

## **Mode d'emploi**

Pour l'obtention du diplôme d'Ingénieur d'Etat en Informatique

### **Option: Systèmes Informatiques**

## **Thème**

**Développement d'une plate forme logicielle  
pour la recherche de l'information dans les  
documents historiques par le Word Spotting**

#### **Réalisé par**

- BOUDRAA Omar  
- OULAD KADDOUR Mohamed

#### **Encadré par**

-Mr SEHAD Abdnour  
-Pr CHIBANI YUCEF

**Promotion : 2010 / 2011**

# **Sommaire**

<b>1. Introduction .....</b>	<b>3</b>
<b>2. Installation .....</b>	<b>4</b>
<b>3. Schéma à suivre .....</b>	<b>5</b>
<b>4. Exemple applicatif .....</b>	<b>(6-10)</b>
<b>5. Modules supplémentaires .....</b>	<b>(11-12)</b>
<b>6. Annexe .....</b>	<b>(13-14)</b>

# 1. Introduction

---

le Word Spotting est une technique de recherche d'information dans les documents anciens qui consiste à repérer les occurrences des mots par appariement. Ce qui augmente le taux de reconnaissance et facilite l'accès aux documents dégradés.

Le principe de repérage est de construire une image de requête pour le mot à rechercher qui sera ensuite comparée par similarité avec une suite d'images, présentant les mots localisés dans l'image du document.

Plusieurs recherches ont été faites afin de développer des méthodes efficaces pour le Word Spotting mais aucune d'elle n'est universelle et il reste toujours un champ d'améliorations.

## 2. Installation:

---

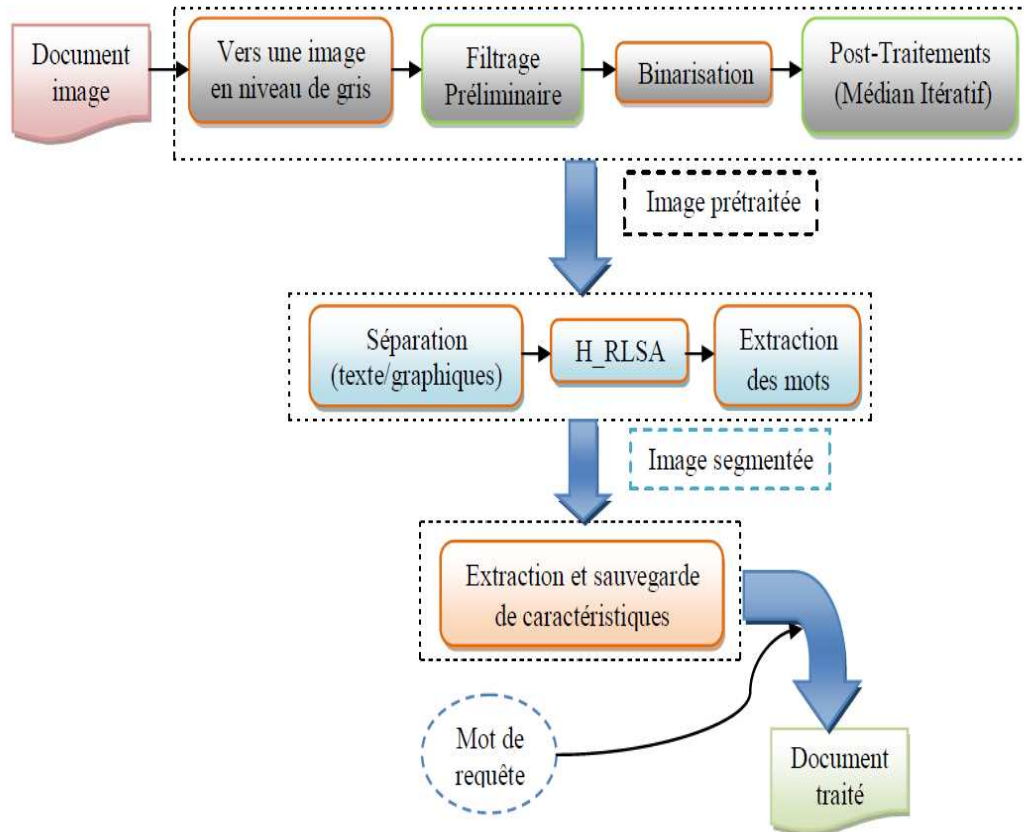
L'installation de cette application est simple et facile à implémenter.

Il suffit de choisir les options offerts (détermination de dossier d'extraction ...etc) le long du chemin d'installation.



### 3. Shéma à suivre

-L'utilisateur doit suivre le chemin suivant:



Tâches indispensables.

Tâches facultatives (restent au choix d'utilisateur (lors de la nécessité)).

