



Mise en pratique avec AGORA-RETRO

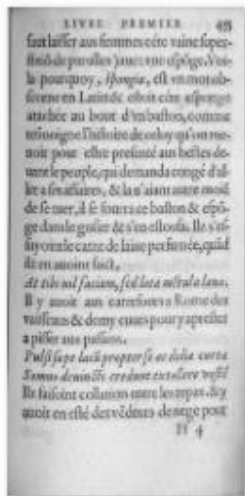


AGORA

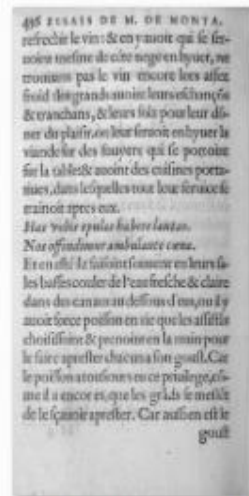
Regardez la vidéo de démonstration en même temps que vous avancez peut vous aider...

En route ...

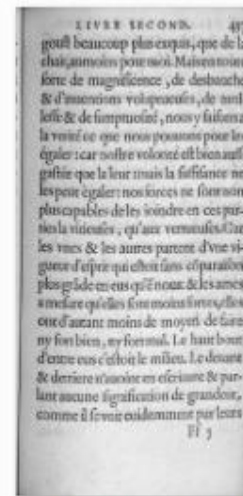
- Écriture d'un scénario d'extraction d'EdC (textuels) sur un ouvrage entier
 - Blocs
 - Lignes
 - Mots
 - Caractères (avec accents, points)
- Ouvrage = Quelques images d'un livre de Montaigne



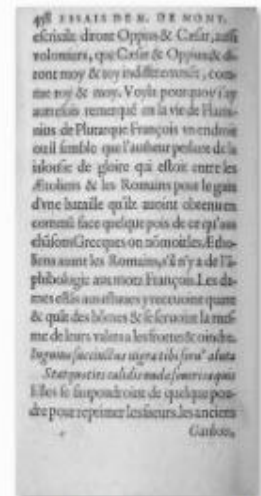
BRom_Montaigne1_0471.jpg



BRom_Montaigne1_0472.jpg



BRom_Montaigne1_0473.jpg



BRom_Montaigne1_0474.jpg

AGORA : interface du logiciel



The screenshot shows the AGORA software interface with the following components:

- Scénario**: A panel on the left containing a menu with options like "CreateLabel.Type=CC" and "FindCC.Parent=DOCUMENT,New=CC,BinarizationBasicThreshold=...". Below the menu is a log window showing technical information such as "Loading map image from cache OK.", "Building map image from loaded map image ...", "Computing min/max for normalisation...", "Computing min/max OK.", "Normalisation between 0 and 10 ...", "Normalisation OK.", and "Storing map ...".
- Image courante**: A large central panel displaying a map image. A red box highlights the top-left corner of the image.
- Arbre d'EOC**: A panel on the right showing a list of EOC (Elementary Object of Context) items. Each item is represented by a colored square (black for NONAME, red for DOCUMENT, green for CC) followed by its coordinates and dimensions, e.g., "CC {X=894,Y=314,Width=9,Height=9}{w=1,[894,0]314,..."
- Labels existants**: A panel at the bottom right showing a table of existing labels. The table has two columns: "Label" and "Count".

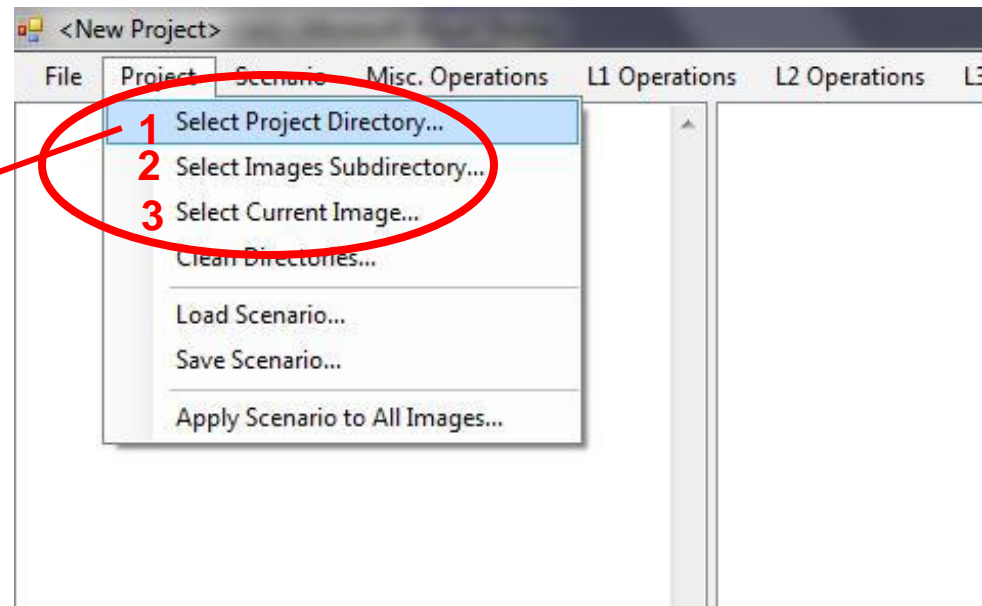
Label	Count
NONAME	
DOCUMENT	
CC	

DEMARRAGE : choix du dossier de projet

- Un projet AGORA → un dossier
 - + 1 sous-dossier des images à segmenter
 - + 1 image « type »
 - Projet.xml, Scenario.xml
 - Sous-dossiers des résultats
 - Alto
 - Etc.

Data_Agora/Test_Montaigne

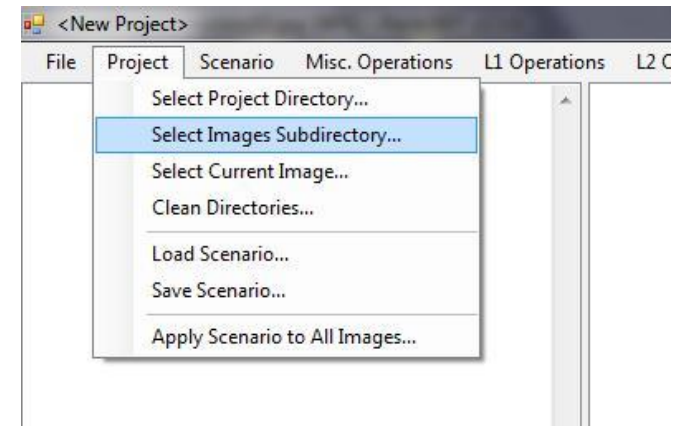
→ A nettoyer : Supprimer tous les répertoires **sauf originalimages**



DEMARRAGE (2)

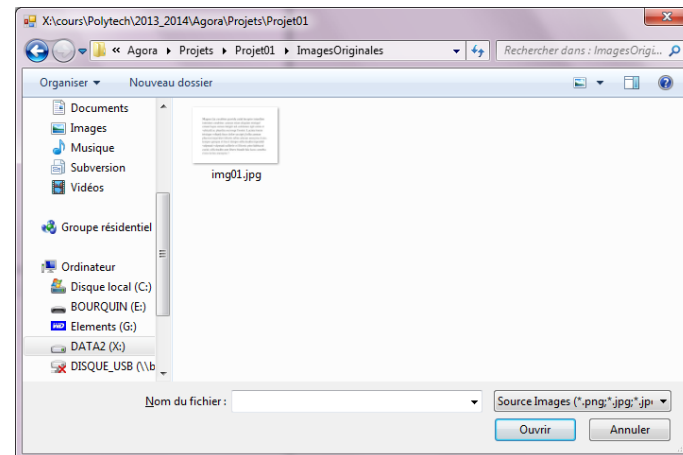
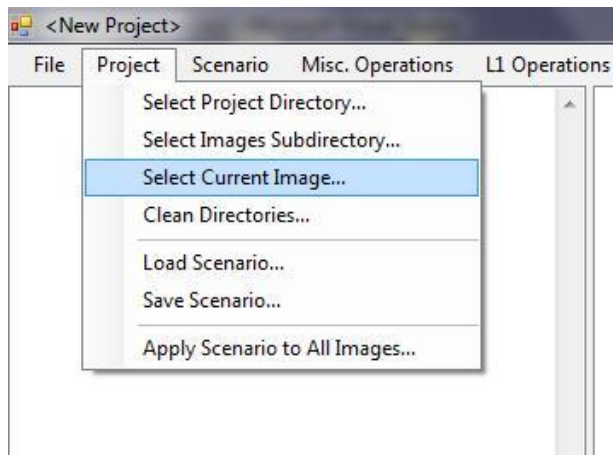
- Choix du sous-dossier des images

