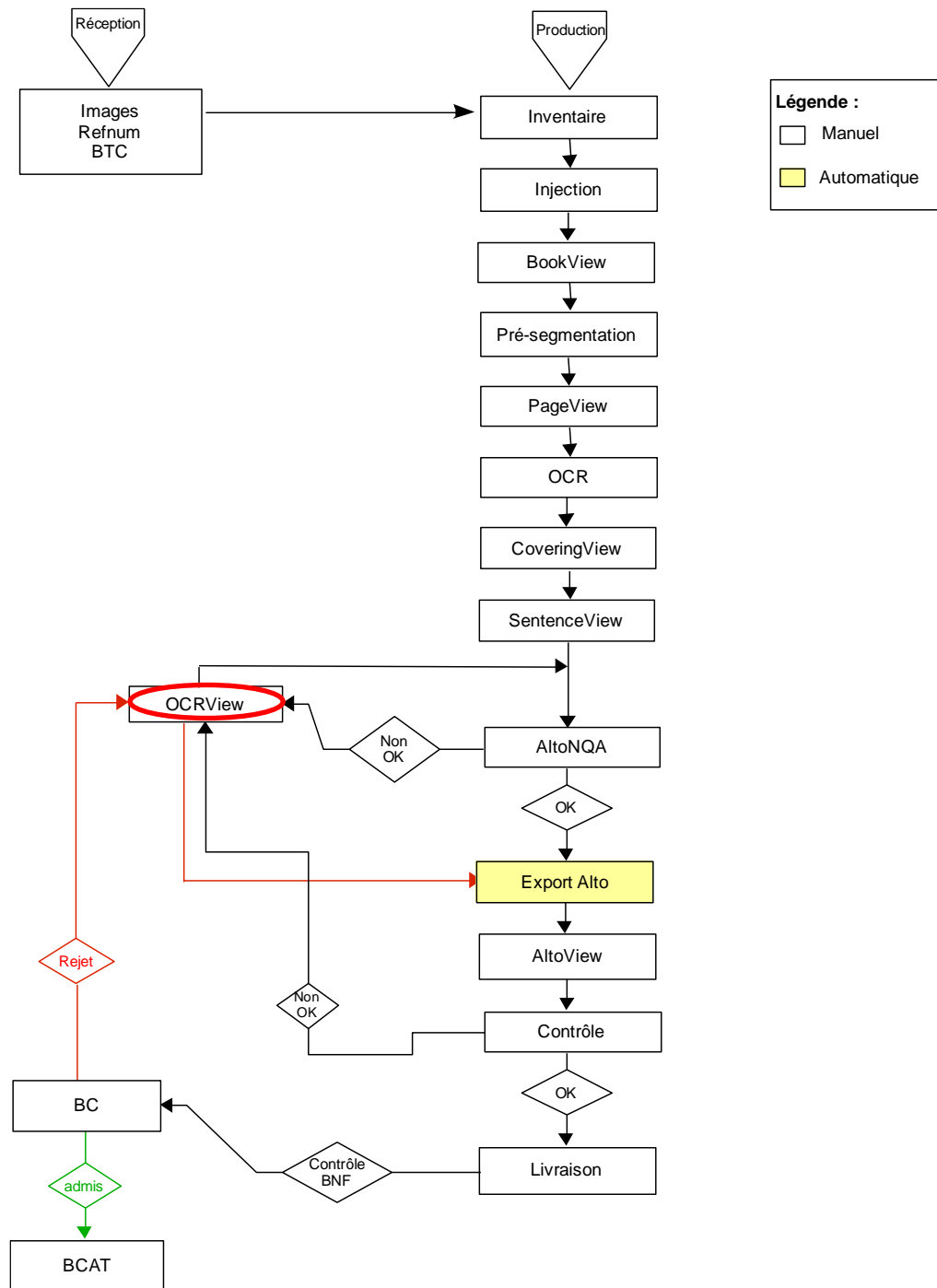
- OcrView for IsakoStudio


- Tâche :

- Vérification et Correction des textes douteux en rouge

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01
		Page 2 sur 7

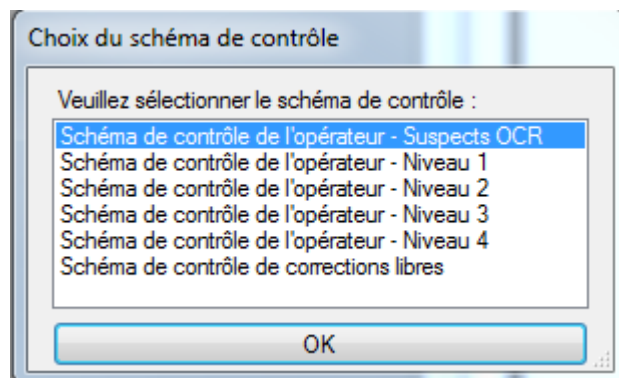
Logigramme



	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01
		Page 3 sur 7

BNFOcrView (cat K, R : 50p/h ; cat L : 30p/h ; cat M : 65p/h ; cat N : 18p/h ; cat O, Q : 60p/h ; cat P : 25p/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_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01
		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 :