## 4. Exemple applicative

---

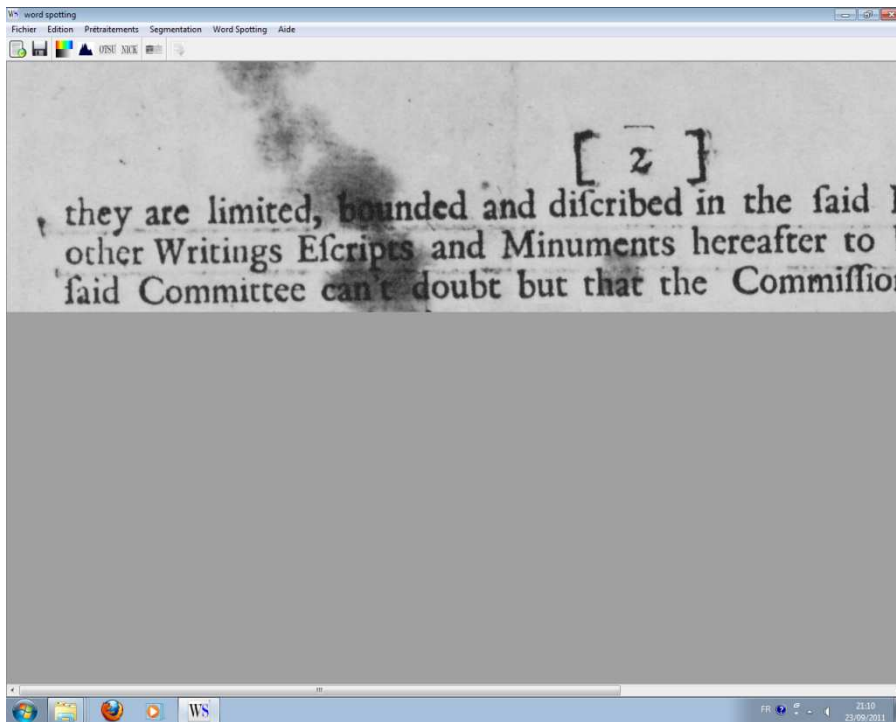


Image originale

(issue de la base DIBCO (Document Image Binarization Contest);voir annexe).

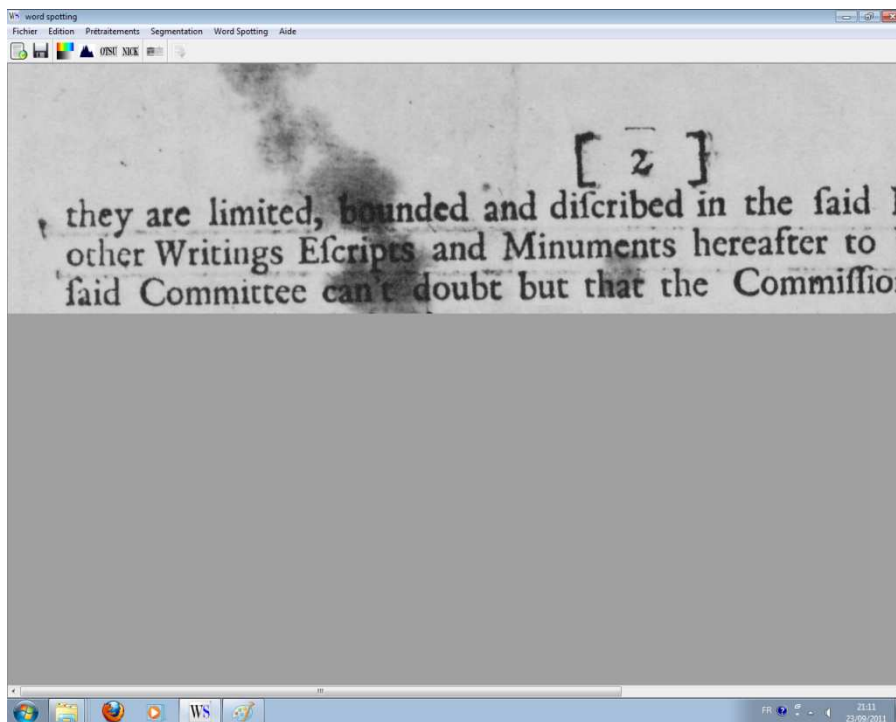
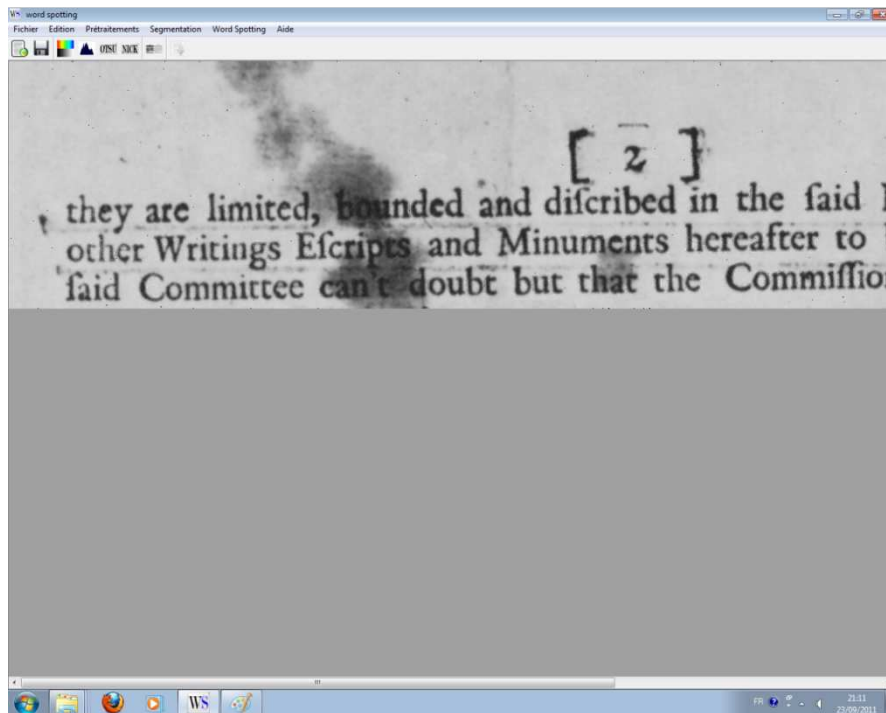


Image en niveau de gris.



