



INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

Projet Imadoc

Interprétation et Reconnaissance d'Images et de Documents

Rennes

————— THÈME 3B —————

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	3
2.1	Objectifs	4
2.2	Méthodologies	5
2.2.1	Stratégies d'analyse de documents	5
2.2.2	Documents structurés	6
2.2.3	Modélisation de l'écriture manuscrite	6
3	Fondements scientifiques	7
3.1	Segmentation par filtrage de Kalman	8
3.2	DMOS, une méthode générique de reconnaissance de documents	9
3.2.1	DMOS : une méthode générique de reconnaissance de documents structurés	9
3.2.2	Formalisme EPF	10
3.2.3	Analyseur et gestion du bruit	11
3.3	Reconnaissance des formes par systèmes d'inférence floue	12
3.4	Modèle et système de perception et d'interprétation (SPI)	13
4	Domaines d'applications	16
4.1	Partitions musicales	16
4.1.1	Rétro-conversion	17
4.1.2	Saisie de partitions musicales	18
4.2	Formules mathématiques	18
4.2.1	Rétro-conversion	18
4.2.2	Saisie de formules mathématiques	19
4.3	Tableaux et formulaires	19
4.4	Documents anciens	20
4.5	Écriture manuscrite	21
4.6	Interfaces homme-machine orientées stylo	22
5	Logiciels	22
6	Résultats nouveaux	23
6.1	Reconnaissance de formules mathématiques	23
6.1.1	Principe	24
6.1.2	Description grammaticale	24
6.1.3	Résultats	24
6.2	Reconnaissances de la structure des registres matricules	25
6.3	Didacticiel d'aide à l'apprentissage de la lecture et de l'écriture	27
6.4	Modélisation du contexte par arbre de décision flou pour la reconnaissance de symboles graphiques	28
6.5	Comparaison de textes et recherche approchée de thèmes musicaux	30
6.5.1	Comparaison de textes	30

6.5.2	Recherche de thèmes musicaux	31
6.6	Histoire de la typographie	31
6.7	Lecture automatique de l'écriture manuscrite	32
6.8	Interface-Utilisateur-Document-Interactif	34
7	Contrats industriels (nationaux, européens et internationaux)	35
7.1	Convention de recherche avec le Ministère de la Culture	35
8	Actions régionales, nationales et internationales	35
8.1	Actions régionales	35
8.2	Actions internationales	35
9	Diffusion de résultats	36
9.1	Animation de la communauté scientifique	36
9.1.1	Comité éditorial de journaux	36
9.1.2	Comités de programme de conférences	36
9.1.3	GdR	36
9.1.4	Sociétés savantes	36
9.1.5	Listes de discussion	37
9.2	Enseignement universitaire	37
9.3	Participation à des colloques, séminaires, invitations	37
10	Bibliographie	38

1 Composition de l'équipe

Responsable scientifique

Jean Camillerapp [professeur Insa]

Assistante de projet

Édith Blin [TR Inria]

Personnel Inria

Jacques André [DR]

Personnel Insa

Éric Anquetil [maître de conférences]

Bertrand Coüasnon [maître de conférences]

Ivan Leplumey [maître de conférences]

Personnel Université de Rennes 1

Guy Lorette [professeur]

Charles Quéguiner [maître de conférences]

Hélène Richy [maître de conférences, en détachement du CNRS]

Laurent Pasquer [ATER depuis le 1/9/2000]

Chercheur doctorant

Nicolas Ragot [boursier université depuis le 1/10/2000]

2 Présentation et objectifs généraux

En dépit d'annonces insistantes et répétées de l'avènement d'une société *sans papier*, la multitude de documents papier existants pose encore d'énormes problèmes de transfert sur un support électronique.

Le document électronique permet une recherche par le contenu, un transport très rapide, un archivage et une gestion beaucoup plus aisée que le papier. Par contre un document sous forme papier permet à l'utilisateur de disposer d'une plus grande autonomie, d'une possibilité de lecture détaillée plus confortable et de possibilités d'annotations plus grandes.

Ces deux formes de représentation des documents sont complémentaires, et il est nécessaire de pouvoir passer facilement d'une forme à une autre. S'il s'agissait, soit d'un simple passage du document papier à une image bitmap, soit d'imprimer le contenu de l'image d'un document électronique, le problème serait trivial. En réalité, les problèmes scientifiques sous-jacents au traitement de documents sont beaucoup plus complexes. Les objectifs sont de développer des systèmes informatiques capables d'effectuer de manière automatique :

- la *synthèse de documents*, c'est-à-dire le passage du document de sa forme conceptuelle à sa réalisation définitive sous forme électronique,

- de l'*analyse de documents*, c'est-à-dire le passage du document papier à la compréhension des informations qu'il contient et par exemple de pouvoir faire le lien entre le texte, les graphiques et les images, de pouvoir faire des recherches par le contenu, de pouvoir jouer une œuvre musicale à partir d'une partition imprimée ou manuscrite [?].