OriginalImages



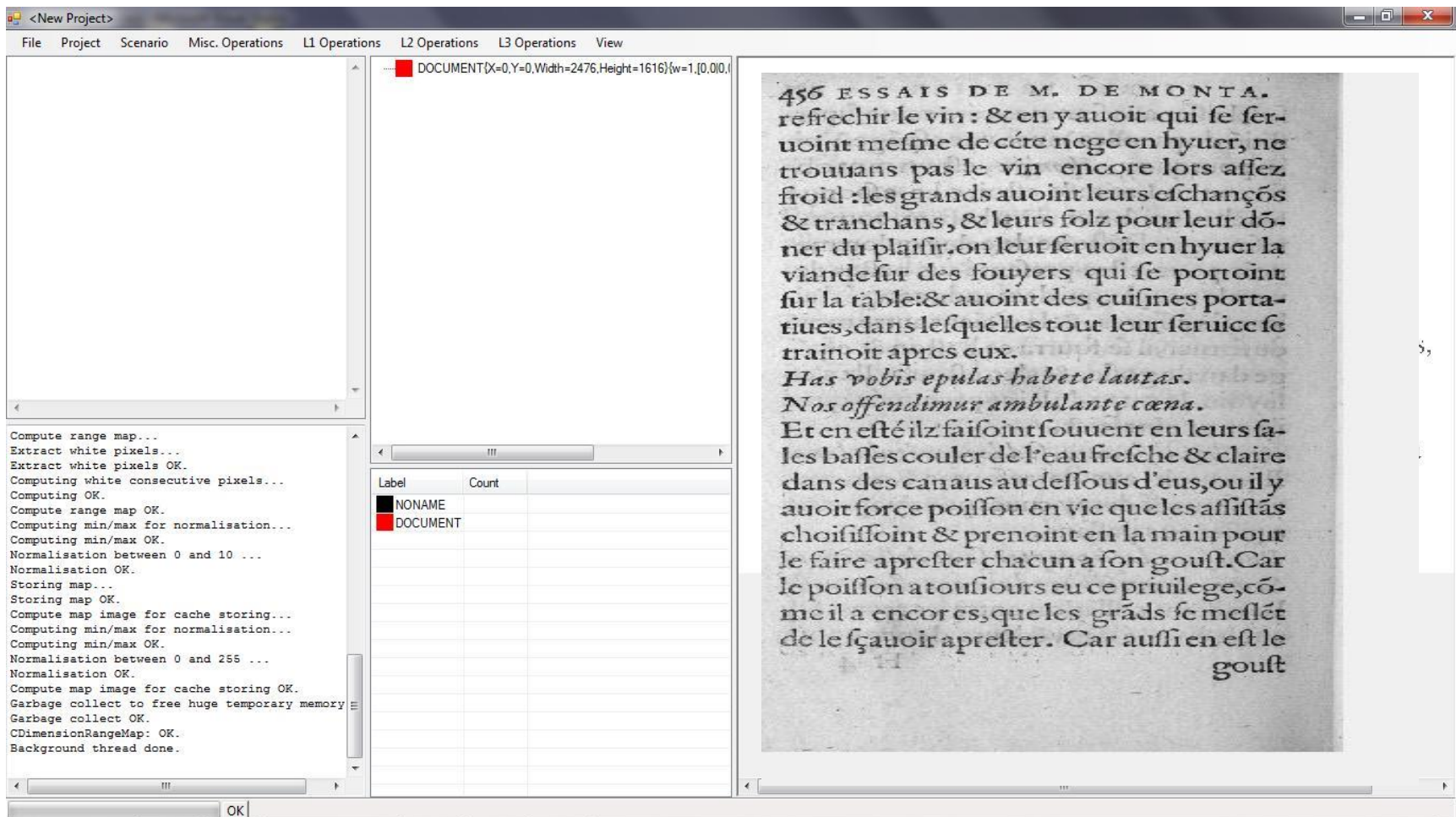
- Choix de l'image courante

BRom_Montaigne1_0471



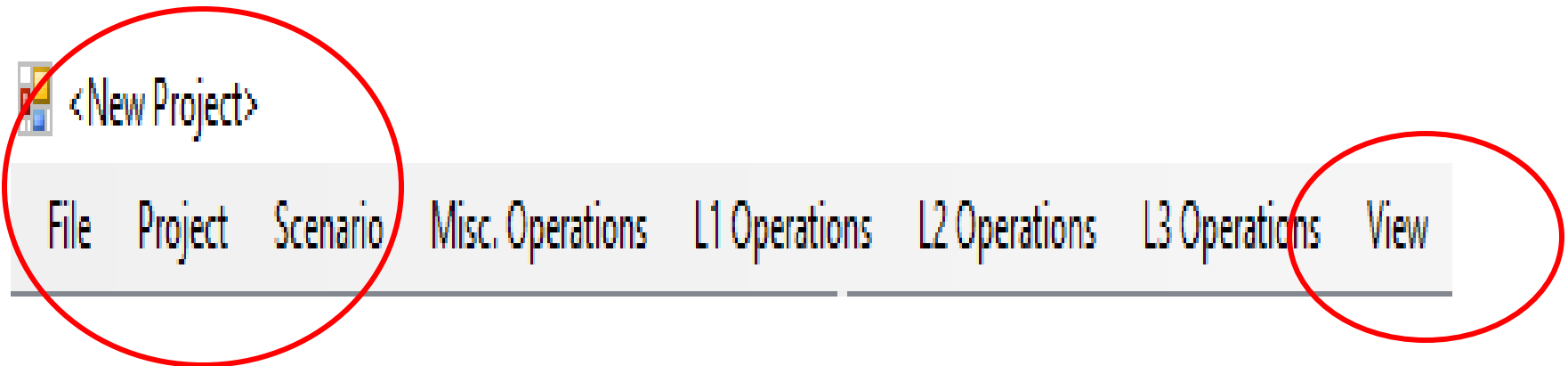
DEMARRAGE (3)

- Suite aux choix précédents, on obtient l'état initial



Prise en main IHM et autre

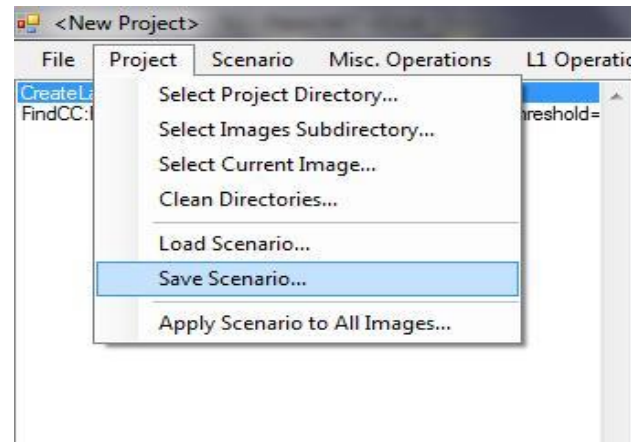
■ IHM



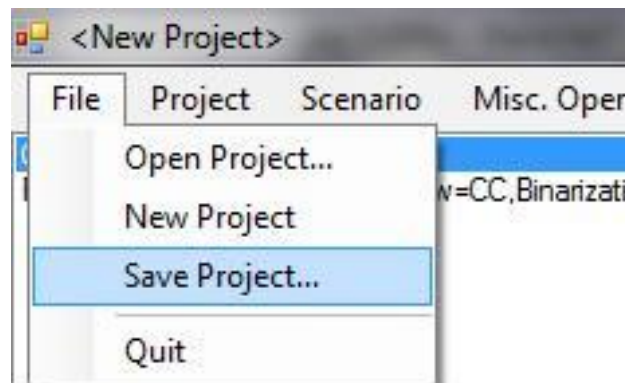
- UNDO !!! → Scenario>> Undo Last Operation

Sauvegarde

■ Sauver le scénario



■ Sauver le projet



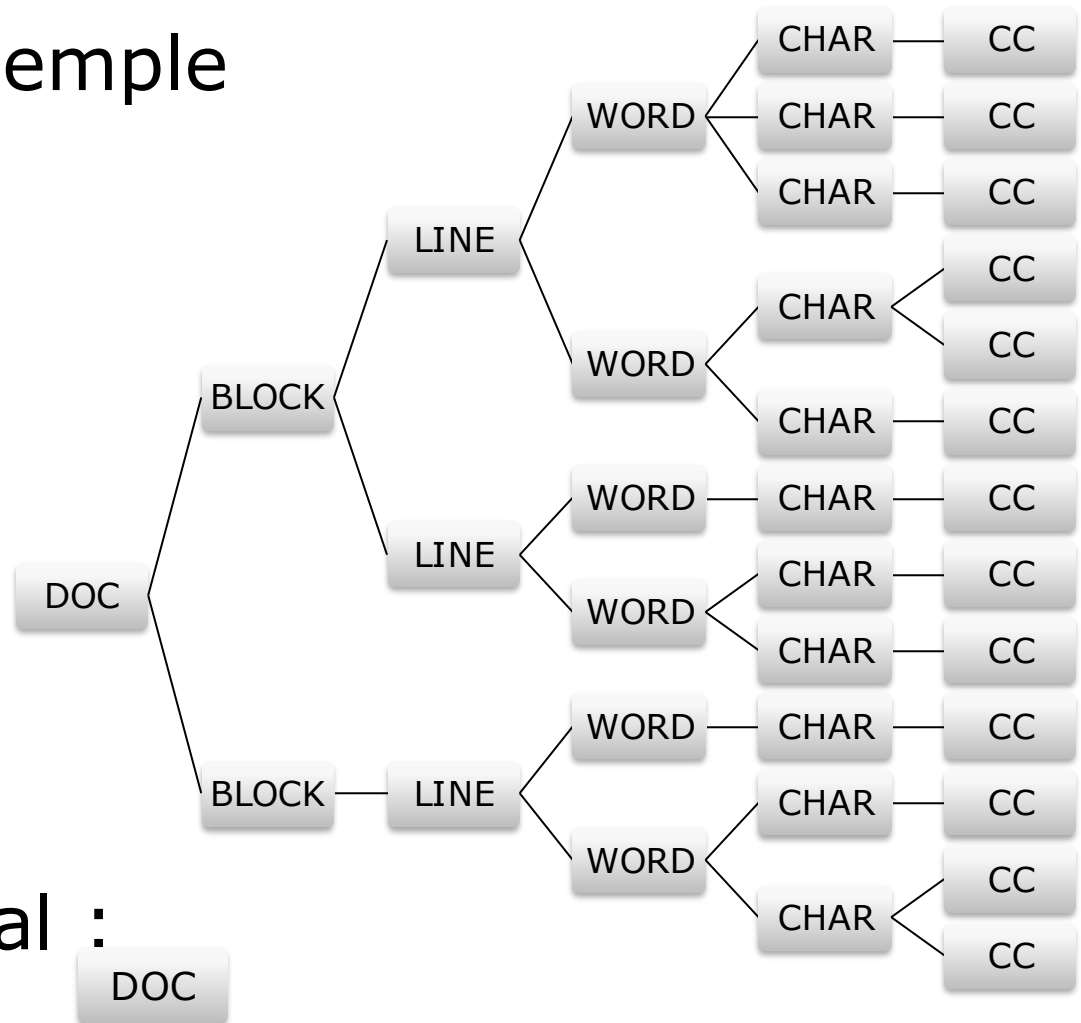
Analyse interactive d'images

- AGORA n'est pas automatique !
 - Interaction forte avec l'utilisateur
 - Construction d'un scénario à partir d'une image
 - Application de ce scénario à toutes les autres images
- Scénario
 - Ensemble d'étapes permettant de construire l'arbre d'EOC
- Etape de scénario
 - Une étape = une opération sur l'arbre d'EOC
 - Création d'EOC
 - Expansion
 - Insertion
 - Suppression d'EOC
 - Modification d'EOC
 - Chaque étape est paramétrable

Analyse interactive d'images

■ Arbre d'EOC : exemple

- Père
- Fils



■ Arbre d'EOC initial :



Analyse interactive d'images

■ EOC

- Possède un label (« type »)
 - NOISE, CHAR, WORD, LINE, etc.
- Possède une liste de caractéristiques (« features »)
 - Actuellement dans AGORA : 5 features
 - **X, Y → Dimensions ou Position horizontale et verticale**
 - RangeMapX, RangeMapY → Dimensions dans la carte des plages
 - ChildNumber → Nombre de fils
 - Chaque feature est un segment
 - 2 nombres « a » et « b » par feature



Creation de labels pour les EoC

Création d'un nouveau type d'EOC (CC, NOISE, CHAR,...)

- Ajout d'EOC de type « CC »

Misc Operations>>Label Creation>> CC

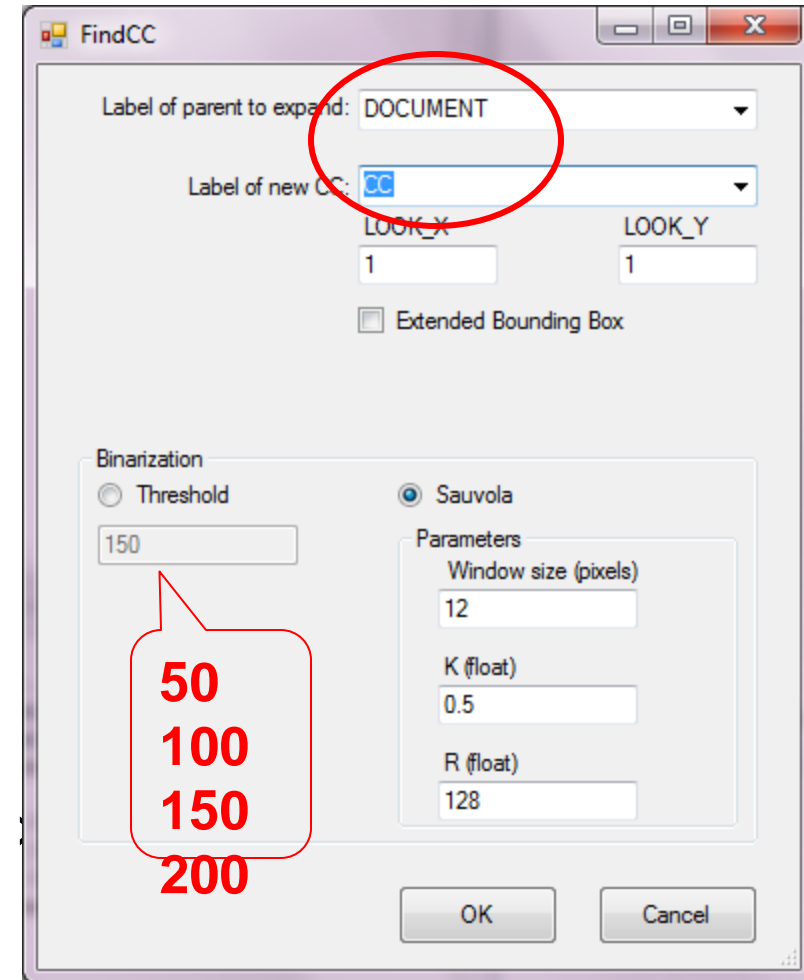
- Label Creation >> Noise

Misc Operations>>Label Creation>> Noise

Etc...