Amélioration de la qualité d'image par le filtre de Wiener  
(l'utilisateur a le droit d'utiliser n'importe quel filtre en dépend du cas à traiter).

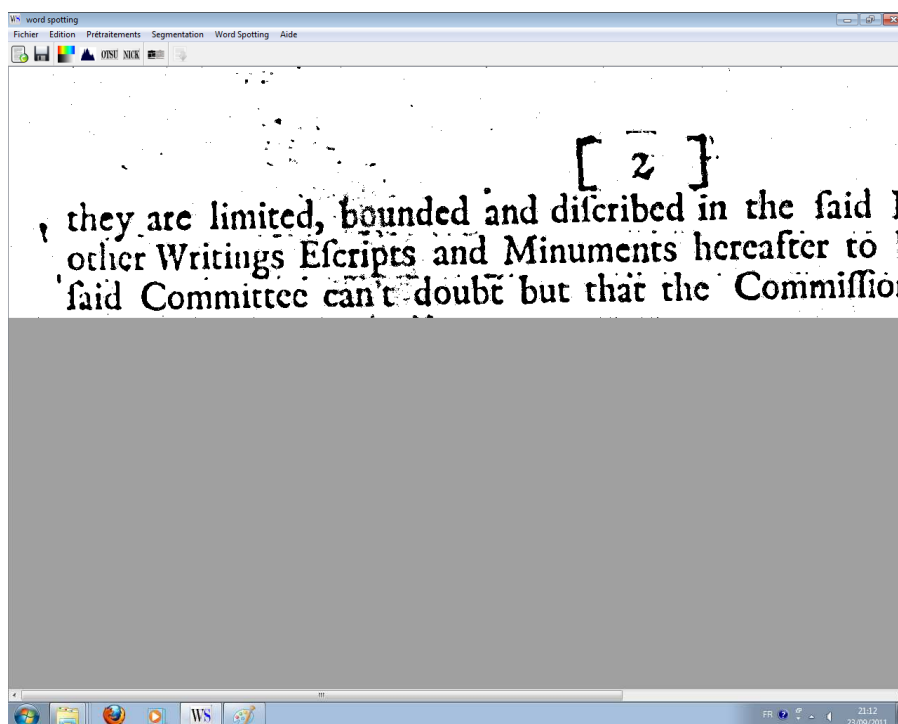
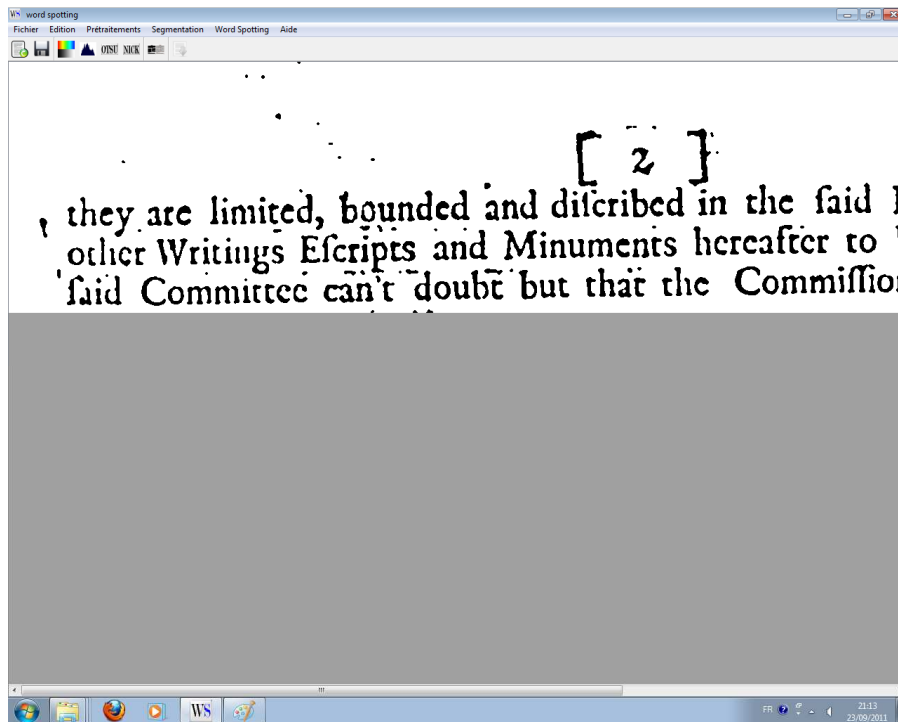
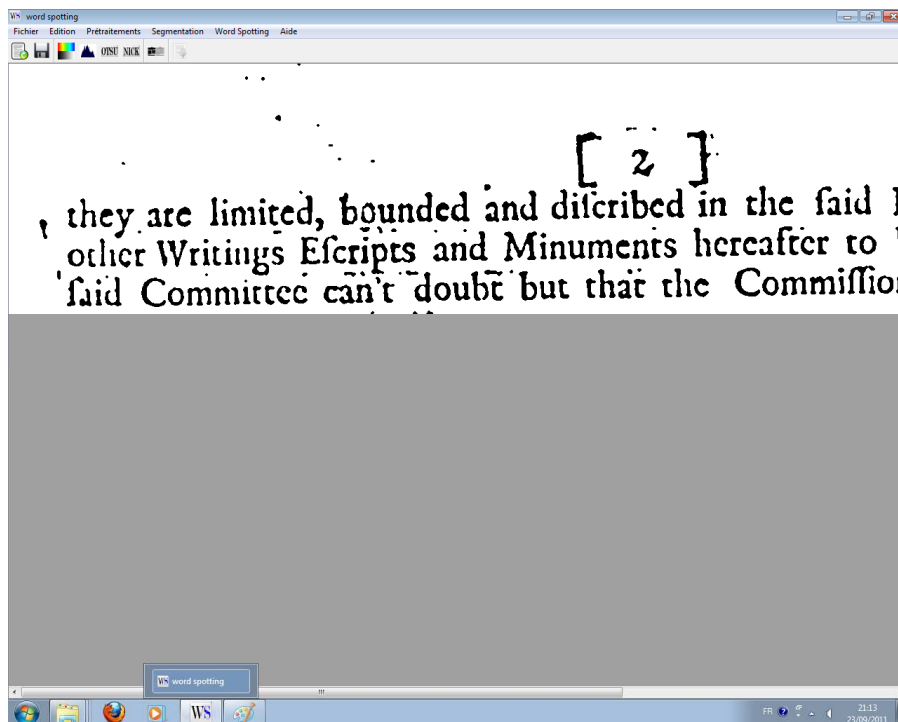