2.1 Objectifs

Les recherches menées au sein du projet Imadoc concernent *l'écrit et le document* sous toutes leurs formes (manuscrit, imprimé, image, graphique, multimédia, etc.) ainsi que les activités qui y sont liées, notamment la production de nouveaux documents hypertextes ou multimédia interactifs, la transformation sous forme électronique élaborée de documents papier existants et leur traitement « intelligent », ainsi que l'*Interaction Homme-Document (IHD)*.

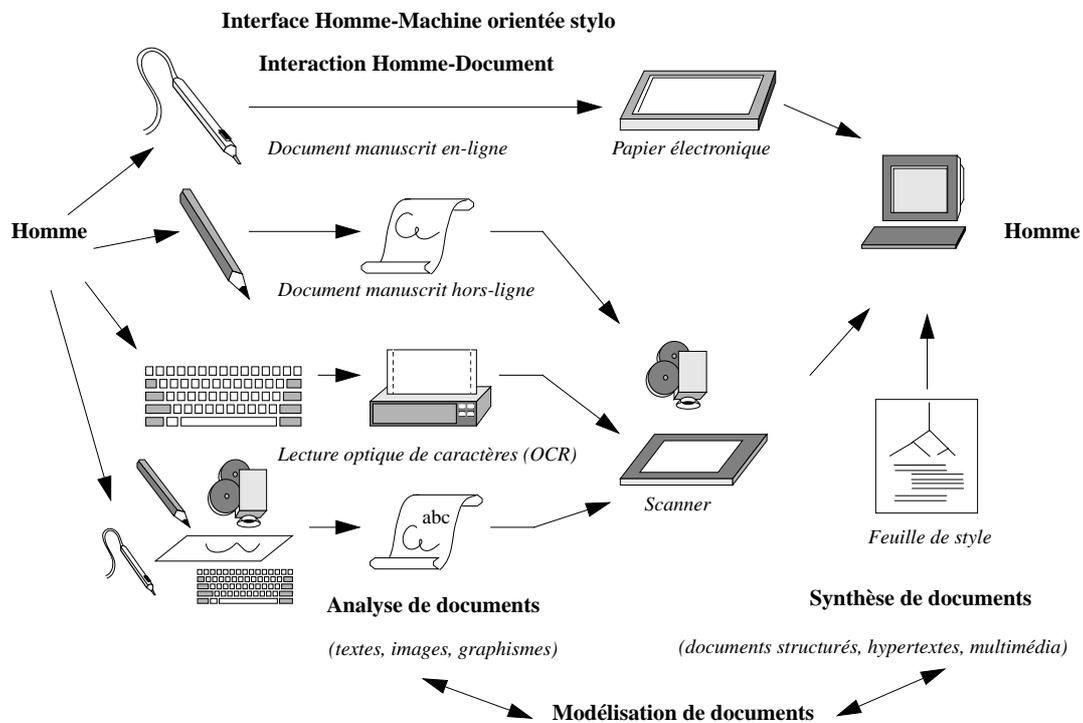


FIG. 1 – *Communication écrite.*

De manière plus générale, les centres d'intérêt du projet touchent à la *communication écrite* sous un double aspect : *analyse de documents* et *interaction homme-document* (cf. figure 1) avec souvent la notion sous-jacente de document structuré.

Analyse de documents : « du document papier à sa compréhension extraite de sa forme électronique »

Afin que l'ordinateur aboutisse à une compréhension fiable des informations contenues dans un document, il s'agit d'étudier des systèmes qui soient capables d'analyser, de manière automatique, les

signaux ou les images de documents ^[OK95], de reconnaître tous les éléments (texte, images, graphiques, figures, ...) qui constituent un document et d'analyser les relations mutuelles entre ceux-ci ^[BW97]. Pour aboutir à une interprétation finale correcte et robuste, il faut notamment que ces systèmes utilisent intelligemment plusieurs types de connaissances (connaissances a priori, connaissances contextuelles, etc.).

De plus, la communication écrite étant une activité humaine, il faut que les systèmes de traitement de documents soient capables de gérer des connaissances imprécises, incertaines et incomplètes, ce qui crée un degré supplémentaire de complexité. Dans l'équipe, ces mêmes principes de recherche sur des connaissances imprécises sont appliqués également à la comparaison de manuscrits d'auteurs d'œuvres littéraires.