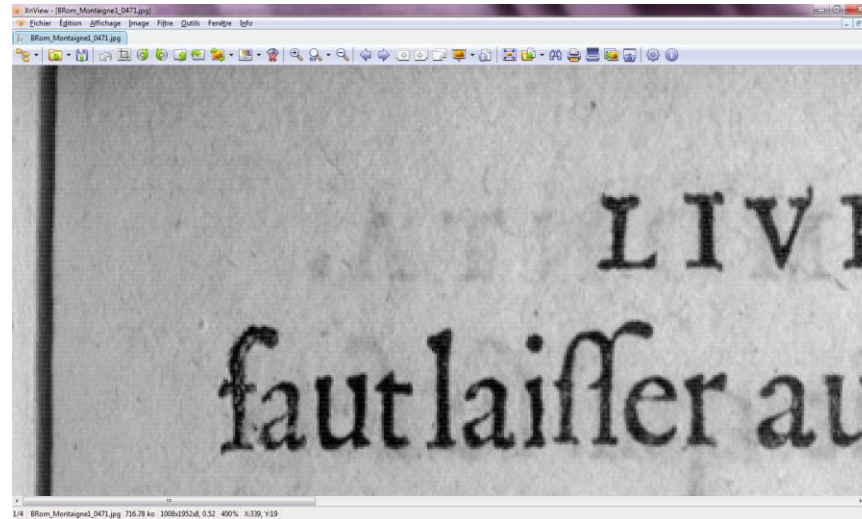
Extraction des EOC

- Extraction des composantes connexes
- L1 operations>>Find CC
- Choix de la binarisation
 - ☐ Binarisation simple
 - Seuil = 150
 - ☐ Binarisation de Sauvola
 - Possible mais lente pour des grandes pages (> 2 Mpixels)
- Démarche par « Essai-Erreur »
- Scenario>>Undo Last

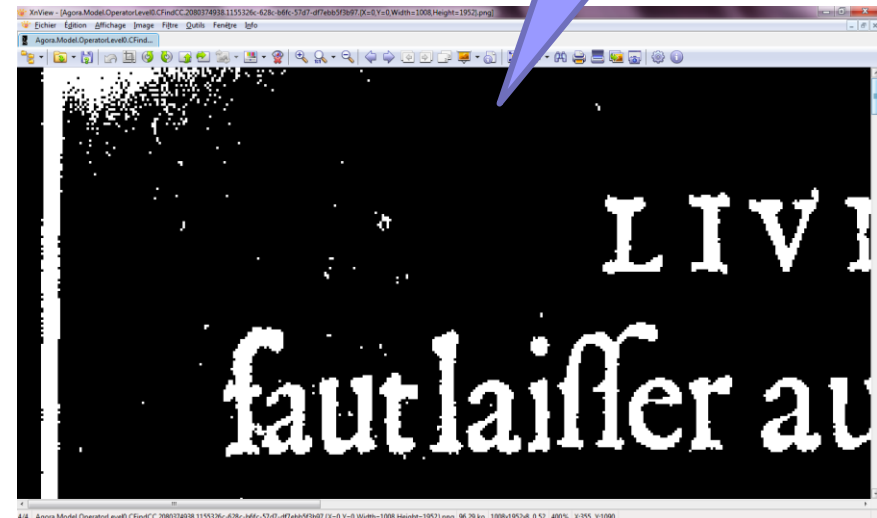


CONTRÔLE VISUEL de la binarisation : Essais-erreurs + voir repertoire « _CACHE »

Sauvola



Seuil



Extraction EOC : résultat

[illegible]

Résultat + IHM

Labels List:

Label	Count
NONAME	
DOCUMENT	
CC	

Label Details (from tree view):

- DOCUMENT {X=0,Y=0,Width=2476,Height=1616}{w=1,[0,0]}
- CC {X=894,Y=314,Width=9,Height=9}{w=1,[894,0]314,1}
- CC {X=1178,Y=314,Width=9,Height=9}{w=1,[1178,0]314,1}
- CC {X=1197,Y=314,Width=39,Height=58}{w=1,[1197,0]}
- CC {X=1450,Y=314,Width=19,Height=15}{w=1,[1450,0]}
- CC {X=1502,Y=314,Width=9,Height=9}{w=1,[1502,0]314,1}
- CC {X=1801,Y=314,Width=9,Height=9}{w=1,[1801,0]314,1}
- CC {X=1949,Y=314,Width=39,Height=58}{w=1,[1949,0]}
- CC {X=1996,Y=314,Width=18,Height=15}{w=1,[1996,0]}
- CC {X=1407,Y=315,Width=19,Height=14}{w=1,[1407,0]}
- CC {X=296,Y=316,Width=71,Height=55}{w=1,[296,0]316,1}
- CC {X=660,Y=323,Width=22,Height=49}{w=1,[660,0]32,1}
- CC {X=912,Y=323,Width=22,Height=49}{w=1,[912,0]32,1}
- CC {X=1678,Y=323,Width=22,Height=49}{w=1,[1678,0]}
- CC {X=1861,Y=323,Width=22,Height=49}{w=1,[1861,0]}
- CC {X=372,Y=333,Width=34,Height=39}{w=1,[372,0]33,1}
- CC {X=408,Y=333,Width=37,Height=56}{w=1,[408,0]33,1}
- CC {X=449,Y=333,Width=39,Height=38}{w=1,[449,0]33,1}

Log Console:

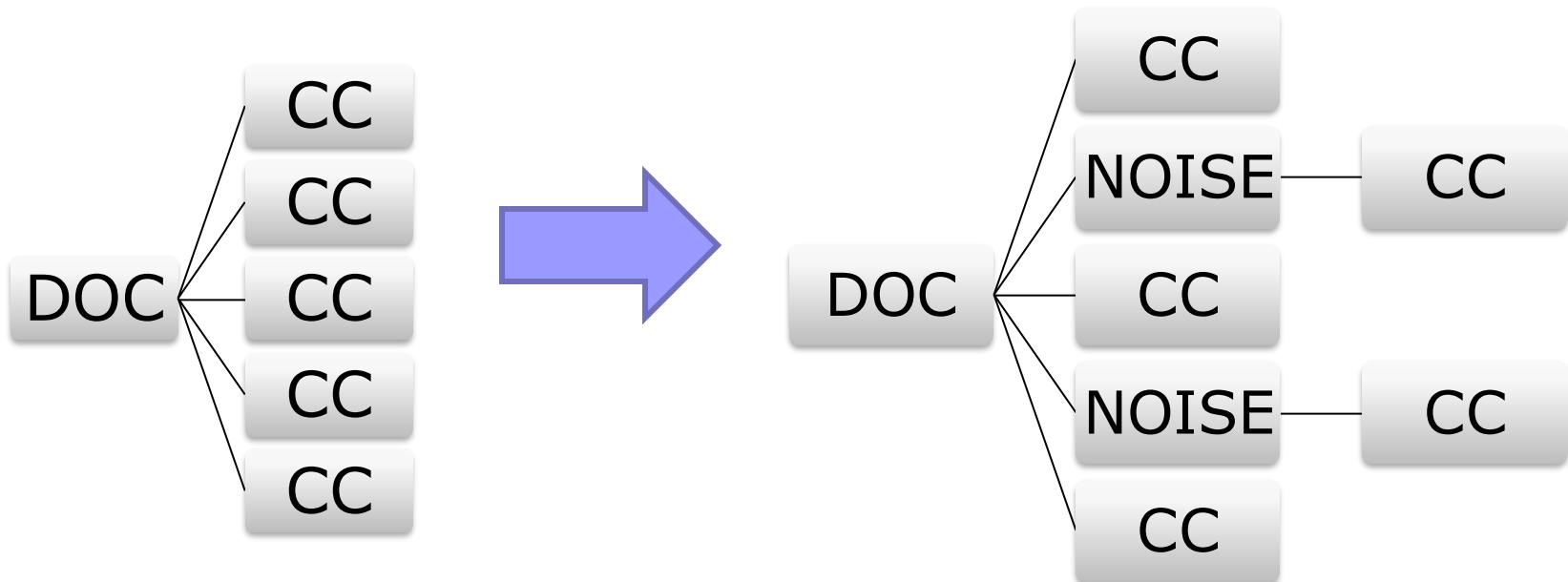
```
Background thread done.
CCreateLabel: Create label CC...
CCreateLabel: OK.
CCreateLabel: Create label CC...
CCreateLabel: OK.
Background thread started...
CFindCC: Find CC...
CFindCC: Expand...
Grayscale...
Grayscale OK.
Threshold Filter...
Threshold OK.
Invert...
Invert OK.
Detect CC...
Detect CC OK.
Adding 525 CC to graph...
Adding CC OK.
CFindCC: OK.
Results = 1
CFindCC: OK.
Background thread done.
```

Callout Box: CLIC sur un LABEL → Visualisation sur l'image

Analyse et Etiquetage des EoC

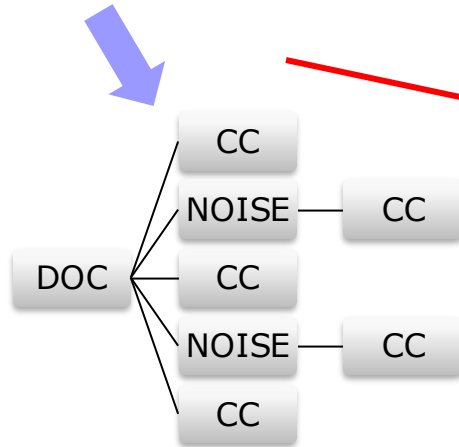
Elimination du bruit

- Bruit = composantes connexes très petites
 - Création d'un label « NOISE »
 - Classification des EOC « CC » en EOC « NOISE »
 - Classification selon les dimensions horizontales et verticales
 - Suppression des EOC « NOISE »



Classification par taille : « pattern »

DOC — CC



L2 >> Classify by Features

Classify By Features

Classify By: ☐ Position ☒ Size ☐ Form

Label of parent: DOCUMENT

Label of new EOC: NOISE

Label of child: CC

☐ Replace child

Features of child:

	a0	b0
▶ EUCLIDIAN_X	0	3
EUCLIDIAN_Y	0	3
RANGEMAP_X	0	0
RANGEMAP_Y	0	0
CHILDS_NUMBER	0	0

Coefficients:

0
1
1
0
0
0

Comparators:

Ch0
...
...
+Infini
+Infini
+Infini

Buttons: Add, Del, Reset, OK, Cancel

C'est ici que vous spécifiez les limites de taille (en X et Y) désirées pour le bruit

Bien mettre des 0 sur les critères inutilisés

Cliquez ici

EOC (a1,b1) < Pattern (a0,b0)

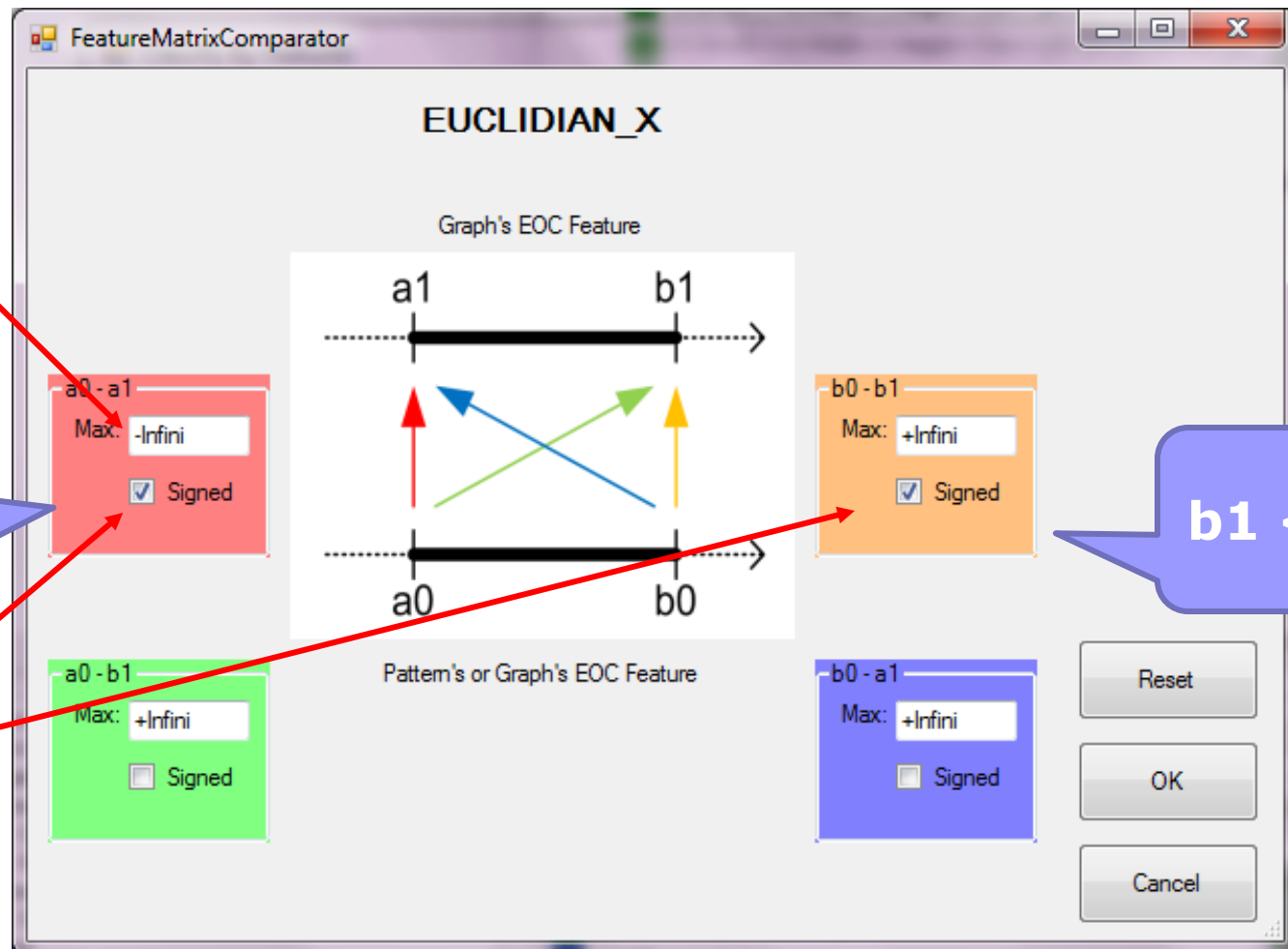
Interface difficile à assimiler surtout sans explication orale....

