



HAL
open science

DMOS : A vous de jouer !

Aurélie Lemaitre, Bertrand B. Couïasnon

► **To cite this version:**

Aurélie Lemaitre, Bertrand B. Couïasnon. DMOS : A vous de jouer !. Symposium International Francophone sur l'Écrit et le Document (SIFED), Jun 2017, Nantes, France. hal-01534885

HAL Id: hal-01534885

<https://inria.hal.science/hal-01534885v1>

Submitted on 9 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DMOS, à vous de jouer !

Aurélie Lemaitre – Bertrand Coüasnon
SIFED 2017

Institut de Recherche en Informatique et Systèmes Aléatoires

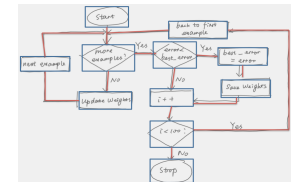
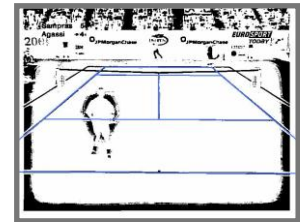
DMOS

- **Méthode de reconnaissance de structure de documents**

- Description et MOdification de la Segmentation
- Développée dans l'équipe Intuidoc de l'IRISA – Rennes

- **Caractéristiques**

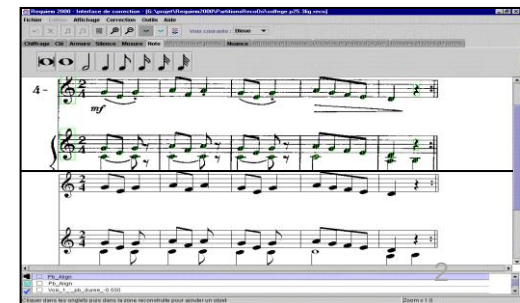
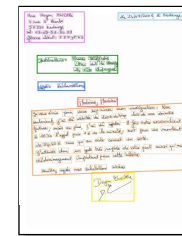
- Méthode syntaxique
 - Expression des connaissances à base de règles
- Méthode générique
 - Applicable à de nombreux types de documents
 - Décrire simplement les connaissances applicatives



- **Large validation**

- Dans de nombreux domaines
- A grande échelle

$$\prod_{i=0}^n \frac{2}{3} - \int_0^{\frac{\pi}{2}} y^2 + 3dy - 1 + 2 \times \sum_{i=0}^n i^2 / \frac{1+1+i}{3+2+yz} \times e^{\frac{1}{3}+2} \times \frac{1+1+i}{2xy + \frac{5}{2}+6} \times e^{\frac{1}{3}+2}$$



A vous de jouer !

- **Mise à disposition de DMOS**
 - Pour des fin recherche
 - Environnement de développement complet

 - **Améliorer vos systèmes de reconnaissance**
 - Fusion d'information issues de (vos !) sources variées
 - Guidée par des connaissances explicites
- ⇒ **Gain en performance et en généricité**

PLAN

- **Exemples d'analyse d'images de documents avec DMOS**
- **Fonctionnement de la méthode DMOS**
- **A vous de jouer !**

Structure de Registres matricules

- Description de structure tabulaire
- Utilisation de connaissances a priori pour compenser les difficultés liées aux documents

Duplicata levé individuel compté le 10 Juin 1912

Tournier

Nom : Tournier Prénoms : Alphonse Marie Louis

Matricule : 504

Classe de mobilisation : 189

ÉTAT CIVIL

Né le 11 Janvier 1880 à La Breuille

SIGNALEMENT

Classe et front : 189

Taille : 1 m. 62

DÉCISION DE CONSEIL DE REVISION ET MOTIFS

Compte dans la "1" partie de la liste de recensement national (.....) positif.

ÉTAT DES SERVICES ET MUTATIONS DIVERSES

Parti dans le 1^{er} Régiment d'Artillerie de Réserve de la Meuse le 22 août 1910.

ÉTAT DES SERVICES

Année	Service	Matricule	Date
1898	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910
1900	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910
1902	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910
1904	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910
1906	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910
1908	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910
1910	1 ^{er} Régiment d'Artillerie de Réserve	504	22 août 1910

Brochard

Nom : Brochard Prénoms : Amos Constant

Matricule : 626

Classe de mobilisation : 189

ÉTAT CIVIL

Né le 14 Septembre 1876 à Grogny

SIGNALEMENT

Classe et front : 189

Taille : 1 m. 52

DÉCISION DE CONSEIL DE REVISION ET MOTIFS

Compte dans la "2" partie de la liste de recensement national (.....) positif.

ÉTAT DES SERVICES ET MUTATIONS DIVERSES

Parti dans le 1^{er} Régiment d'Artillerie de Réserve de la Meuse le 22 août 1910.