Image binaire (Nick).

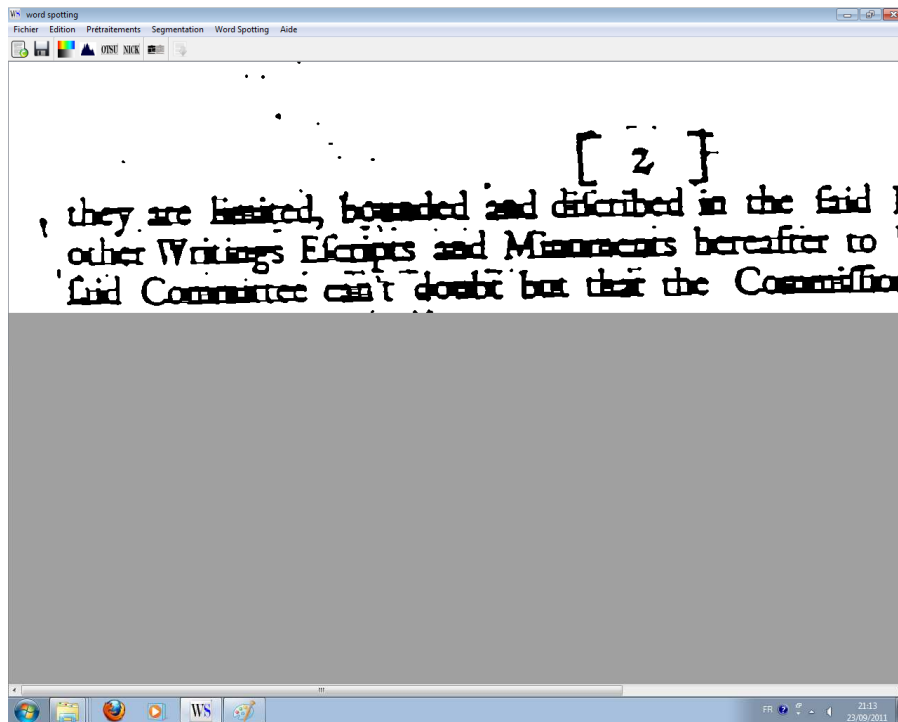


Exténuation de parasites (filtre médian itératif (7 itérations)).