Il est possible de juste recopier les valeurs ci dessous sans chercher a comprendre...

Remplacer
+Infini par
-Infini

a0 < a1

Cochez



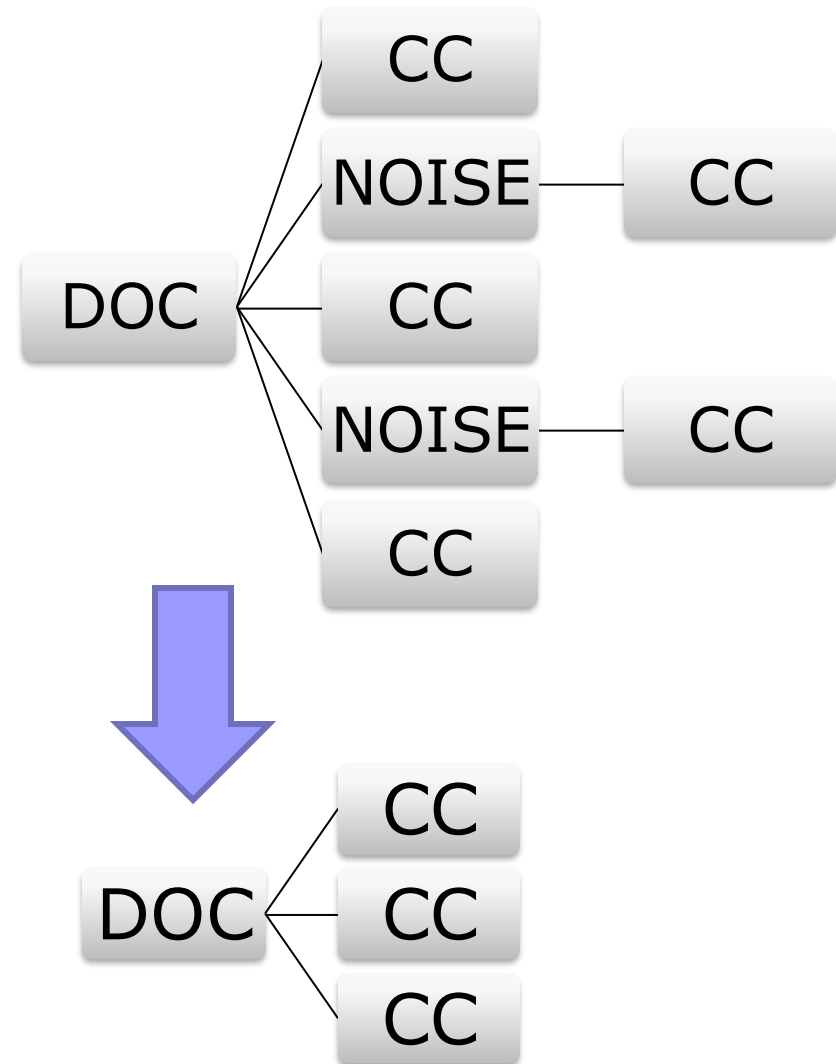
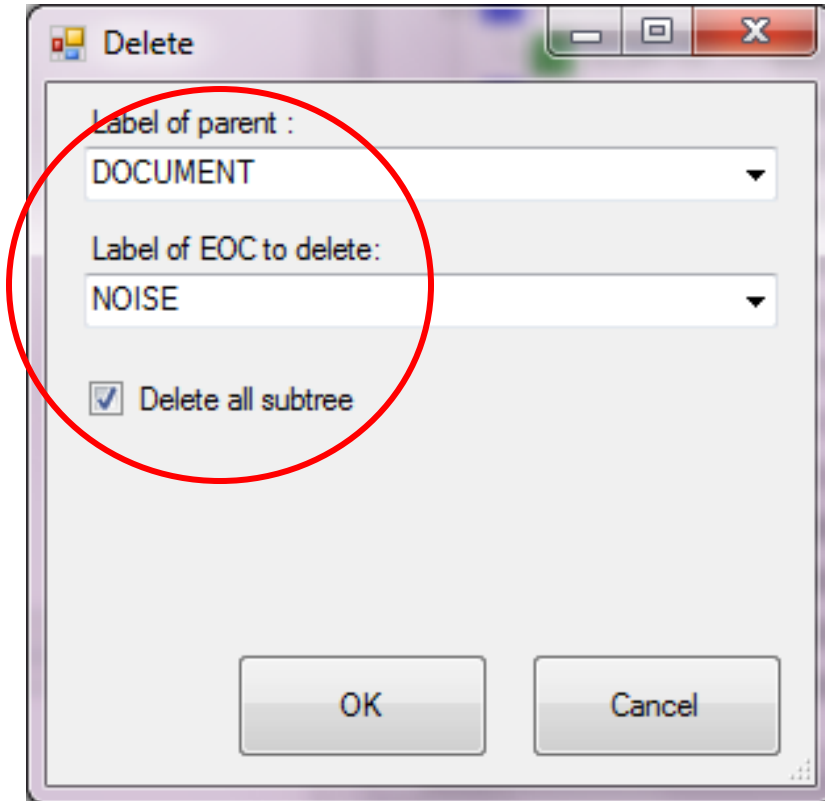
b1 < b0

Classification : résultat

The screenshot displays a software application window titled 'proj301.xml'. The interface is divided into several sections:

- Menu Bar:** File, Project, Scenario, Misc. Operations, L1 Operations, L2 Operations, L3 Operations, View.
- Left Sidebar:** Contains a tree view of the document structure. The root node is 'CreateLabel: Type=CC'. It has several children, including 'FindCC: Parent=DOCUMENT, New=CC, BinarizationBasicThreshold=150', 'CreateLabel: Type=NOISE', and 'Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=NOI'. The tree view shows a hierarchical structure of nodes, with some nodes expanded to show their children.
- Central Text Area:** Displays the document content, which is a page from a book. The text is in French and includes the title 'LIVRE PREMIER' and the beginning of a chapter. The text is: 'faut laisser aus femmes cete vair ititiō de parolles)auec vne espō la pourquoy, *spongia*, est vn scœne en Latin:& estoit cete c atachée au bout d'vn baston, tesmoigne l'histoire de celuy qu noit pour estre presenté aus b uant le peuple, qui demanda cor ler a ses affaires, & la n'ayant aut de se tuer, il se fourra ce baston'.
- Bottom Status Bar:** Shows the current status of the application, including 'CFindCC: OK.', 'Background thread done.', 'Background thread started...', 'CFindCC: Find CC...', 'CFindCC: Expand...', 'Detect CC...', 'Detect CC OK.', 'Adding 1164 CC to graph...', 'Adding CC OK.', 'CFindCC: OK.', 'Results = 1', 'CFindCC: OK.', 'Background thread done.', 'CCreateLabel: Create label NOISE...', 'CCreateLabel: OK.', 'Background thread started...', 'CClassifyByFeatures: Classify...', 'CInsert: NOISE <- { CC }', '100', '200', '300', 'Results = 302', 'CInsert: OK.', 'CClassifyByFeatures: OK.', 'Background thread done.'

Suppression des EOC « NOISE »



Detection / Etiquetage d'autres types d'EoC

- Classification par la taille (idem NOISE)
 - **Caractères : taille** spécifiée
 - Taille X < 80
 - Taille Y < 80
 - **Accents** : caractères de taille « petite »
 - Taille X < 25
 - Taille Y < 20
 - **Images** : tout le reste 😊

Caractères : résultat

proj301.xml

Classify By Features

Classify By...
☐ Position
☒ Size
☐ Form

Label of parent: DOCUMENT
 Label of new EOC: CHAR
☐ Replace child
 Label of child: CC

Features of child:

	a0	b0
EUCLIDIAN_X	0	80
EUCLIDIAN_Y	0	80
RANGEMAP_X	0	0
RANGEMAP_Y	0	0
CHILDS_NUMBER	0	0

Coefficients:

0
1
1
1
1

Comparators:

Ch0
...
...
+Infini
+Infini
+Infini

Reset
OK
Cancel

Add Del

Results = 854
 CInsert: OK.
 CClassifyByFeatures: OK.
 Background thread done.
 Background thread started...
 CClassifyByFeatures: Classify...
 CInsert: CHAR <- { CC }
 100
 200
 300
 400
 500
 600
 700
 800
 Results = 861
 CInsert: OK.
 CClassifyByFeatures: OK.
 Background thread done.

Label Count

Label	Count
NONAME	
DOCUMENT	
CC	
NOISE	
CHAR	

X=302 Y=1077

ACCENTS : résultat

testjyyy.xml

File Project Scenario Misc. Operations L1 Operations L2 Operations L3 Operations View

CreateLabel: Type=CC
FindCC: Parent=DOCUMENT, New=CC, BinarizationBasicThreshold=150, BinarizationS
CreateLabel: Type=NOISE
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=NOISE, Replace=
Delete: Parent=DOCUMENT, Child=NOISE, DeleteAll=True, Loop=True
CreateLabel: Type=CHAR
CreateLabel: Type=IMAGE
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=CHAR, Replace=f
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=IMAGE, Replace=
CreateLabel: Type=LINE
CreateLabel: Type=ACCENT
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CHAR, New=ACCENT, Rep

#####CDelete: Delete child OK.
CCreateLabel: Create label CHAR...
CCreateLabel: OK.
CCreateLabel: Create label IMAGE...
CCreateLabel: OK.
Background thread started...
CClassifyByFeatures: Classify...
#####CClassifyByFeatures: OK.
Background thread done.
Background thread started...
CClassifyByFeatures: Classify...
#####CClassifyByFeatures: OK.
Background thread done.
Background thread started...
CClassifyByFeatures: Classify...
CClassifyByFeatures: OK.
Background thread done.
CCreateLabel: Create label LINE...
CCreateLabel: OK.
CCreateLabel: Create label ACCENT...
CCreateLabel: OK.
Background thread started...
CClassifyByFeatures: Classify...
#CClassifyByFeatures: OK.
Background thread done.

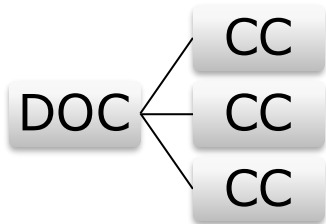
Label Count

NONAME	
DOCUMENT	
CC	
NOISE	
CHAR	
IMAGE	
LINE	
ACCENT	

X=222 Y=478

la pourquoy, s
scoene en Latin: &
atachée au bout
tesinoigne l'histoir
noit pour estre p

IMAGES : « tout le reste »



Classify By Features

Classify By...
☐ Position
☒ Size
☐ Form

Label of parent: DOCUMENT
 Label of new EOC: IMAGE
 Label of child: CC

☐ Replace child

Features of child:

	a0	b0
▶ EUCLIDIAN_X	0	0
EUCLIDIAN_Y	0	0
RANGEMAP_X	0	0
RANGEMAP_Y	0	0
CHILDS_NUMBER	0	0

Coefficients:

0
0
0
0
0
0

Comparators:

Ch0
+Infini
+Infini
+Infini
+Infini
+Infini

Add Del

Reset OK Cancel

Toutes les CC
ayant encore Doc
comme père dans
l'arbre deviennent
Image

→ Aucun critère
sélectionné

IMAGES : résultat

The screenshot shows a software application window titled "proj301.xml". The interface is divided into several sections:

- Menu Bar:** File, Project, Scenario, Misc. Operations, L1 Operations, L2 Operations, L3 Operations, View.
- Left Sidebar:** Contains a tree view of document elements. The selected element is "IMAGE(X=0,Y=0,Width=20,Height=1524){w=1,[0.010.010.0]}", which is highlighted in pink. Other elements include "CHAR" and "CC" objects with various coordinates and dimensions.
- Bottom Status Bar:** Displays the coordinates "X=302 Y=1077".
- Main Text Area:** Displays a scanned page of text from a French manuscript. The text is in French and includes Latin phrases. The page is numbered "455" in the top right corner.