ÉTAT DES SERVICES

Année	Service	Matricule	Date
1898	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910
1900	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910
1902	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910
1904	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910
1906	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910
1908	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910
1910	1 ^{er} Régiment d'Artillerie de Réserve	626	22 août 1910

Sefeuette

Nom : Sefeuette Prénoms : François Eugène

Matricule : 535

Classe de mobilisation : 189

ÉTAT CIVIL

Né le 28 août 1876 à Buxeuil

SIGNALEMENT

Classe et front : 189

Taille : 1 m. 72

DÉCISION DE CONSEIL DE REVISION ET MOTIFS

Compte dans la "1" partie de la liste de recensement national (.....) positif.

ÉTAT DES SERVICES ET MUTATIONS DIVERSES

Parti dans le 1^{er} Régiment d'Artillerie de Réserve de la Meuse le 22 août 1910.

ÉTAT DES SERVICES

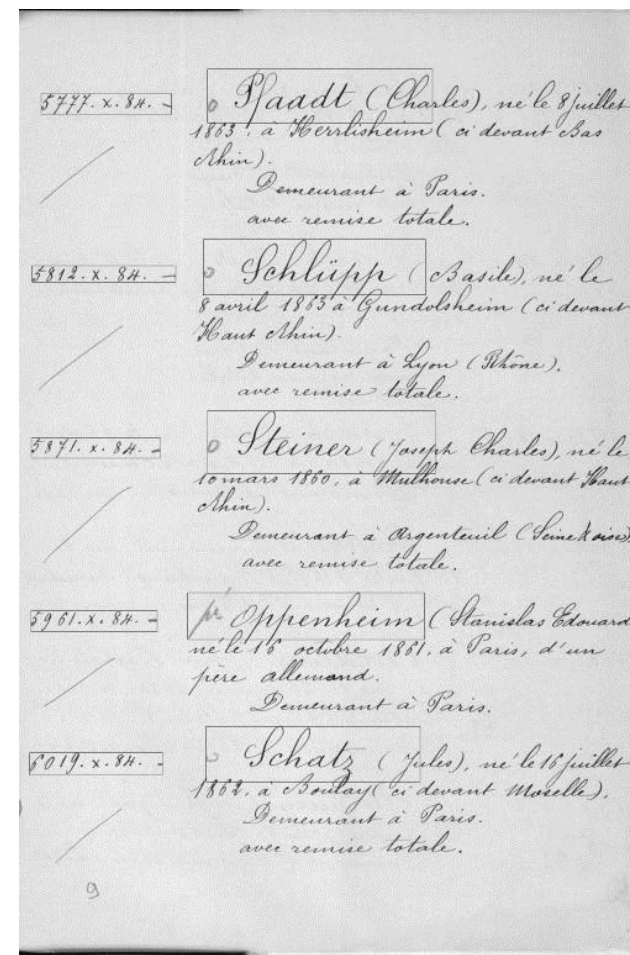
Année	Service	Matricule	Date
1898	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1900	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1902	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1904	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1906	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1908	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1910	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910

ÉTAT DES SERVICES

Année	Service	Matricule	Date
1898	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1900	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1902	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1904	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1906	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1908	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910
1910	1 ^{er} Régiment d'Artillerie de Réserve	535	22 août 1910

Structure de décrets de naturalisation

- Description du contenu logique du document
- Utilisation de deux résolutions d'image



Reconnaissance de structure dans la base RIMES


- Localisation de lignes de texte par vision perceptive
 - Vue de loin : les lignes de texte sautent aux yeux (segments)
 - Vue de près : détail des caractères
- Intégration stochastique de scores de reconnaisseurs
 - Localisation ligne d'ouverture selon une liste de mots clés

HARTEL Michael
 2 rue du Tillard
 68480 Fels
 Tél : 03 72 16 27 71

las 3 Saifs
 Service communal
 Rue Haute
 46 340 Salviac

Objet: Modification de commande.

Madame, Monsieur,
 Je vous contacte au sujet de l'adresse
 commune de CD viengs que j'ai posée
 le 03/06/2006. La référence client est
 HVRN9447. Je souhaite doubler la
 quantité des CDs commandés.

Cordialement,
 HARTEL Michael


Objet:

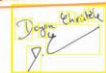
Madame, Monsieur,
 Je vous contacte
 commune de CD vi

[Faded handwritten text with some lines highlighted in green]

Objet: Modification de commande.