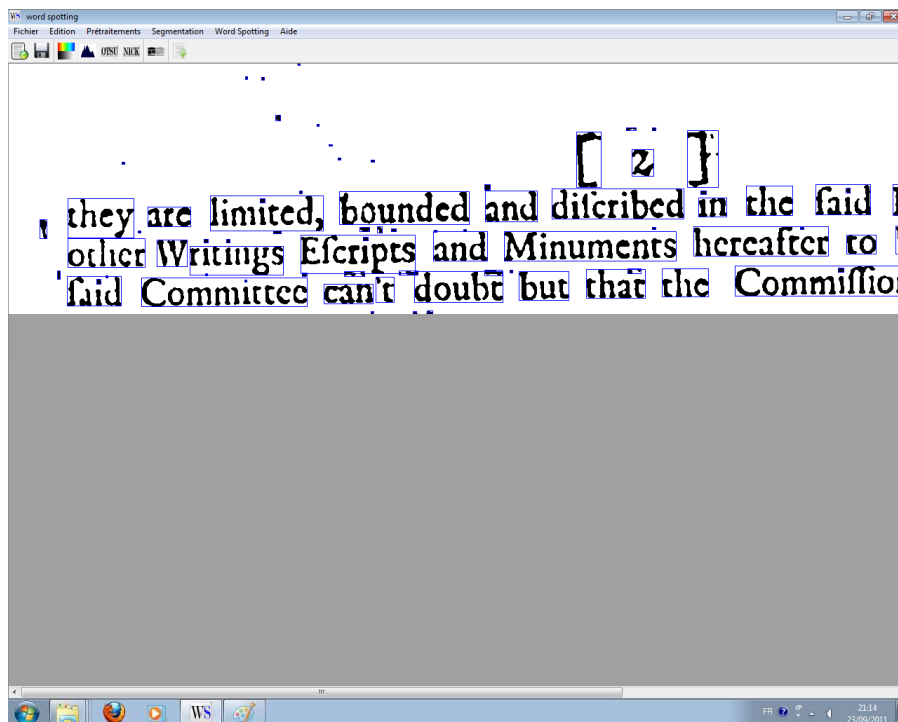


Séparation (texte/graphiques)  
(pas de graphiques).



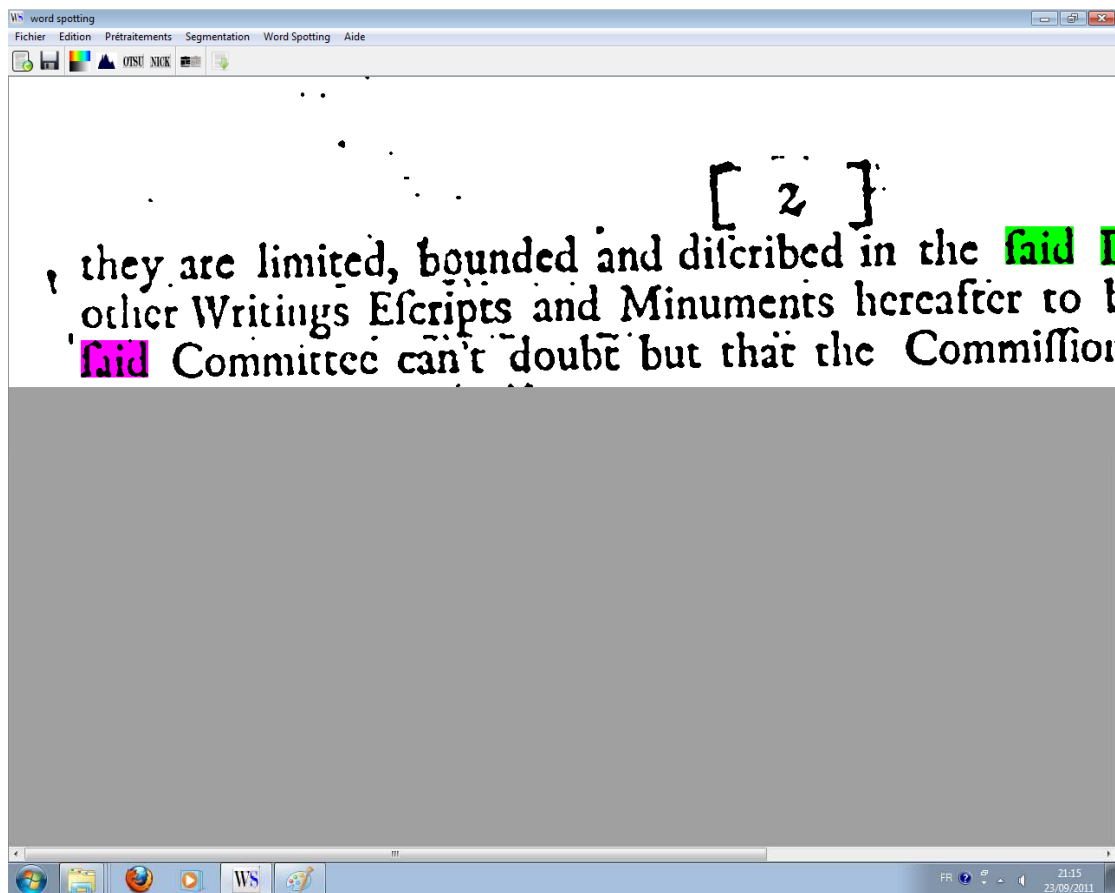


H\_RLSA.



Extraction des mots

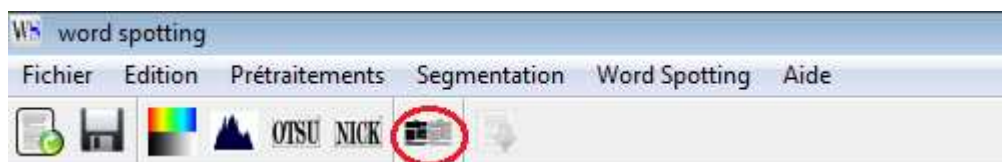
(cette étape devra suivie par l'extraction et la sauvegarde de caractéristiques).