The text in the main area is as follows:

LIVRE PREMIER 455
 faut laisser aus femmes cete vaine super-
 flitiō de parolles) avec vne espōge. Voi-
 la pourquoy, *spongia*, est vn mot ob-
 scene en Latin: & estoit cete esponge
 atachée au bout d'un baston, comme
 tesmoigne l'histoire de celuy qu'on me-
 noit pour estre presenté aus bestes de-
 uant le peuple, qui demanda congé d'al-
 ler a ses affaires, & la n'ayant autre moié
 de se tuer, il se fourra ce baston & espō-
 ge dans le gosier & s'en estoufa. Ilz s'es-
 fuyoint le catze de laine perfumée, quād
 ilz en auoint fait,
At tibi nil faciam, sed lota mētula lana.
 Il y auoit aus carrefours a Rome des
 vaisseaus & demy cuues pour y aprester
 a pisser aus passans.
Pulsi saepe lacū propter se ac dolia curta
Somno deuincti credunt extollere vestē
 Ilz faisoient collation entre les repas. & y
 auoit en esté des vèdeurs de nege pour
 Ff 4

Classification par la position

Classify By Features

Classify By...
☐ Position
☒ Size
☐ Form

Label of parent: DOCUMENT
 Label of new EDC: NOISE
 Label of child: CC

☐ Replace child

Features of child:

	a0	b0
EUCLIDIAN_X	0	1
EUCLIDIAN_Y	3	1
RANGEMAP_X	0	0
RANGEMAP_Y	0	0
HILL NUMBER	0	0

Efficients:

	0
	1
	0
	0
	0

Comparators:

Ch0
...
...
+Infini
+Infini
+Infini

Add Del

Reset OK Cancel

IDEM TAILLE

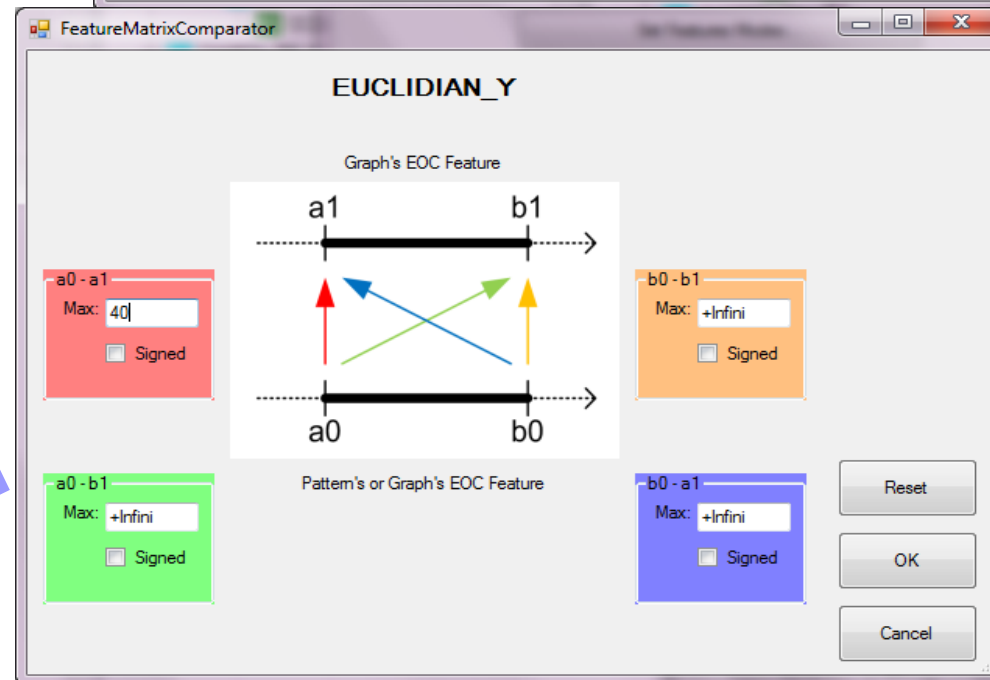
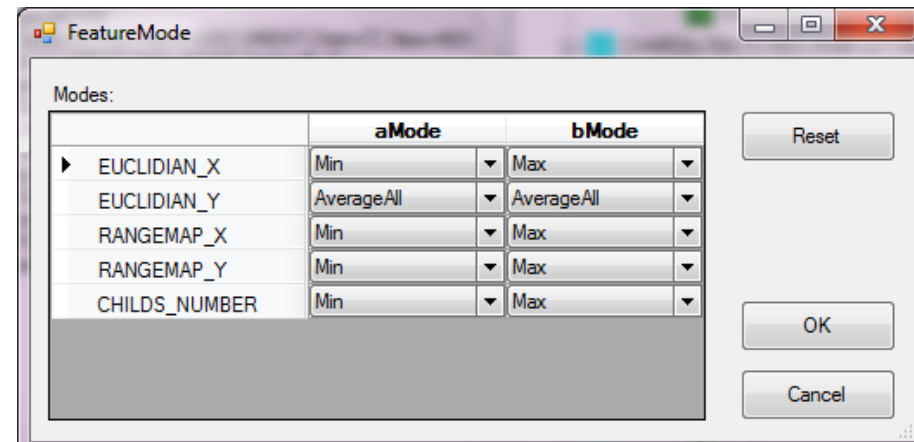
Rassemblement / Fusion d'EoC

Création de lignes de texte

- Création d'un label « LINE »
- CHAR → LINE
- Un EOC LINE aura comme Y = moyenne des Y fils
 - Set Feature mode

- Fusion Line/CHAR → Line
 - Contexte = Doc
 - LINE = LINE + LINE si alignement Y satisfaisant

- Simplification finale
(delete line dans line)



Lignes : résultat

The screenshot displays the 'proj301.xml' application window. The interface is divided into several sections:

- Menu Bar:** File, Project, Scenario, Misc. Operations, L1 Operations, L2 Operations, L3 Operations, View.
- Command Log (Left):** Shows a series of commands and their results, including 'CreateLabel: Type=CC', 'FindCC: Parent=DOCUMENT, New=CC, BinarizationBasic Threshold=150', 'Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=NOI', 'Delete: Parent=DOCUMENT, Child=NOISE, DeleteAll=True', 'CreateLabel: Type=CHAR', 'Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=CHA', 'CreateLabel: Type=IMAGE', 'Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=IMA', 'CreateLabel: Type=ACCENT', 'Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CHAR, New=/', 'SetFeatures: Parent=DOCUMENT, Child=ACCENT, Mode={a=Average, /', 'Insert: Parent=DOCUMENT, Childs={CHAR, ACCENT, }, New=CHAR, Fea', 'CreateLabel: Type=LINE', and 'Fusion: Parent=DOCUMENT, Childs=CHAR, New=LINE, Comparator={0='. It also shows 'Results = 635' and 'Results = 1270'.
- Tree View (Center):** A hierarchical list of detected elements with their bounding boxes and labels. For example, 'CHAR(X=88,Y=1515,Width=10,Height=48){w=1,[88.0]', 'CC(X=88,Y=1515,Width=10,Height=48){w=1,[88.0]', 'CHAR(X=135,Y=1516,Width=25,Height=50){w=1,[135]', 'CC(X=135,Y=1516,Width=25,Height=50){w=1,[135]', etc.
- Legend (Bottom Center):** A table with two columns: 'Label' and 'Count'. The labels are NONAME, DOCUMENT, CC, NOISE, CHAR, IMAGE, ACCENT, and LINE.
- Document Image (Right):** A scan of a page from a book. The text is in French and is enclosed in red rectangular bounding boxes. The text reads: 'LIVRE PREMIER 455', 'faut laisser aus femmes cete vaine super-', 'fittio de parolles)avec vne espōge. Voi-', 'la pourquoy, *spongia*, est vn mot ob-', 'scoene en Latin:& estoit cete esponge', 'atachée au bout d'vn baston, comme', 'tesmoigne l'histoire de celuy qu'on me-', 'noit pour estre presenté aus bestes de-', 'uant le peuple, qui demanda congé d'al-', 'ler a ses affaires, & la n'ayant autre moie', 'de se tuer, il se fourra ce baston & espō-', 'ge dans le gosier & s'en estoufa. Ilz s'es-', 'fuyoint le catze de laine perfumée, quād', 'ilz en auoint faict,'.

Rassemblement / Fusion d'EoC

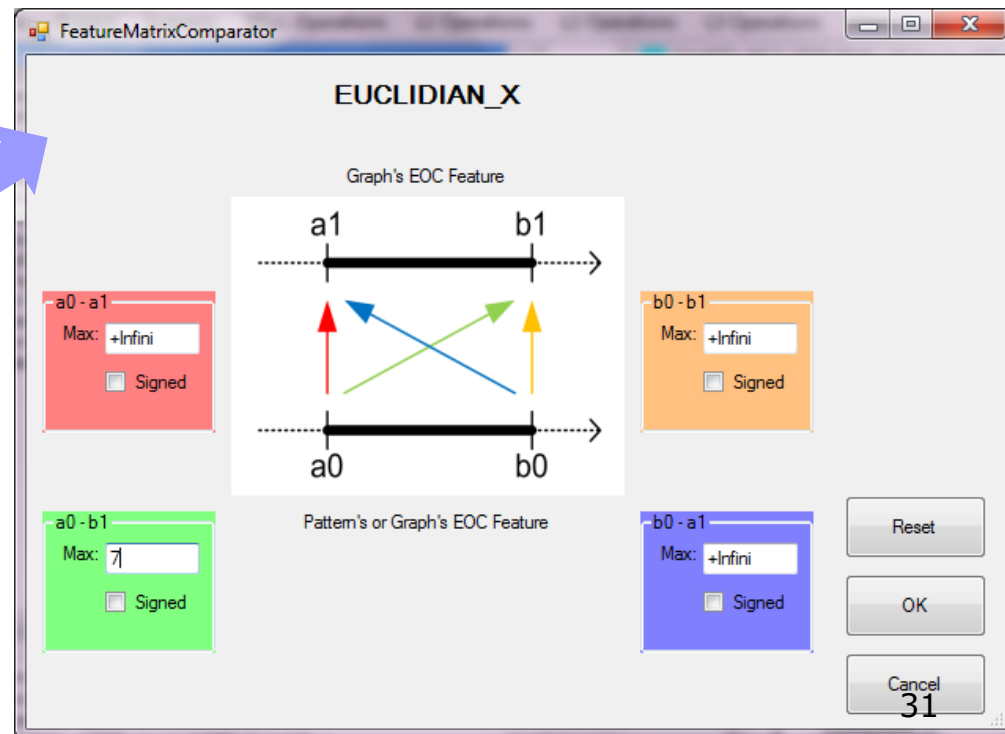
Construction des MOTS

Similaire LINE mais :

- Contexte = LINE
- On fusionne les CHAR qui sont proches en X (7 pixels)

Les Etapes :

- CHAR \rightarrow WORD
- Fusion WORD = WORD+WORD
si X satisfaisant
- Intersection entre WORD
inters. en X \rightarrow 1 seul WORD
- Simplification finale



MOTS : intersection + simplification

Intersection

Label of parent: LINE

Label of new EOC: WORD

Labels of childs: WORD WORD

Coefficients:

	0
▶ EUCLIDIAN_X	1
EUCLIDIAN_Y	0
RANGEMAP_X	0
RANGEMAP_Y	0
CHILDS_NUMBER	0

Reset

OK

Cancel

Delete

Label of parent : WORD

Label of EOC to delete: WORD

☐ Delete all subtree

OK

Cancel

Intersection + simplification : résultat

proj301.xml

File Project Scenario Misc. Operations L1 Operations L2 Operations L3 Operations View

CreateLabel: Type=CC
FindCC: Parent=DOCUMENT, New=CC, BinarizationBasicThreshold=150
CreateLabel: Type=NOISE
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=NOI
Delete: Parent=DOCUMENT, Child=NOISE, DeleteAll=True
CreateLabel: Type=CHAR
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=CHAR
CreateLabel: Type=IMAGE
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=IMA
CreateLabel: Type=ACCENT
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CHAR, New=ACCENT
SetFeatures: Parent=DOCUMENT, Child=ACCENT, Mode={a=Average, f=Intersection, Parent=DOCUMENT, Childs={CHAR, ACCENT, }, New=CHAR, Feat
CreateLabel: Type=LINE
Fusion: Parent=DOCUMENT, Childs=CHAR, New=LINE, Comparator=0
CreateLabel: Type=WORD
Fusion: Parent=LINE, Childs=CHAR, New=WORD, Comparator=0
Delete: Parent=WORD, Child=WORD, DeleteAll=False

CFusion: Phase 3 : Simplification...
CDelete: Delete child WORD...
100
200
300
400
500
600
700
800
Results = 878
CDelete: OK.
CFusion: OK.
Background thread done.
Background thread started...
CIntersect: ChildsEOC -> NewEOC...
CInsert: WORD <- { WORD + WORD }
Results = 33
Results = 33
CInsert: OK.
CIntersect: OK.
Background thread done.
CDelete: Delete child WORD...
Results = 66
CDelete: OK.