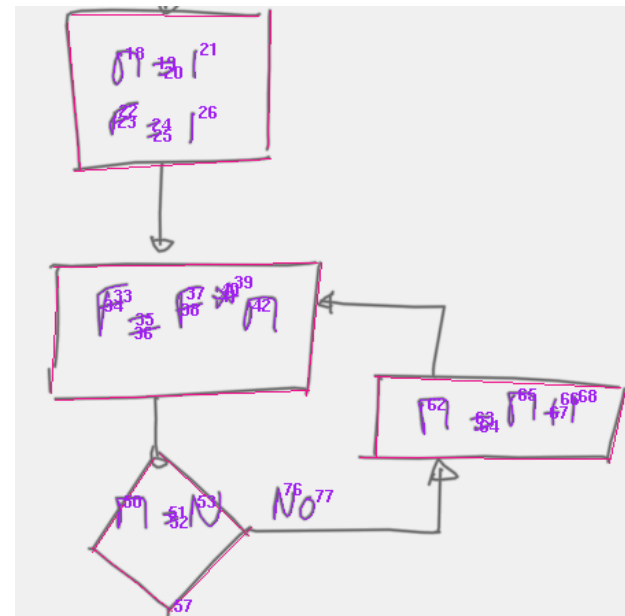
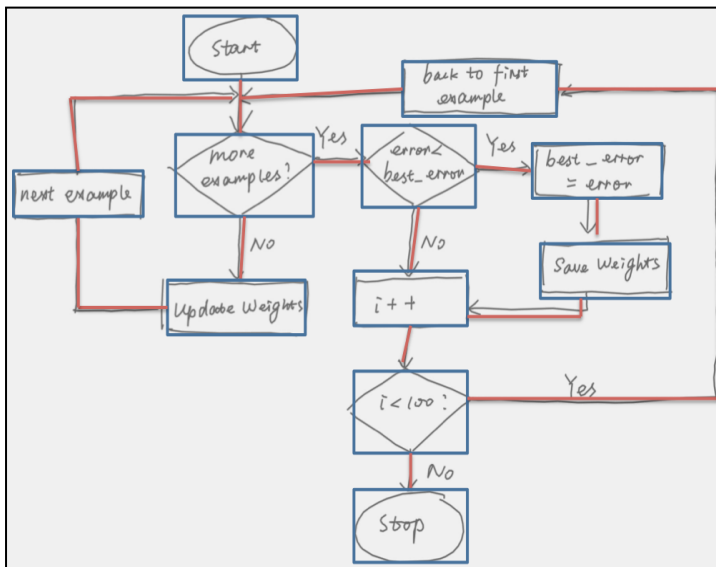
Madame, Monsieur,

Je vous contacte au sujet de l'adresse commune de CD viengs que j'ai posée le 03/06/2006. La référence client est HVRN9447. Je souhaite doubler la quantité des CDs commandés.

Cordialement,
 HARTEL Michael


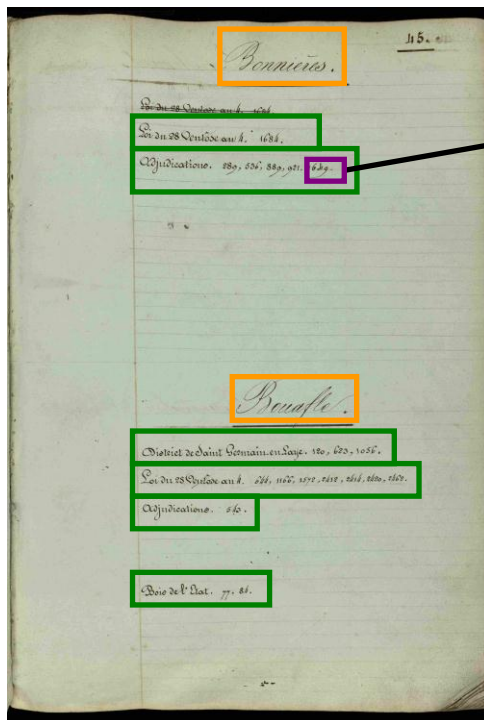
Reconnaissance de flowcharts en ligne

- **Description logique des contenus**
 - Ordre des boites
- **Interaction avec les classifieurs de strokes**
 - Stratégie de prise de décision en fonction
 - du contexte logique local
 - et du résultat du classifieur



Documents de séquestres révolutionnaires

- **Modélisation de la collection**
 - Transmission d'informations entre pages
- **Transcription assistée de mots manuscrits**
 - Mutualisation des demandes d'interaction
 - Intervention asynchrone de l'utilisateur



Table

Année	Contenu	Statut	Page	Observations
1784	2 oct 1785	2 de 20	100	
1785	1785	100		
1786	1786	100		
1787	6 oct 1787	100		
1788	1788	100		
1789	1789	100		
1790	1790	100		
1791	1791	100		
1792	1792	100		
1793	1793	100		
1794	1794	100		
1795	1795	100		
1796	1796	100		
1797	1797	100		
1798	1798	100		
1799	1799	100		
1800	1800	100		