Les recherches sur la reconnaissance et l'analyse du document manuscrit ^[Pla93] se subdivisent traditionnellement en deux catégories selon l'origine du document : *l'analyse statique* ou *hors-ligne*, dans laquelle le document est à l'origine sur support papier, puis est numérisé au moyen d'un scanner, et *l'analyse dynamique* ou *en-ligne* dans laquelle l'auteur du document est en interaction avec la machine au travers d'une interface orientée stylo.

Interaction Homme-Document

Cette interaction se fait à l'aide de la conception de nouvelles interfaces homme-machine (IHM) orientées stylo et de nouvelles modalités d'interaction, *papier électronique*, *reconnaissance du geste graphique*, qui sont plus naturelles, plus conviviales et plus ergonomiques que l'utilisation du clavier et de la souris pour l'annotation, la modification et la correction de documents.

Celles-ci font partie des applications pour lesquelles les interfaces manuscrites peuvent offrir un meilleur confort d'utilisation, en permettant une interaction plus naturelle et plus proche de la correction manuscrite traditionnelle.

2.2 Méthodologies

2.2.1 Stratégies d'analyse de documents

Si l'action de lire, de reconnaître ou d'interpréter un document peut sembler simple et naturelle, il s'avère que cette activité est difficile à décrire et à formaliser sous la forme d'une application informatique, car elle fait intervenir de manière subjective la gestion de plusieurs contextes d'analyse. Dans ce domaine, on se trouve en présence d'un paradoxe bien connu : *pour reconnaître il faut segmenter le signal d'entrée, mais pour bien segmenter il faut avoir reconnu* ^[Say73].

L'existence de ce paradoxe implique donc qu'on ne peut pas se restreindre à une analyse ascendante, c'est-à-dire du signal vers le document interprété, ou au contraire à une analyse complètement descendante. Les mécanismes de reconnaissance que nous utilisons vont donc enchaîner des actions de segmentation et des actions de regroupement en mettant en œuvre des stratégies de génération et de validation d'hypothèses. L'approche développée dans l'équipe consiste également à essayer de mettre en

[OK95] L. O'GORMAN, R. KASTURI, *Document Image Analysis*, IEEE Computer Society Press, 1995.

[BW97] H. BUNKE, P. WANG, *Handbook of Character Recognition and Document Analysis*, World Scientific, 1997.

[Pla93] R. PLAMONDON, « Special Issue on Handwriting Processing and Recognition », *Pattern Recognition* 26(3), 1993.

[Say73] M. SAYRE, « Machine Recognition of Handwritten words. A project report », *Pattern Recognition* 5, sep 1973, p. 213-228.

évidence les informations contextuelles utilisées et d'en obtenir une description informatique simple et bien séparée des procédures qui les utilisent.

La première segmentation est une *segmentation physique*. Elle s'appuie sur des propriétés simples du signal : composantes connexes d'une image binaire, points singuliers dans la trajectoire du stylo, etc. Ces propriétés qui n'utilisent qu'une information contextuelle très pauvre et pratiquement indépendante de l'application permettent d'aboutir à une décomposition assez sûre du signal d'entrée, décomposition sur laquelle les étapes suivantes de regroupement et de *segmentation logique* pourront s'appuyer, sans remise en question majeure.

2.2.2 Documents structurés

Bien que déjà ancienne, la notion de document structuré est sous-jacente aux « langages » tels que HTML, XML, etc. sur le Web. L'approche générale est comparable à celle utilisée en génie logiciel pour l'édition des programmes : un document est d'abord considéré comme une structure abstraite construite selon le modèle d'une structure générique (d'une grammaire, dans le cas des programmes). La première structuration évidente concerne la hiérarchie arborescente des documents : livre, chapitre, section, sous-section, paragraphe, liste, élément de liste, etc. Il en est de même pour les objets de tout type contenus dans le document et qui s'intègrent à sa structure globale : formules mathématiques, tableaux, schémas, tables d'index, bibliographie, partitions musicales, registres militaires, etc.

En même temps qu'on cherche à représenter les documents et leur contenu [11] selon un modèle structurel abstrait, on prête une grande attention aux problèmes de fond et de forme (mise en page ou en écran, fontes, code typographique, etc.) tout en essayant de moderniser ces concepts qui ont parfois mal vieilli.