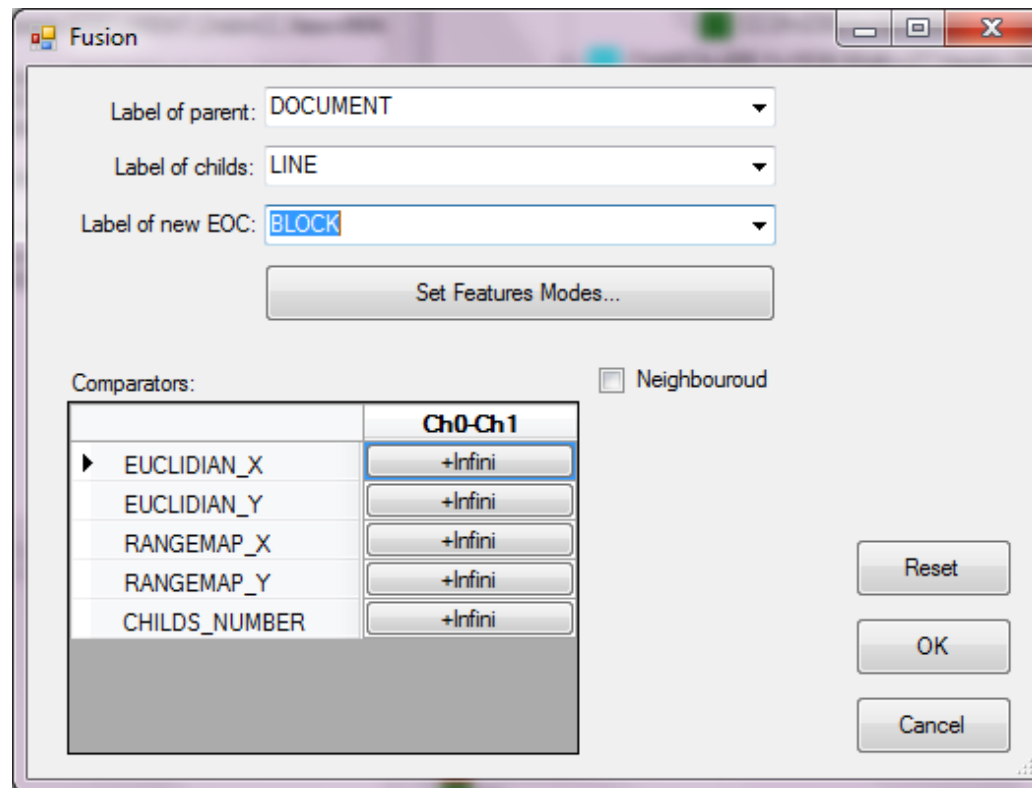
Label Count

Label	Count
NONAME	
DOCUMENT	
CC	
NOISE	
CHAR	
IMAGE	
ACCENT	
LINE	
WORD	

X=764 Y=335

Idem pour les blocs de texte

- Rassemblement de toutes les lignes
- Fusion LINE → BLOCK
 - Pas de critère



Blocs de texte : résultat

proj301.xml

File Project Scenario Misc. Operations L1 Operations L2 Operations L3 Operations View

CreateLabel: Type=CC
FindCC: Parent=DOCUMENT, New=CC, BinarizationBasicThreshold=150
CreateLabel: Type=NOISE
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=NOI
Delete: Parent=DOCUMENT, Child=NOISE, DeleteAll=True
CreateLabel: Type=CHAR
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=CHA
CreateLabel: Type=IMAGE
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CC, New=IMA
CreateLabel: Type=ACCENT
Classify: Type=ClassifyBySize, Parent=DOCUMENT, Child=CHAR, New=/
SetFeatures: Parent=DOCUMENT, Child=ACCENT, Mode={a=Average/
Insert: Parent=DOCUMENT, Childs={CHAR, ACCENT, }, New=CHAR, Fea
CreateLabel: Type=LINE
Fusion: Parent=DOCUMENT, Childs=CHAR, New=LINE, Comparator={0=
CreateLabel: Type=WORD
Fusion: Parent=LINE, Childs=CHAR, New=WORD, Comparator={0={Fals
Intersection: Parent=LINE, Childs={WORD, WORD, }, New=WORD, Coeff
Delete: Parent=WORD, Child=WORD, DeleteAll=False
CreateLabel: Type=BLOCK
Fusion: Parent=DOCUMENT, Childs=LINE, New=BLOCK, Comparator={0=

CDelete: Delete child WORD...
Results = 66
CDelete: OK.
CCreateLabel: Create label BLOCK...
CCreateLabel: OK.
Background thread started...
CFusion: Phase 1 : ChildsEOC -> NewEOC...
CInsert: BLOCK <- { LINE }
Results = 23
Results = 23
CInsert: OK.
CFusion: Phase 2 : Fusion...
CInsert: BLOCK <- { BLOCK + BLOCK }
Results = 21
Results = 22
CInsert: OK.
CInsert: BLOCK <- { BLOCK + BLOCK }
CInsert: OK.
CFusion: Phase 3 : Simplification...
CDelete: Delete child BLOCK...
Results = 44
CDelete: OK.
CFusion: OK.
Background thread done.

CHAR{X=280,Y=128,Width=18,Height=31}{w-
CC{X=280,Y=128,Width=18,Height=31}{w-
CHAR{X=255,Y=130,Width=22,Height=29}{w-
CC{X=255,Y=130,Width=22,Height=29}{w-
CHAR{X=226,Y=107,Width=41,Height=51}{w-
CC{X=226,Y=107,Width=41,Height=51}{w-
CHAR{X=214,Y=112,Width=8,Height=47}{w-
CHAR{X=214,Y=130,Width=8,Height=29}{
CC{X=214,Y=130,Width=8,Height=29}{
ACCENT{X=214,Y=112,Width=7,Height=8
CHAR{X=214,Y=112,Width=7,Height=
CC{X=214,Y=112,Width=7,Height=
CHAR{X=188,Y=130,Width=22,Height=30}{w-
CC{X=188,Y=130,Width=22,Height=30}{w-
CHAR{X=173,Y=110,Width=11,Height=50}{w-
CC{X=173,Y=110,Width=11,Height=50}{w-
CHAR{X=152,Y=128,Width=15,Height=32}{w-
CC{X=152,Y=128,Width=15,Height=32}{w-
CHAR{X=126,Y=128,Width=23,Height=32}{w-
CC{X=126,Y=128,Width=23,Height=32}{w-

Label	Count
NONAME	
DOCUMENT	
CC	
NOISE	
CHAR	
IMAGE	
ACCENT	
LINE	
WORD	
BLOCK	

X=764 Y=335

LIVRE PREMIER 455
faut laisser aus femmes cete vaine super-
flitiō de parolles) avec vne espōge. Voi-
la pourquoy, *spongia*, est vn mot ob-
scene en Latin: & estoit cete espōge
atachée au bout d'vn baston, comme
tesmoigne l'histoire de celuy qu'on me-
noit pour estre presenté aus bestes de-
uant le peuple, qui demanda congé d'al-
ler a ses affaires, & la n'ayant autre moiē
de se tuer, il se fourra ce baston & espō-
ge dans le gosier & s'en estoufa. Ilz s'es-
fuyoint le catze de laine perfumée, quād
ilz en auoint faict,
At tibi nil faciam, sed lota mētula lana.
Il y auoit aus carrefours a Rome des
vaiffeaus & demy cuues pour y aprester
a pisser aus passans.
Pulsi sepe lacū propter se ac dolia curta
Somno deuincti credunt extollere vestē
Ilz faisoient collation entre les repas. & y
auoit en esté des vèdeurs de negre pour
Ff 4

Un peu compliqué tous ça ?

- Sauvegarde du projet
- Sauvegarde du scenario
- Application du scenario sur un ouvrage complet...

- File ➔ Open Project1.xml
- Analyse des résultats générés

- Mise en place d'Assistants (*work in progress...*)

Assistant « Text Segmentation » ➔

Assistant « text segmentation »

TextSegmentation

Label of parent: DOCUMENT

Label of CC: CC

Prefix for new EOC labels: TXT_

Noise

Max Width: 5

Max Height: 5

Image...

Accent

Max Width: 40

Max Height: 20

Image...

Line

Char

Max Width: 80

Max Height: 40

Image...

Accent Y Spacing Max: 20

Image...

AD Max Width: 80

AD Max Height: 60

Image...

Block

Max spacing between lines: 200

Image...

Word

Max X spacing between chars: 15

Image...

Left RM X Max: -10

Right RM X Min: 10

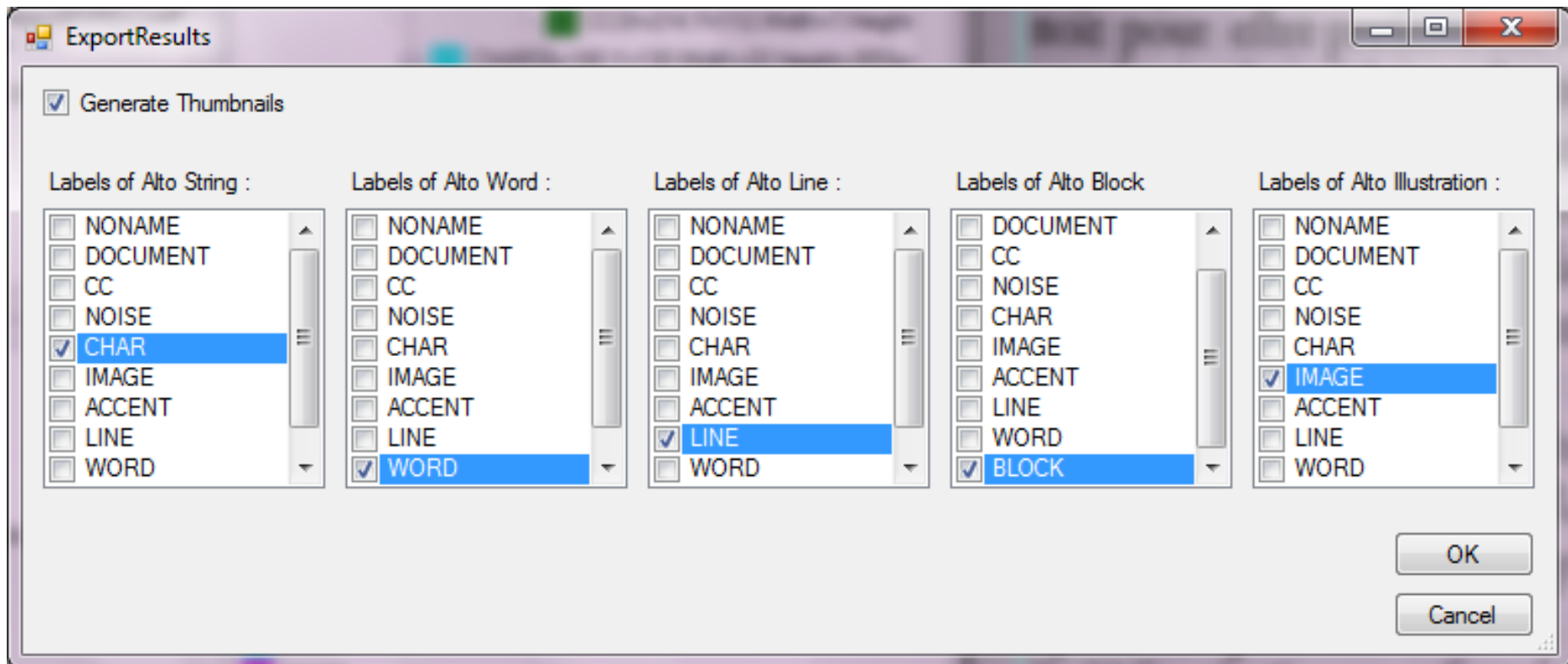
Document

Min X spacing between blocks: 100

Image...

integer ad

Exportation des résultats

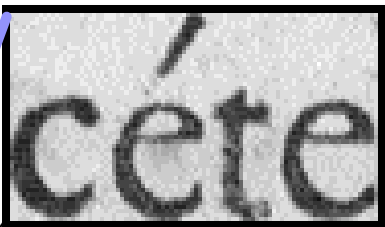


- Sauvegarde du projet
- Sauvegarde du scenario
- Application du scenario sur un ouvrage complet...