Registre

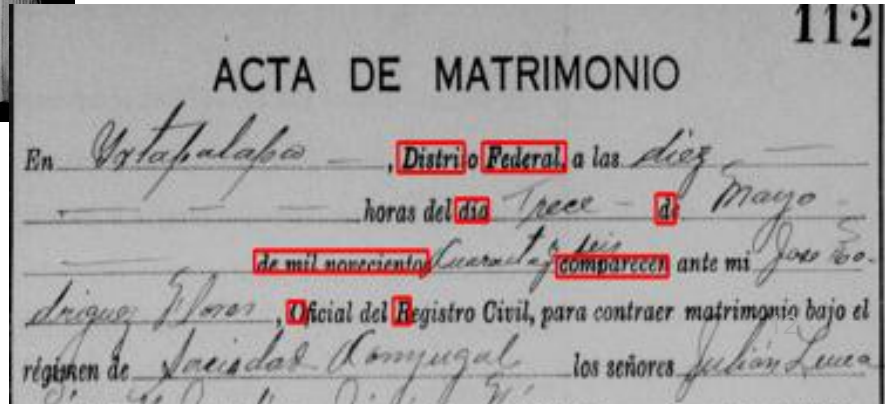
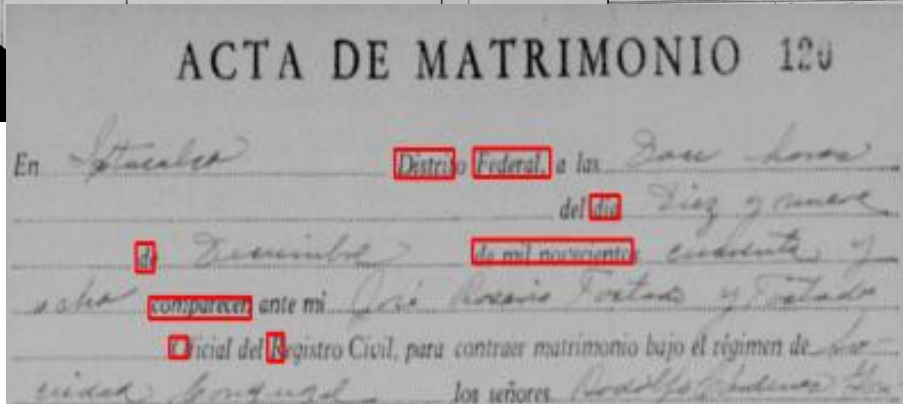
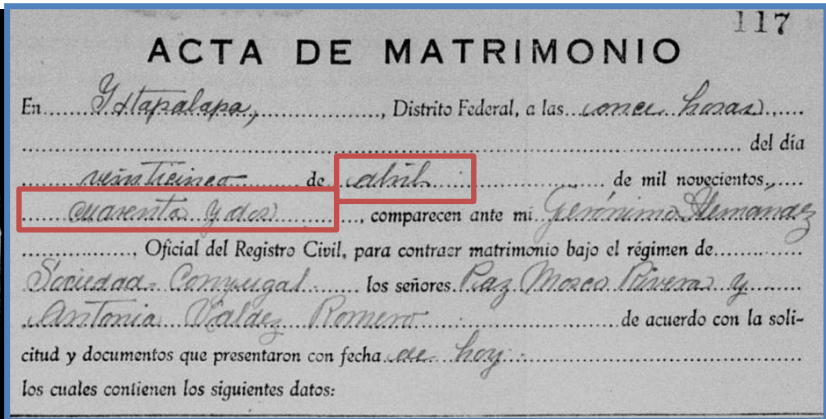
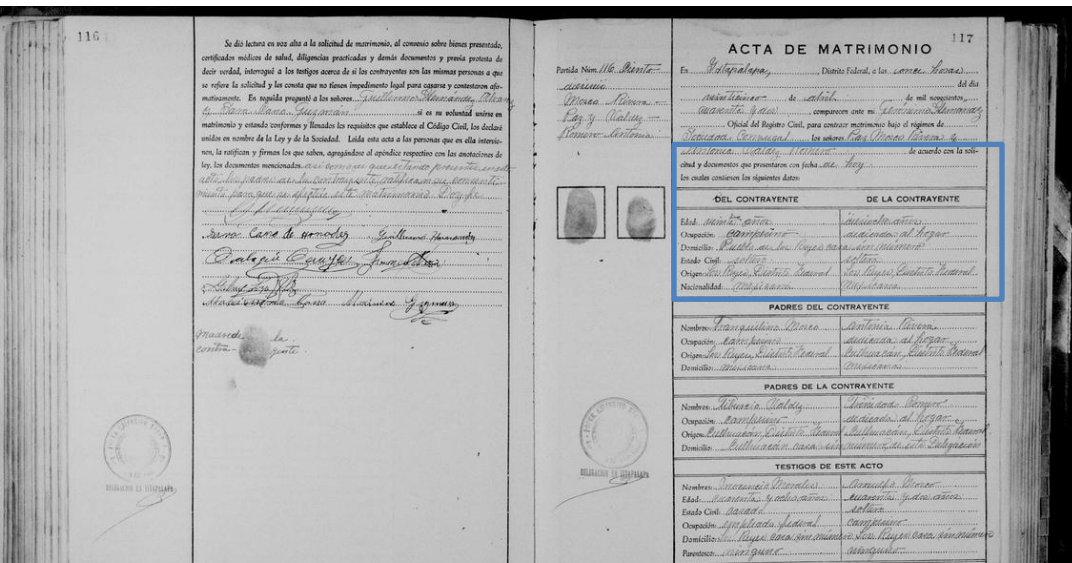
Travaux sur la presse ancienne