Enfin, tout comme la structure logique d'un document permet d'induire sa structure physique à partir d'une feuille de style, elle permet aussi de vérifier la cohérence des documents, par exemple sa correction typographique. La correction de documents électroniques fait partie des applications pour lesquelles les interfaces manuscrites peuvent offrir un meilleur confort d'utilisation, en permettant une interaction plus naturelle et plus proche de la correction manuscrite traditionnelle. Cette étude est donc un point de convergence des études sur la reconnaissance des documents et de celles sur les documents structurés.

2.2.3 Modélisation de l'écriture manuscrite

Dans le modèle que nous proposons, le tracé d'un mot se décompose en un ensemble d'allographes¹ concaténés. Dans chaque allographe, on distingue les constituants associés au corps de l'allographe et les zones (amorce et terminaison) situées en début et en fin du tracé de l'allographe, qui composent la liaison entre deux allographes juxtaposés. Ces zones de liaison peuvent être invisibles et se réduire, dans certains cas, à un simple lever de stylet.

Dans une phase d'analyse plus fine, on extrait de chaque allographe les structures fondamentales (traits descendants significatifs), leur environnement graphique local dit contexte morphologique (deuxième niveau de modélisation), et les zones de liaison (troisième niveau de modélisation) qui peuvent être, là aussi, invisibles et correspondre à un lever de stylet (cf. figure 2).

1. Le mot allographe désigne l'un des tracés possibles d'une lettre.

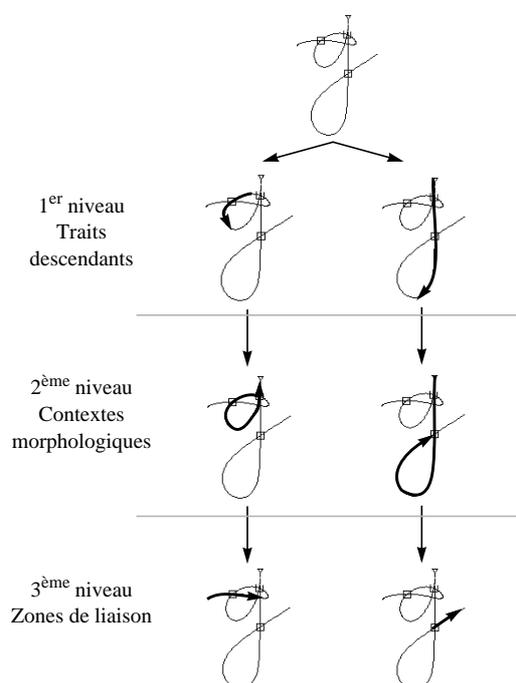


FIG. 2 – Les trois niveaux de modélisation pour la classe de la lettre g.

Cette décomposition reflète la modélisation des lettres manuscrites dont les paramètres sont établis à partir d'un apprentissage automatique. Plus précisément, la détection et la modélisation des différentes zones pertinentes de tracé (appelées primitives) sont établies par *classification non supervisée* de type *possibiliste* (cf. section 3.3) à partir d'un ensemble de graphèmes extraits du tracé de lettres isolées. Ces graphèmes sont construits à partir de la détection de points d'ancrage visuel correspondant aux singularités du tracé : points de forte variation angulaire, intersections, etc. [?]. Les primitives ainsi modélisées sont représentées sous la forme de *prototypes* formalisés de manière qualitative par des *règles floues*. Chaque classe de lettres est alors décrite par un ensemble de *systèmes d'inférence floue* (cf. section 3.3) hiérarchisés d'après la *robustesse* des connaissances modélisées. On obtient ainsi une modélisation compacte et explicite de chaque classe de lettres. Les différents constituants du tracé, mis en évidence dans ces trois niveaux de décomposition, à travers le modèle logique que nous venons de décrire, composent une base d'informations sur laquelle peut s'appuyer l'analyse logique du tracé d'un mot pour développer une méthodologie de reconnaissance.

3 Fondements scientifiques

Résumé : *Pour faire face au paradoxe reliant segmentation et reconnaissance nous cherchons à combiner des mécanismes de segmentation et des mécanismes de reconnaissance (section 2.2.1). La détection des segments de droite est un élément important dans le traitement de certains types de documents (formulaires, partitions musicales, ...). Le filtrage de Kalman s'avère être une approche relativement robuste aux erreurs de seg-*

mentation, il est à la base de leur reconnaissance. L'enchaînement de la segmentation et de la reconnaissance se fait selon deux approches : la méthode DMOS pour la prise en compte des connaissances sur la structure du document, l'utilisation des cycles perceptifs pour l'intégration de plusieurs niveaux de connaissances. Les systèmes d'inférence floue permettent d'ajuster les paramètres des modèles lors d'une phase d'apprentissage. Ils ont l'avantage de laisser accessibles et interprétables leur processus de reconnaissance.