Sortie Alto

```
X:\AgoraTests\test01\alto\00000.xml - Notepad++
Fichier Edition Recherche Affichage Encodage Langage Paramétrage Macro Exécution Compléments Documents ?
00000.xml
240 <alto:Styles>
241   <alto:ParagraphStyle ID="BLOCK" />
242   <alto:ParagraphStyle ID="LINE" />
243   <alto:ParagraphStyle ID="CHAR" />
244 </alto:Styles>
245 <alto:Layout>
246   <alto:Page ID="00000" HEIGHT="1952" WIDTH="1008" PHYSICAL_IMG_NR="0">
247     <alto:PrintSpace HEIGHT="1672" WIDTH="889" HPOS="68" VPOS="48">
248       <alto:TextBlock ID="00000.0" STYLEREFS="BLOCK" HEIGHT="1672" WIDTH="889" HPOS="68" VPOS="48">
249         <alto:TextLine ID="00000.0.0" STYLEREFS="LINE" HEIGHT="56" WIDTH="713" HPOS="244" VPOS="48">
250           <alto:TextLine ID="00000.0.1" STYLEREFS="LINE" HEIGHT="72" WIDTH="852" HPOS="86" VPOS="107">
251             <alto:String ID="00000.0.1.0.0" STYLEREFS="CHAR" HEIGHT="54" WIDTH="25" HPOS="86" VPOS="108" CONTENT="*" />
252             <alto:String ID="00000.0.1.1.0" STYLEREFS="CHAR" HEIGHT="32" WIDTH="23" HPOS="126" VPOS="128" CONTENT="*" />
253             <alto:String ID="00000.0.1.1.1" STYLEREFS="CHAR" HEIGHT="32" WIDTH="15" HPOS="152" VPOS="128" CONTENT="*" />
254             <alto:String ID="00000.0.1.1.2" STYLEREFS="CHAR" HEIGHT="50" WIDTH="11" HPOS="173" VPOS="110" CONTENT="*" />
255             <alto:String ID="00000.0.1.1.3" STYLEREFS="CHAR" HEIGHT="30" WIDTH="22" HPOS="188" VPOS="130" CONTENT="*" />
256             <alto:String ID="00000.0.1.1.4" STYLEREFS="CHAR" HEIGHT="47" WIDTH="8" HPOS="214" VPOS="112" CONTENT="*" />
257             <alto:String ID="00000.0.1.1.5" STYLEREFS="CHAR" HEIGHT="51" WIDTH="41" HPOS="226" VPOS="107" CONTENT="*" />
258             <alto:String ID="00000.0.1.1.6" STYLEREFS="CHAR" HEIGHT="29" WIDTH="22" HPOS="255" VPOS="130" CONTENT="*" />
259             <alto:String ID="00000.0.1.1.7" STYLEREFS="CHAR" HEIGHT="31" WIDTH="18" HPOS="280" VPOS="128" CONTENT="*" />
260             <alto:String ID="00000.0.1.2.0" STYLEREFS="CHAR" HEIGHT="30" WIDTH="23" HPOS="312" VPOS="129" CONTENT="*" />
261             <alto:String ID="00000.0.1.2.1" STYLEREFS="CHAR" HEIGHT="30" WIDTH="25" HPOS="339" VPOS="129" CONTENT="*" />
262             <alto:String ID="00000.0.1.2.2" STYLEREFS="CHAR" HEIGHT="30" WIDTH="18" HPOS="367" VPOS="129" CONTENT="*" />
263             <alto:String ID="00000.0.1.3.0" STYLEREFS="CHAR" HEIGHT="50" WIDTH="26" HPOS="396" VPOS="109" CONTENT="*" />
264             <alto:String ID="00000.0.1.3.1" STYLEREFS="CHAR" HEIGHT="30" WIDTH="22" HPOS="412" VPOS="128" CONTENT="*" />
265             <alto:String ID="00000.0.1.3.2" STYLEREFS="CHAR" HEIGHT="30" WIDTH="44" HPOS="438" VPOS="128" CONTENT="*" />
266             <alto:String ID="00000.0.1.3.3" STYLEREFS="CHAR" HEIGHT="32" WIDTH="42" HPOS="486" VPOS="127" CONTENT="*" />
267             <alto:String ID="00000.0.1.3.4" STYLEREFS="CHAR" HEIGHT="30" WIDTH="21" HPOS="534" VPOS="129" CONTENT="*" />
268           </alto:TextLine>
269         </alto:TextBlock>
270       </alto:PrintSpace>
271     </alto:Page>
272   </alto:Layout>
273 </alto>
```

```
<alto:TextLine ID="00000.0.1" STYLEREFS="LINE" HEIGHT="72" WIDTH="852" HPOS="86" VPOS="107">
  <alto:String ID="00000.0.1.0.0" STYLEREFS="CHAR" HEIGHT="54" WIDTH="25" HPOS="86" VPOS="108" CONTENT="*" />
  <alto:String ID="00000.0.1.1.0" STYLEREFS="CHAR" HEIGHT="32" WIDTH="23" HPOS="126" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.1.1" STYLEREFS="CHAR" HEIGHT="32" WIDTH="15" HPOS="152" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.1.2" STYLEREFS="CHAR" HEIGHT="50" WIDTH="11" HPOS="173" VPOS="110" CONTENT="*" />
  <alto:String ID="00000.0.1.1.3" STYLEREFS="CHAR" HEIGHT="30" WIDTH="22" HPOS="188" VPOS="130" CONTENT="*" />
  <alto:String ID="00000.0.1.1.4" STYLEREFS="CHAR" HEIGHT="47" WIDTH="8" HPOS="214" VPOS="112" CONTENT="*" />
  <alto:String ID="00000.0.1.1.5" STYLEREFS="CHAR" HEIGHT="51" WIDTH="41" HPOS="226" VPOS="107" CONTENT="*" />
  <alto:String ID="00000.0.1.1.6" STYLEREFS="CHAR" HEIGHT="29" WIDTH="22" HPOS="255" VPOS="130" CONTENT="*" />
  <alto:String ID="00000.0.1.1.7" STYLEREFS="CHAR" HEIGHT="31" WIDTH="18" HPOS="280" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.2.0" STYLEREFS="CHAR" HEIGHT="30" WIDTH="23" HPOS="312" VPOS="129" CONTENT="*" />
  <alto:String ID="00000.0.1.2.1" STYLEREFS="CHAR" HEIGHT="30" WIDTH="25" HPOS="339" VPOS="129" CONTENT="*" />
  <alto:String ID="00000.0.1.2.2" STYLEREFS="CHAR" HEIGHT="30" WIDTH="18" HPOS="367" VPOS="129" CONTENT="*" />
  <alto:String ID="00000.0.1.3.0" STYLEREFS="CHAR" HEIGHT="50" WIDTH="26" HPOS="396" VPOS="109" CONTENT="*" />
  <alto:String ID="00000.0.1.3.1" STYLEREFS="CHAR" HEIGHT="30" WIDTH="22" HPOS="412" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.3.2" STYLEREFS="CHAR" HEIGHT="30" WIDTH="44" HPOS="438" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.3.3" STYLEREFS="CHAR" HEIGHT="32" WIDTH="42" HPOS="486" VPOS="127" CONTENT="*" />
  <alto:String ID="00000.0.1.3.4" STYLEREFS="CHAR" HEIGHT="30" WIDTH="21" HPOS="534" VPOS="129" CONTENT="*" />
  <alto:String ID="00000.0.1.3.5" STYLEREFS="CHAR" HEIGHT="30" WIDTH="18" HPOS="560" VPOS="129" CONTENT="*" />
  <alto:String ID="00000.0.1.4.0" STYLEREFS="CHAR" HEIGHT="30" WIDTH="22" HPOS="591" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.4.1" STYLEREFS="CHAR" HEIGHT="51" WIDTH="23" HPOS="616" VPOS="108" CONTENT="*" />
  <alto:String ID="00000.0.1.4.2" STYLEREFS="CHAR" HEIGHT="32" WIDTH="13" HPOS="641" VPOS="127" CONTENT="*" />
  <alto:String ID="00000.0.1.4.3" STYLEREFS="CHAR" HEIGHT="29" WIDTH="23" HPOS="658" VPOS="129" CONTENT="*" />
  <alto:String ID="00000.0.1.5.0" STYLEREFS="CHAR" HEIGHT="31" WIDTH="24" HPOS="695" VPOS="128" CONTENT="*" />
  <alto:String ID="00000.0.1.5.1" STYLEREFS="CHAR" HEIGHT="31" WIDTH="18" HPOS="722" VPOS="129" CONTENT="*" />
```





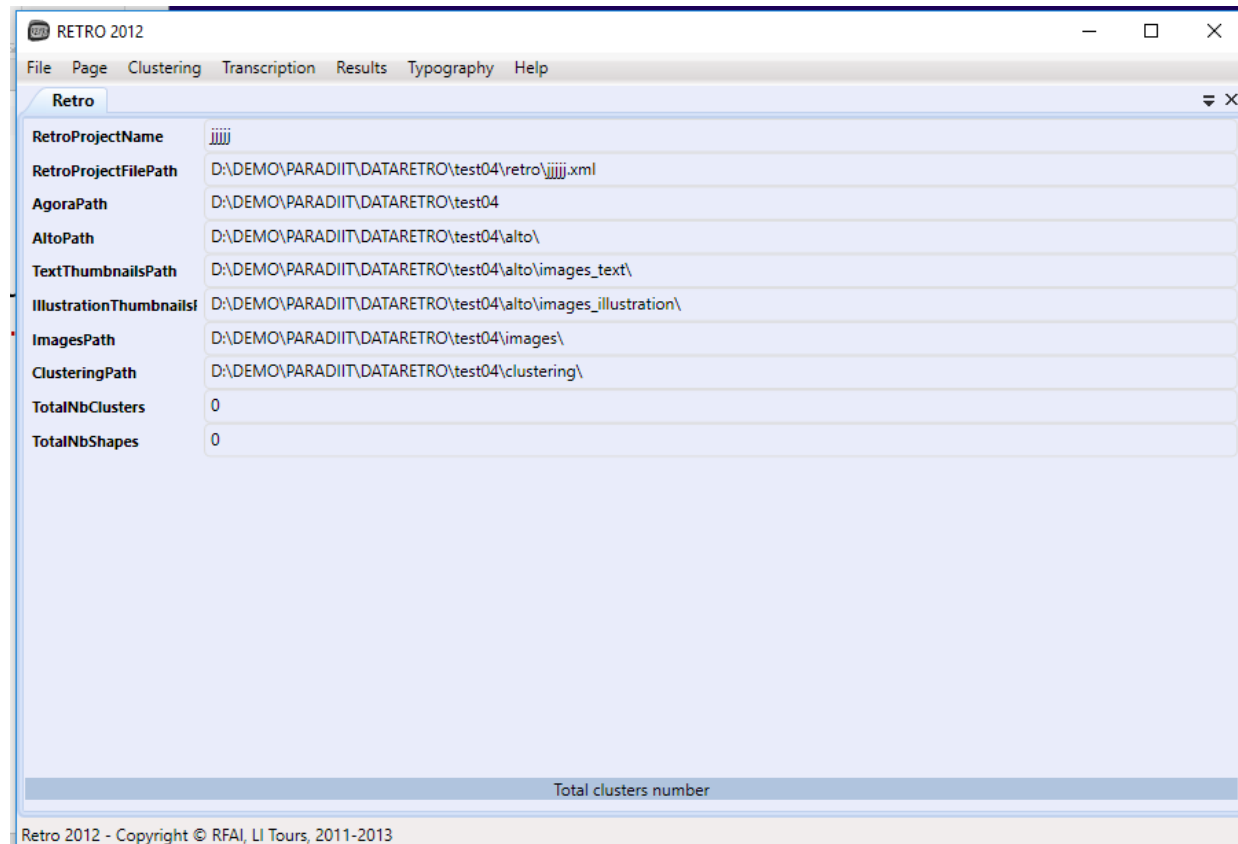
RETRO

Regardez la vidéo de démonstration en même temps que vous avancez est indispensable...