- Description de structure hiérarchique
 - Article, titre, paragraphes, tableaux, illustrations
- Fusion de sortie d'OCR à différentes résolutions



Analyse de pages d'actes de mariage mexicains

- Localisation de champs logiques basée sur extraction de POI,
- Inférences de règles par l'analyse de redondances dans des masses de données

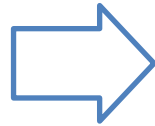


PLAN

- Exemples d'analyse d'images de documents avec DMOS
- **Fonctionnement de la méthode DMOS**
- A vous de jouer !

Schéma global

Indices
visuels

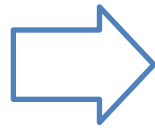


Description
des
connaissances

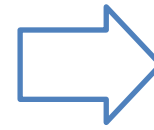


Phase de
compilation

Indices
visuels



Description
des
connaissances

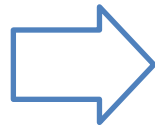


Phase de
compilation

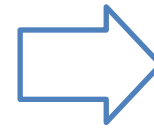
■ Nature des indices visuels

- Primitives extraites à différentes résolutions d'images
 - Composantes connexes, segments, lignes de texte
- Sorties de systèmes de reconnaissance de structure élaborés
- Résultats d'OCR
- Sorties de reconnaisseurs d'écritures
- Informations provenant d'autres pages de la collection
- Champs localisés par word spotting
- Résultats d'interactions avec l'utilisateur
- Données numériques issues de documents pdf
- ...

Indices
visuels



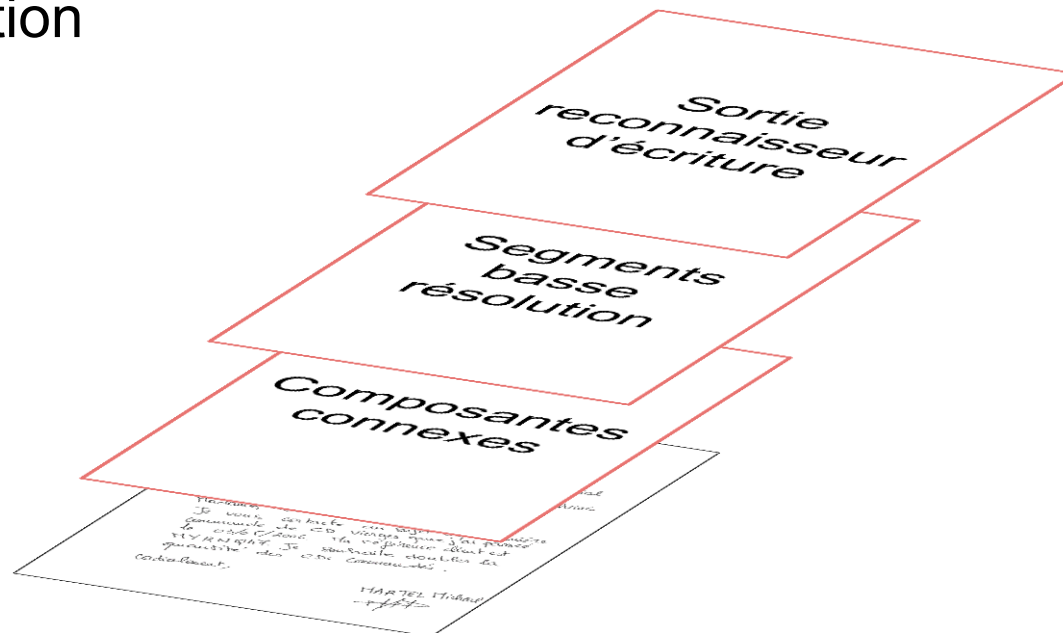
Description
des
connaissances



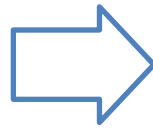
Phase de
compilation

■ Organisation en calques perceptifs

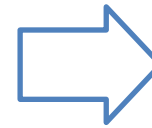
- Structure contenant les primitives
- Coordonnées dans un unique repère (image)
- Autant de calques que de points de vue sur le documents
- Chaque règle précise le calque dans lequel elle pioche l'information