Repérage du mot (faid).

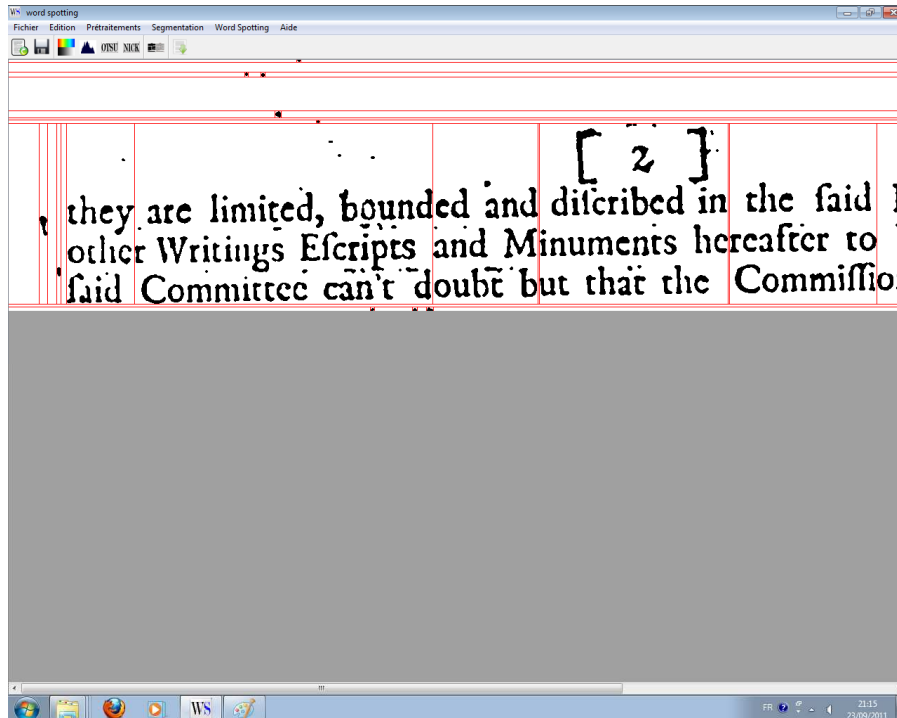
### **Note :**

-l'utilisateur peut cliquer sur le bouton de « segmentation » dans le menu principal au lieu de passer par les trois étapes de segmentation dans la barre d'outils.

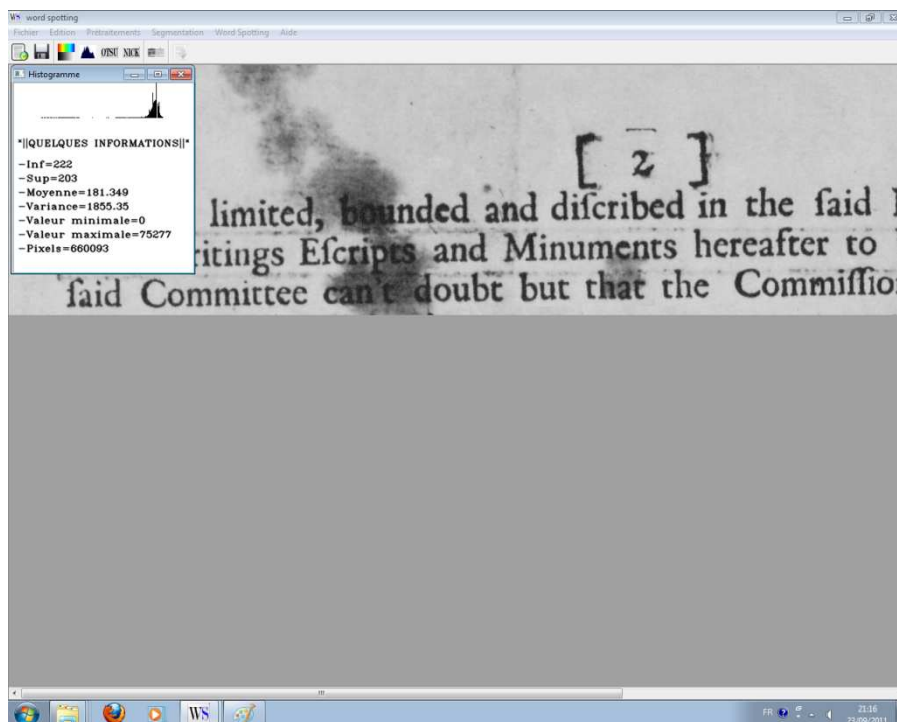


## 5. Modules supplémentaires

---



Segmentation par le profil de projection (avec un niveau hiérarchique =2).