RETRO : Comparaison de formes sans reconnaissance

Analyse de la redondance

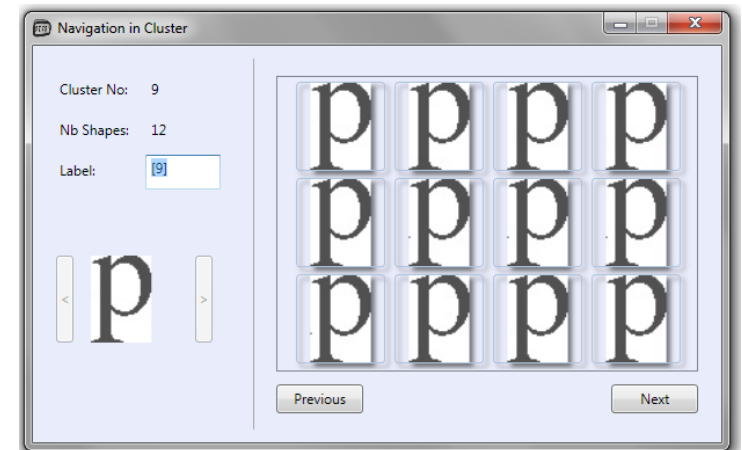
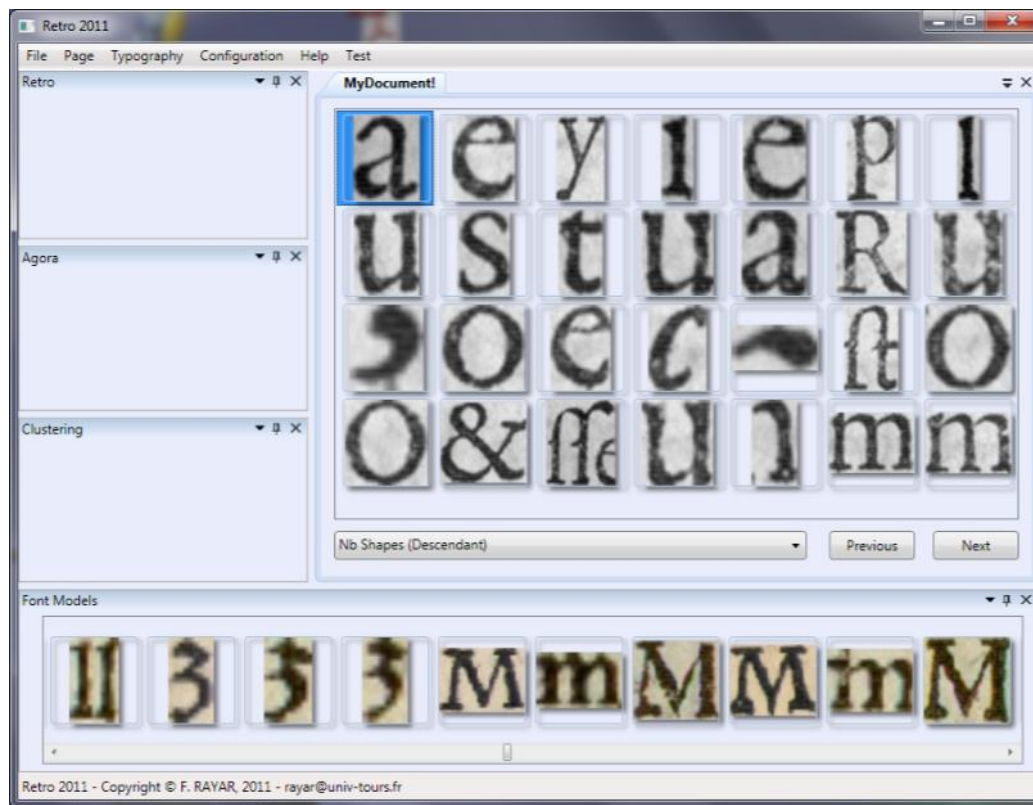
- Interfaces de gestion du clustering & de vérification des clusters
- Lancer Retro → New Project → RetroData\test04\projet01.xml



RETRO : Comparaison de formes sans reconnaissance

Analyse de la redondance

- Interfaces de gestion du clustering & de vérification des clusters
- Clustering → Process Clustering → Load Clusters → View Clusters

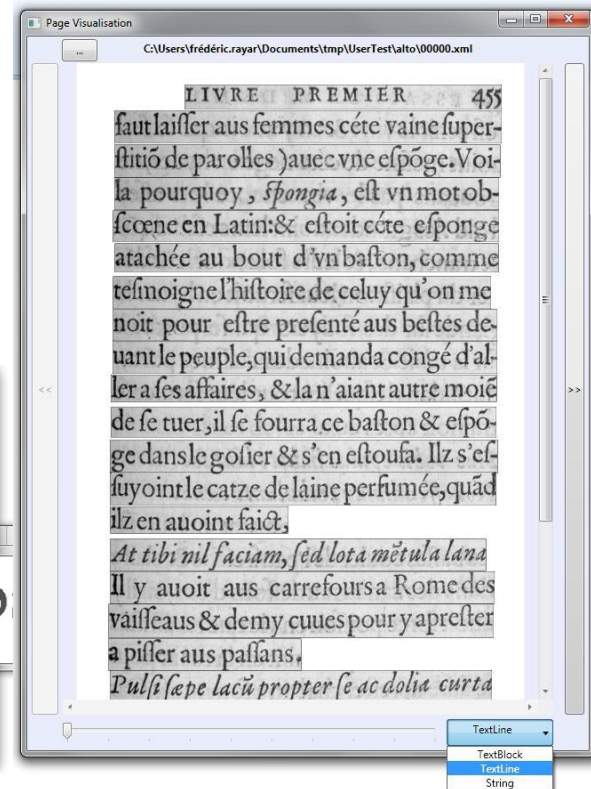
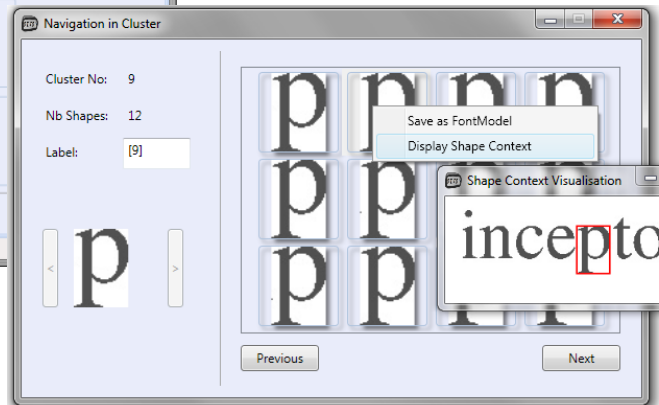
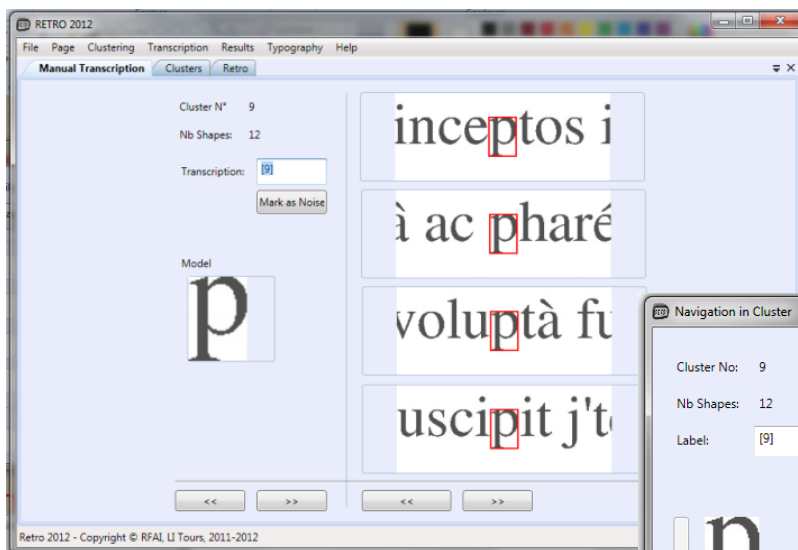


```
<?xml version="1.0" encoding="utf-8"?>
<cluster size="36" label="">
  <cc id="00000.0.0.0.1"/>
  <cc id="00000.0.0.0.1.1"/>
  ...
  <cc id="00000.0.9.2.6"/>
</cluster>
```

RETRO : Comparaison de formes sans reconnaissance

Analyse de la redondance

- Interfaces de visualisation des clusters + contexte + création de modèles typographiques
- Interfaces de transcription manuelle (et automatique)
- Interfaces de visualisation des fichiers Alto et EoC (Pages)
- **Voir la vidéo de démonstration de RETRO**

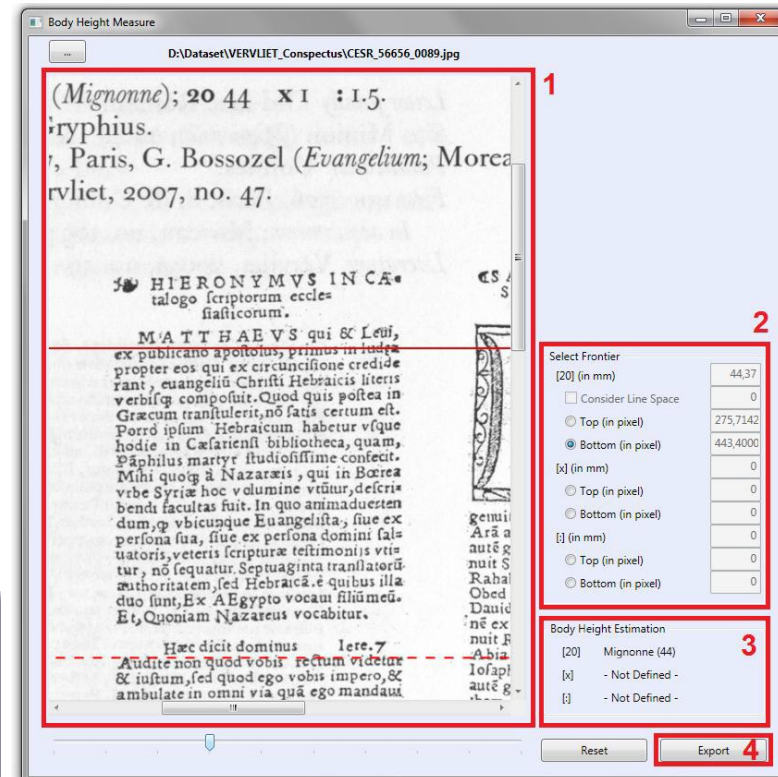


RETRO : Comparaison de formes sans reconnaissance

Analyse typographique

■ Interfaces d'analyse typographique

- Création de modèles à partir d'images
- Visualisation de base de modèles
- Mesure et identification de Fontes
 1. Select a page
 2. Selection of the frontier (top/bottom for [20], [x], [:] height), and display of measured values (in pixels and mm)
 3. Estimated Body height designation
 4. Possibility to export computed information in xml



```
<?xml version="1.0" encoding="utf-16"?>
<!--RETRO Body Height file-->
<BodyHeight>
  <Filename Path="D:\Dataset\VERVLIET_Conspectus\CESR_56656_0089.jpg" />
  <Typography FrenchName="Mignonne" FrenchCode="MI" EnglishName="Minion" />
  <[20] Value="44,37" />
  <[x] Value="0" />
  <[:] Value="0" />
</Typography>
</BodyHeight>
```

```
<?xml version="1.0" encoding="utf-16"?>
<!--RETRO Model file-->
<Model>
  <Metadata>
    <Book Author="" Place="" PrinterOrPublisher="" Date="" Format="" />
    <Copy Library="" CallNumber="" Digitalization="" Copyright="" CataloguerName="" />
  </Metadata>
  <Transcription Character="" Unicode="" />
  <Image Filename="" Folder="" Page="" Resolution="" />
  <Thumbnail Name="" Width="" Height="" PositionX="" PositionY="" />
  <Typography IsSmallCap="" Type="" Alphabet="" Family="" SubFamily="" BodyHeight="" Thickness="" />
  <Description References="" Engraver="" Comments="" />
</Model>
```