Indices
visuels



Description
des
connaissances



Phase de
compilation

Langage grammatical EPF

- Connaissances sur le contenu physique
- Organisation logique des documents

pageCourrier ::=

AT(hautGauchePage) &&

coordonneesExp &&

AT(milieuPage) &&

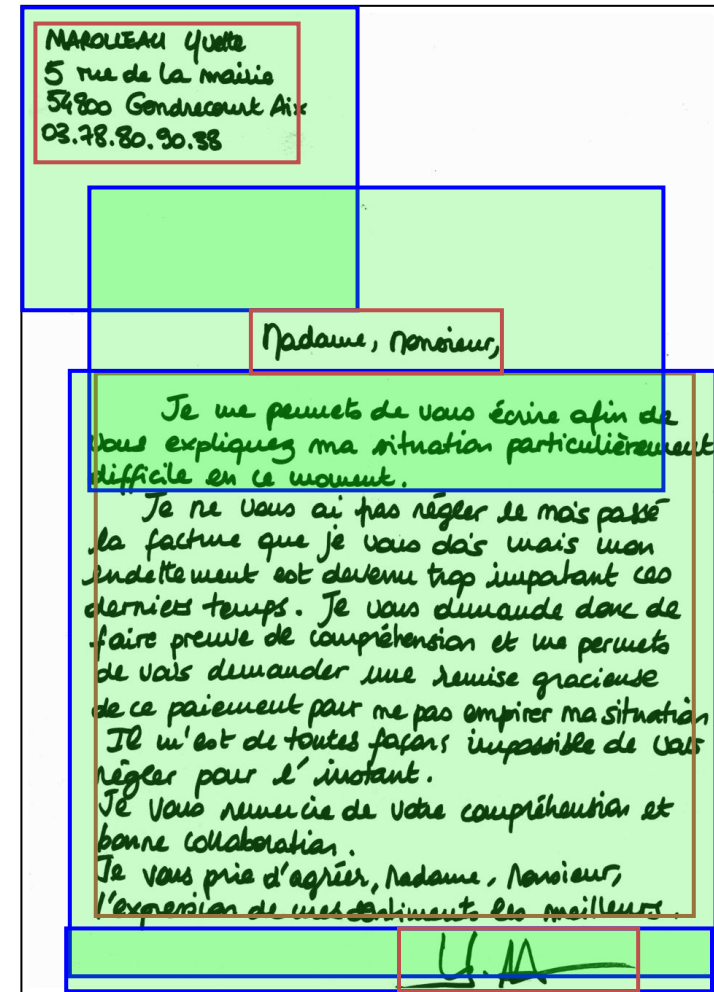
ouverture &&

AT(sousOuverture) &&

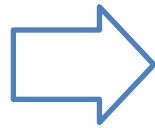
corps &&

AT(sousCorps) &&

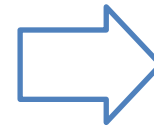
signature.



Indices
visuels



Description
des
connaissances



Phase de
compilation

■ Langage grammatical EPF

- Stratégie de coopération entre différentes sources d'indices visuels disponibles
 - Ex : mécanismes perceptifs

ligneDeTexte ::=

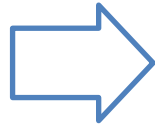
USE_LAYER(resolutionBasse) FOR (segment) &&

AT(zoneSegment) &&

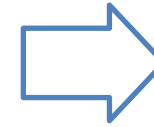
USE_LAYER(resolutionHaute) FOR (suiteDeLettres).



Indices
visuels



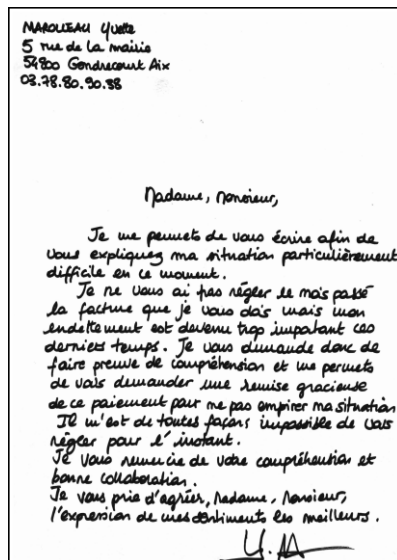
Description
des
connaissances



Phase de
compilation

■ Compilateur du langage EPF

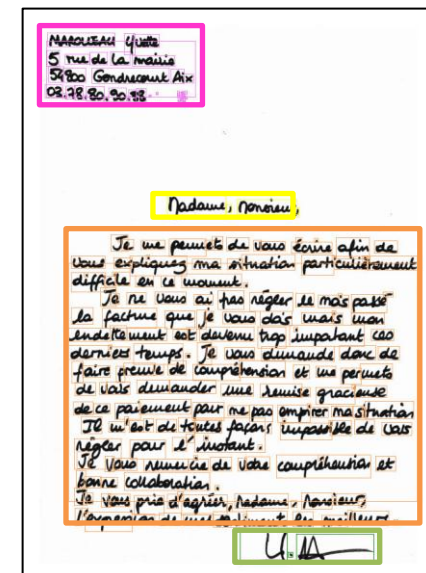
- Génération automatique du système de reconnaissance
 - pour le type de document associé à la description grammaticale



Description EPF

Compilation

Système
reconnaissance



PLAN

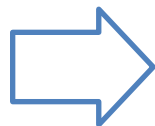
- Exemples d'analyse d'images de documents avec DMOS
- Fonctionnement de la méthode DMOS
- **A vous de jouer !**

Quel intérêt ?