Histogramme étiré d'image en niveau de gris.

sqlite											
ID	TYPE	X	Y	LARGEUR	HAUTEUR	PROFIL VERTICAL	PROFIL SUPERIEUR	PROFIL INFÉRIEUR	HISTOGRAMME VERTICAL	TRANSITIONS	TRANSITIONS DE LA LIGNE CENTRALE
0	BRUIT	1152	355	1	1	0.576	0	0	1	0	0
1	BRUIT	219	354	2	1	0 0	0 0	0 0	1 1	0 0	0 0
2	MOT	576	353	6	3	0.418 0.425 0.42 0....	0.333 0 0 0 0.333 0....	0.667 0.667 0.667 0....	0.667 0.667 1 1 0.667 0.667	0.5 1 0 0 0.5 0.5	0 0 0 0 0 0
3	MOT	596	352	9	4	0.52 0.431 0.384 0....	0.75 0.25 0 0 0.25 0....	0.75 0.75 0.75 0.75 ...	0.25 0.75 1 1 0.75 0.5 0.5 ...	1 1 0 0 1 1 1 1 1 1	0 1 0 0 0 0 0 0 1
4	MOT	516	352	5	4	0.429 0.391 0.378 0...	0.75 0.5 0 0 0.5	0.75 0.75 0.75 0.75 ...	0.25 0.5 0.75 1 0.5	0.5 0.5 1 0 0.5	0 1 0 0 0
5	BRUIT	1543	351	1	1	0	0	0	1	0	0
6	BRUIT	141	351	1	2	0	0	0.5	1	0	0
7	BRUIT	1205	347	1	1	0	0	0	1	0	0
8	BRUIT	1127	347	1	1	0.541	0	0	1	0	0
9	BRUIT	1431	346	2	1	0.494 0.565	0 0	0 0	1 1	0 0	0 0
10	BRUIT	1159	344	1	1	0	0	0	1	0	0
11	BRUIT	353	344	1	1	0.157	0	0	1	0	0
12	BRUIT	1368	343	1	1	0.659	0	0	1	0	0
13	BRUIT	502	342	1	1	0.302	0	0	1	0	0
14	BRUIT	1203	340	2	1	0 0	0 0	0 0	1 1	0 0	0 0
15	BRUIT	555	340	1	1	0.31	0	0	1	0	0
16	BRUIT	636	338	1	1	0.494	0	0	1	0	0
17	BRUIT	611	338	1	1	0.345	0	0	1	0	0
18	BRUIT	566	337	1	1	0.31	0	0	1	0	0
19	BRUIT	568	334	1	1	0.294	0	0	1	0	0
20	BRUIT	1402	330	1	1	0.475	0	0	1	0	0
21	BRUIT	1200	330	1	1	0.522	0	0	1	0	0
22	BRUIT	1132	330	1	1	0.522	0	0	1	0	0
23	MOT	562	329	10	4	0.285 0.266 0.26 0....	0.25 0.25 0 0.25 0.2...	0.25 0.5 0.5 0.75 0.7...	0.25 0.5 0.75 0.5 0.75 0.75 ...	0.667 0.667 0.3...	0 1 0 1 1 0 1 0 0 0
24	BRUIT	906	327	1	1	0.529	0	0	1	0	0
25	BRUIT	802	327	1	1	0	0	0	1	0	0
26	BRUIT	571	327	1	1	0.259	0	0	1	0	0
27	MOT	550	327	5	2	0.235 0.275 0.239 0...	0 1 0 1 0	0 0 0 0 0.5	0.5 0 0.5 0 1	1 0 1 0 0	0 0 0 0 1
28	BRUIT	1765	326	1	1	0.533	0	0	1	0	0
29	BRUIT	1580	326	1	1	0.271	0	0	1	0	0
30	BRUIT	556	323	1	1	0.231	0	0	1	0	0

Affichage de caractéristiques.