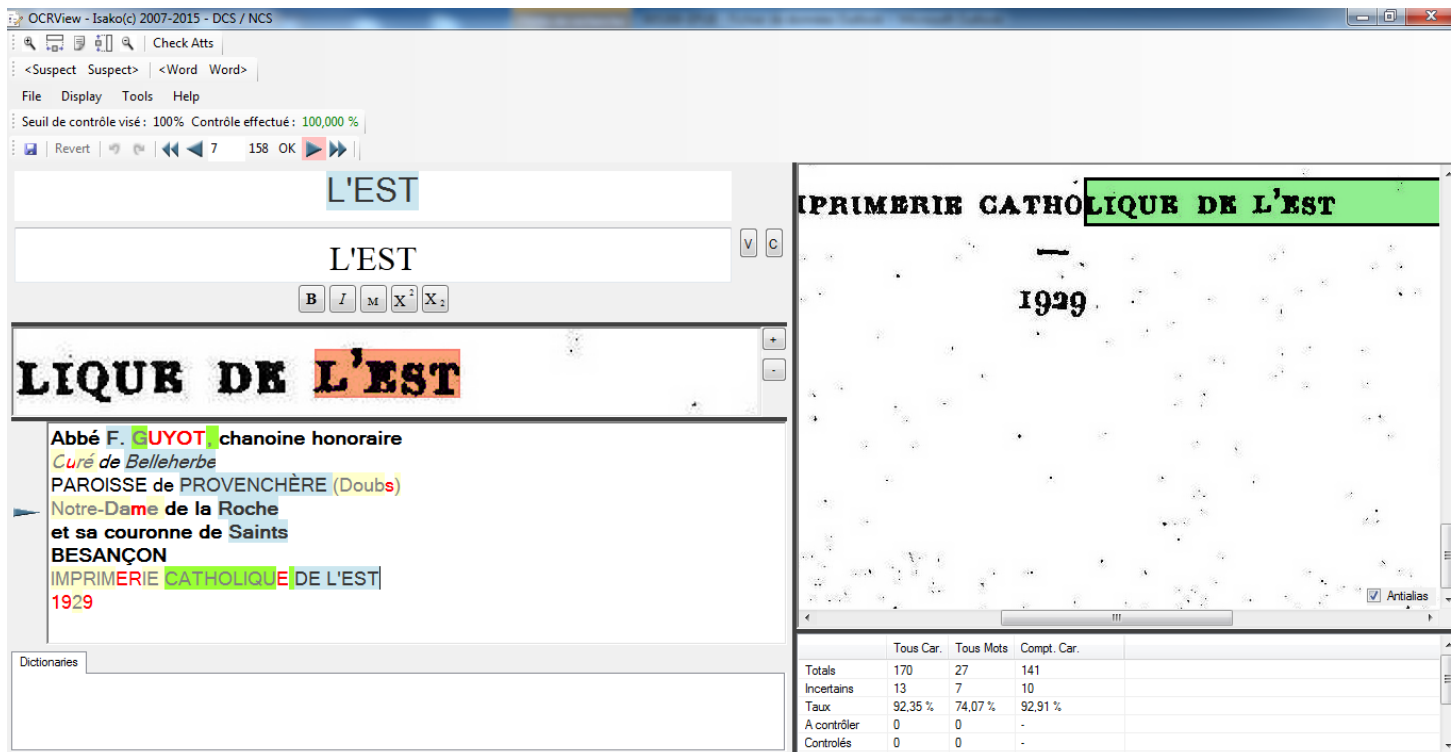
[IL], [II], [il], [ii], [11], [1I], [1i], [Y], [1I], [m], [nn], [ri], [cl], [<char-alpha>']

RegEx-B

[e], [c]

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01
		Page 5 sur 7

Il faut valider (bouton «v») tous les mots en rouge, même s'ils ne sont pas erronés pour le marquage que c'est déjà vu.



	Tous Car.	Tous Mots	Compt. Car.
Totals	170	27	141
Incertains	13	7	10
Taux	92,35 %	74,07 %	92,91 %
A contrôler	0	0	-
Contrôlés	0	0	-

- En cas de saut de lignes, on peut les rajouter à partir du menu ci-dessous
- Erreur segmentation, possibilité de correction dans :


Menu : Tools/Graphic Editor

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01
		Page 6 sur 7



- Description des codes couleurs :

Menu : Display/Caption

	CONSIGNES DE TRAITEMENT	Réf : CO_PRD_5309
		Date : 11/11/2014
	BNF - OCR Etape : OcrView	Ver : 01 Page 7 sur 7

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	