- **Fusion d'information issues de (vos !) sources variées**
 - Indices visuels extraits de l'image
 - Résultats de reconnaissance d'écriture, de symboles
 - Hypothèses de segmentation
 - Sorties de systèmes complets
 - Mécanismes d'interaction utilisateur
- **Guidée par des connaissances explicites**
 - Description grammaticale de la connaissance physique, logique et sémantique du contenu
 - Fusion permise grâce au mécanisme de calques perceptifs

⇒ **Gain en performance et en généricité**

Indices
visuels



Description
des
connaissances



Phase de
compilation

Environnement de développement complet

- **Machine virtuelle**
 - **Environnement de développement Eclipse EPF - λ Prolog**
 - **Bibliothèque d'extraction de primitives standard**
 - Binarisation, segments, composantes connexes, ligne de texte...
 - **Compilateur DMOS**
 - **Visualisateur de résultats**
 - Plans graphiques
 - Etapes intermédiaires, résultat final
- ⇒ **Mise à disposition de la communauté pour des fins recherche**

DocRead - CHART/src/CHART_flowchart.gram - Eclipse

File Edit Navigate Search Project Run Window Help

Quick Access C/C++ DebugDocRead DocRead

BMS_doptim.par BMS_position.pm BMS_condTERM.pm BMS_actes.pm CHART_flowchart.gram

DocRead Vie Navigator

BMS [BMS master]
 BOX [BOX 3.0FindBestFirst]
 CHART [CHART master]
 bin
 data
 CHART_images.par
 writer1_1.inkml
 writer1_1.mem
 writer1_1.res.lg
 doc
 lib
 main
 CHART_action.pm
 CHART_action.pm.pmcdep
 CHART_read.pm
 CHART_read.pm.pmcdep
 MainDMOS.cpp
 MenuAction.c
 MenuAction.h
 obj
 OutilsDeveloppement
 src
 CHART_condTERM.pm
 CHART_condTERM.pm.pmcdep
 CHART_flowchart.gram
 CHART_flowchart.pm
 CHART_flowchart.syst
 CHART_flowchart.type
 CHART_flowchart.type.pmcdep
 CHART_outils_attributs.pm
 CHART_outils_attributs.pm.pmcdep
 CHART_position.pm
 CHART_position.pm.pmcdep
 CHART_struct_prod.pm

```

255
256
257 %Suite diagramme dépend du symbole courant
258
259 suiteDiagramme E1@(terminator _ _ _ _ CCE) [Fleche|Reste] ::=
260   AT(sousSymbole E1) &&
261   uneFleche Fleche &&
262   AT(autreBoutFleche E1 Fleche) &&
263   :>(acces_Position Anc Zone) &&
264   USE_LAYER (nomResol "CHART_construits") FOR(uneBoite Anc Zone B Symbdjvu)&&
265   ``(!)&&
266   sauveFleche Fleche CCE B Symbdjvu &&
267   enchaîneDiagramme B Reste.
268
269
270
271 suiteDiagramme (terminator _ _ _ _ ) [] ::=
272   ``(!).
273
274 suiteDiagramme E1@(connection _ _ _ _ CCE) [Fleche|Reste] ::=
275   AT(sousSymbole E1) &&
276   uneFleche Fleche &&
277   AT(autreBoutFleche E1 Fleche) &&
278   :>(acces_Position Anc Zone) &&
279   USE_LAYER (nomResol "CHART_construits") FOR(uneBoite Anc Zone B Symbdjvu)&&
280   ``(!) &&
281   sauveFleche Fleche CCE B Symbdjvu &&
282   enchaîneDiagramme B Reste.
283
284
285 suiteDiagramme (connection _ _ _ _ ) [] ::=
286   ``(!).
287
288 suiteDiagramme E1@(data _ _ _ _ CCE) [Fleche|Reste] ::=
289   AT(sousSymbole E1) &&
290   uneFleche Fleche &&
291   AT(autreBoutFleche E1 Fleche) &&
292   :>(acces_Position Anc Zone) &&
293   USE_LAYER (nomResol "CHART_construits") FOR(uneBoite Anc Zone B Symbdjvu)&&
294   ``(!) &&
295   sauveFleche Fleche CCE B Symbdjvu &&
  