## 6. Annexe

---

### Les différentes images de la base DIBCO

#### 1. Images imprimées (Printed)

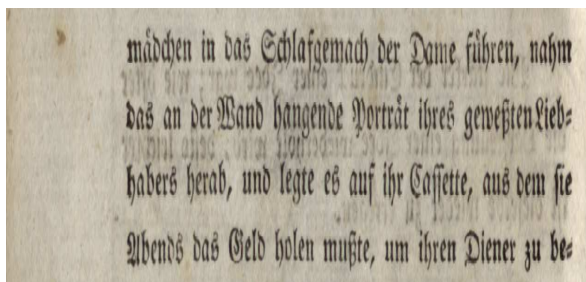


Figure A1 : PO1

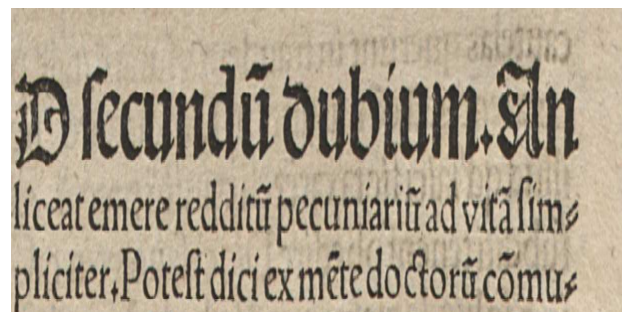


Figure A2 : PO2

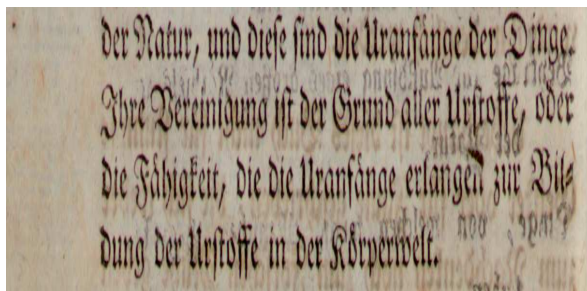


Figure A3 : PO3

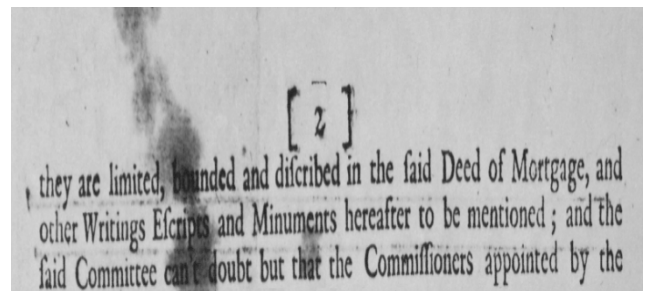


Figure A4 : PO4



Figure A5 : PO5

## 1. Images manuscripts (Handwritten)

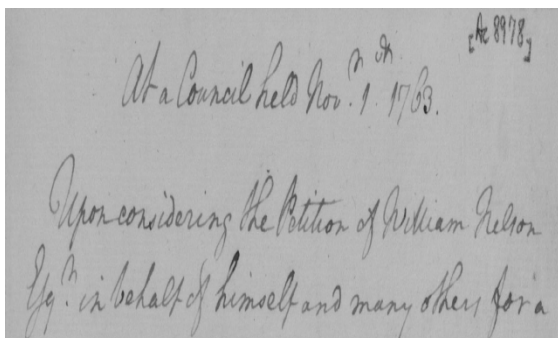


Figure A6 : H01

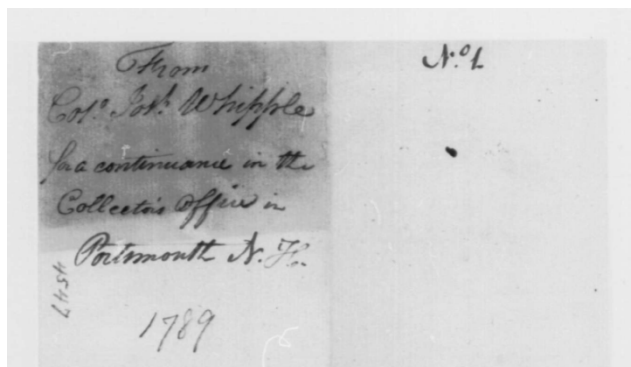


Figure A7 : H02

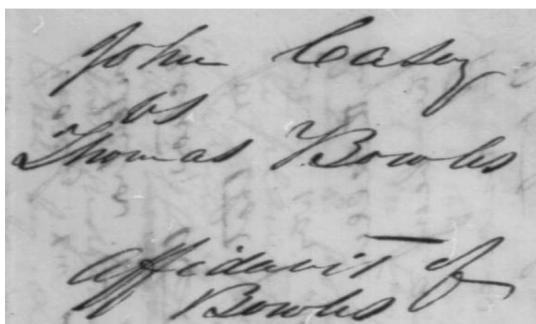


Figure A8 : H03

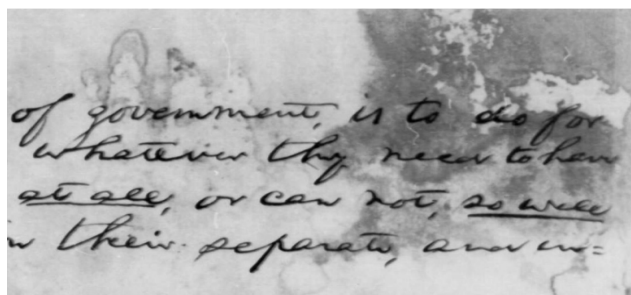


Figure A9 : H04

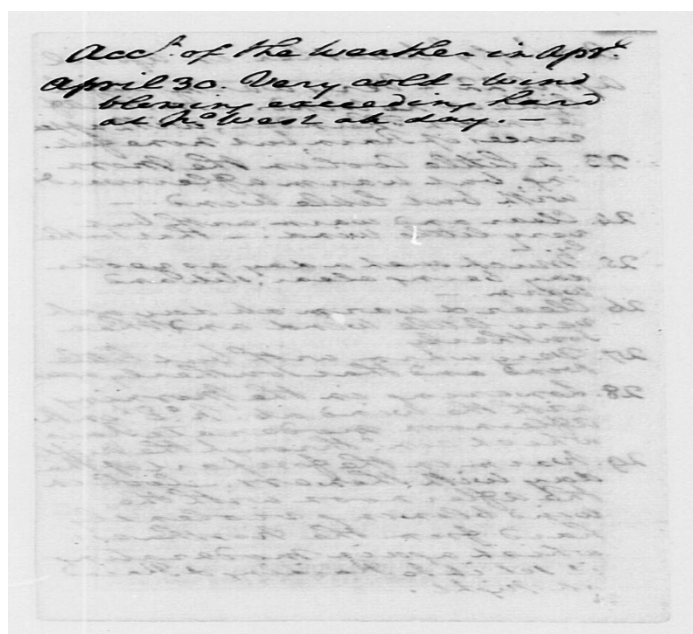


Figure A10 : H05

Référence : <http://users.iit.demokritos.gr/>