```

Outline

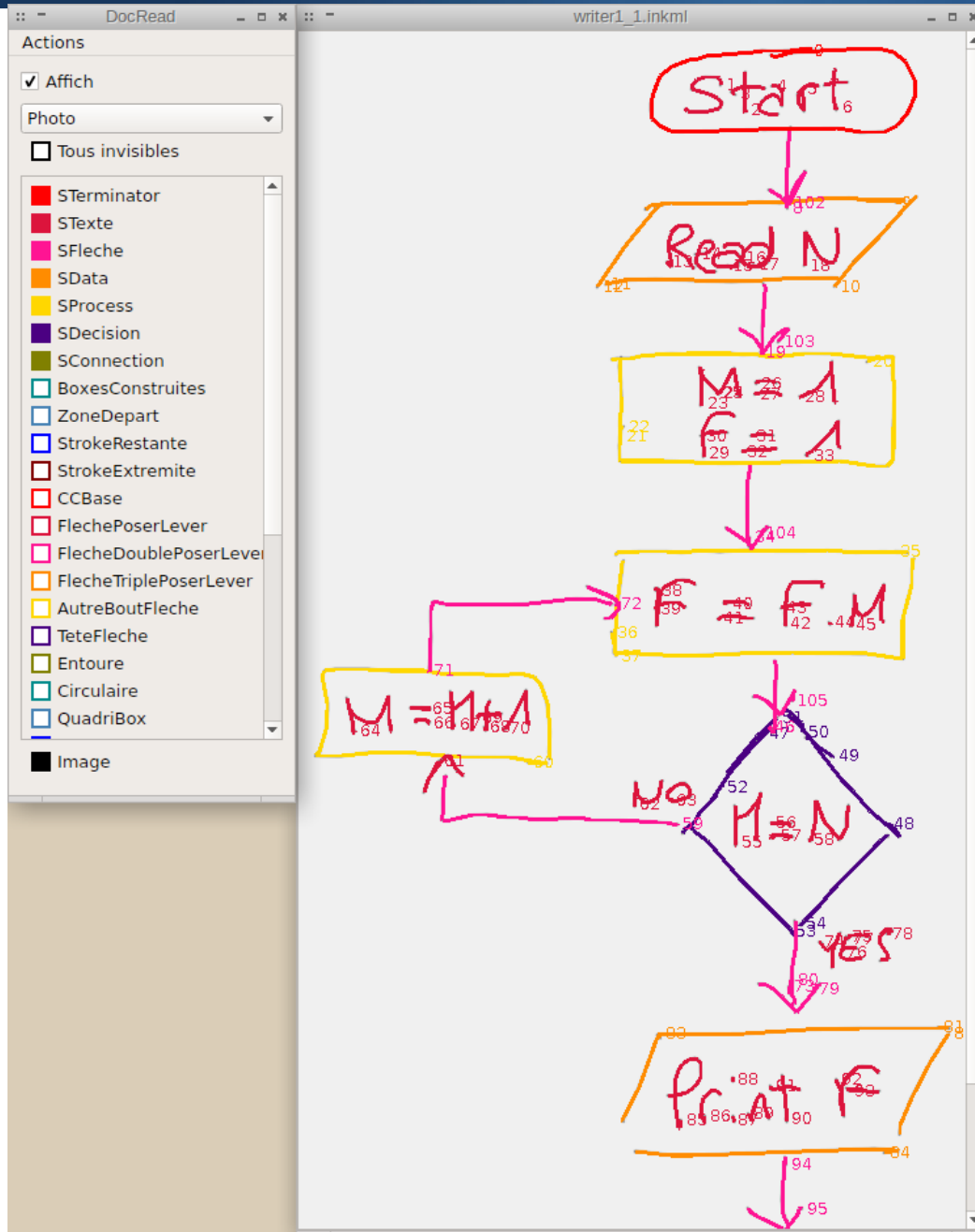
- analyseStrokesRestantes
- associeLabelsBox (2)
- associeSegmentsListeStroke (2)
- associeSegmentsStroke (2)
- chargerCalqueReco (2)
- consommeSegConstruits (2)
- consommeSegmentsEnferme
- debutAlignementCC (2)
- diagramme (chart LChart)
- duTexte (2)
- duTexteEnferme Symb T
- ecritResultat Fichier
- enchaîneDiagramme (2)
- extraireSymboles (2)
- extraitStrokesCotes (2)
- extraitStrokesListeSeg (3)
- extremiteDeFleche F Str
- finAlignementCC (2)
- finCote (2)
- finFleche C C2
- finTexte (3)
- flowchart Fichier InputMem (
- integreMemBox (2)
- integreMemStrokes (2)
- libellerStrokesRestantes (2)
- morceauxExtremite (2)
- morceauxExtremiteBoite (2)
- saveFleche F Prec Suiv Svr

Problems Tasks Console Properties Search

<terminated> Chart1_5 [DocRead Application] /home/dmosuser/WorkspaceDocRead/CHART/bin/v-chartread_PC

Densite de numérisation : 300
 Appel de l'analyse des boites
 Création de l'ensemble des résolutions
 Première Image
 Seconde Image

Writable Insert 271 : 45



Conclusion

■ DMOS

- Fusion d'informations hétérogènes guidée par des règles
- Système complet de reconnaissance de documents

■ Mise à disposition pour un but de recherche

- Du système DMOS
- De l'environnement de développement adapté
- De notre expertise pour guider votre utilisation du système
 - En incorporant les résultats de vos briques/systèmes technologiques

■ N'hésitez pas à nous contacter !