3.1 Segmentation par filtrage de Kalman

Mots clés : filtrage de Kalman.

Résumé : Les traitements d'images liés à la reconnaissance de documents ont souvent besoin d'extraire des structures filiformes, c'est-à-dire des objets dont la largeur est très faible par rapport à leur longueur. Dans les images de documents, ces structures ne sont généralement pas isolées mais se croisent ou se chevauchent, ce qui en complique la détection. Pour les identifier, nous avons proposé d'utiliser le formalisme du filtrage de Kalman qui fournit un cadre cohérent permettant de bien faire la séparation entre la définition des caractéristiques des objets à reconnaître et les procédures de segmentation.

Les structures linéaires tiennent une grande place dans les documents et elles sont généralement à tendance horizontale ou verticale ; aussi peut-on se ramener à une analyse de l'image soit colonne par colonne, soit ligne par ligne. Les structures linéaires présentent une cohérence forte, il est donc possible de prédire leurs caractéristiques dans une colonne (ou dans une ligne) à partir des positions dans les colonnes précédentes. Nous utilisons pour cela le formalisme de Kalman dont on pourra trouver un exposé dans [Sor85] et une description de son application au traitement de documents dans [?].

Dans la colonne k , chaque structure linéaire est représentable par un vecteur d'état $S(k)$ qui comporte l'ordonnée dans la colonne, l'épaisseur, la pente et éventuellement une luminosité pour les images en niveaux de gris.

Ce vecteur évolue en fonction de la colonne analysée k selon l'équation :

$$S(k+1) = A.S(k) + W$$

dans laquelle A est la matrice caractérisant l'évolution du vecteur d'état S d'une colonne à l'autre et W un bruit de moyenne nulle.

Cette structure linéaire, si elle était isolée, produirait dans l'image une trace qui se traduirait dans chaque colonne par un empan - ensemble de pixels connexes de la colonne - généralement plus sombre que son voisinage. Les caractéristiques de cet empan fournissent le vecteur de mesure $X(k)$ qui est relié à l'état $S(k)$ par l'équation :

$$X(k) = C.S(k) + N$$

dans laquelle C est la matrice permettant de passer de l'espace des états à celui des mesures et N le bruit de mesure.

[Sor85] H. SORENSON, *Kalman filtering*, IEEE Press, 1985.

Pour l'analyse de la colonne suivante, le filtrage de Kalman permet de prédire l'état $\hat{S}(k+1)$ et les caractéristiques de l'empan associé $\hat{X}(k+1)$, ainsi que la matrice de covariance $\hat{H}(k+1)$ de l'erreur de prédiction.

La véritable difficulté de cette approche ne réside pas dans ces équations, mais dans le processus d'association des empan extraits d'une colonne de l'image avec les empan prédits. En effet, les structures linéaires se coupent et se chevauchent, ce qui perturbe les traces dans l'image. Il n'y a donc pas une correspondance biunivoque entre les états caractérisant les structures linéaires et les mesures. Nous avons donc ajouté un mécanisme de contrôle qui, en utilisant les matrices de covariance H , est capable de détecter les empan qui proviennent de la fusion de plusieurs structures, et qui associe les empan ou des parties d'empan avec les empan prédits. Ce mécanisme de contrôle détecte également l'apparition de nouvelles structures en créant un nouvel état quand il existe des mesures qui ne sont pas interprétables par les prédictions.

Ce formalisme a été appliqué pour :

- la détection des lignes de portées lors de l'analyse des partitions musicales ;
- la détection des traits délimitant des tableaux ;
- la détection et la suppression des lignes de base dans des images de chèques en niveaux de gris.

3.2 DMOS, une méthode générique de reconnaissance de documents

Mots clés : analyse structurelle, grammaire, Definite Clause Grammar (DCG), segmentation, gestion des connaissances a priori, reconnaissance de documents, partitions musicales, formules mathématiques, tableaux, formulaires.

Résumé : *Afin de tenter de résoudre les problèmes de segmentation que l'on rencontre en reconnaissance de documents, nous proposons une méthode générique de reconnaissance de document, baptisée DMOS. Cette méthode est constituée du formalisme EPF permettant de modéliser grammaticalement la connaissance a priori, et d'un analyseur associé capable de remettre en cause, au cours de l'analyse, la structure analysée tout en gérant correctement le bruit très présent en reconnaissance de documents.*

Cette méthode permet de réaliser une segmentation contextuelle des documents à forte structure. L'introduction du contexte améliore la qualité de la segmentation, donc de la reconnaissance. De plus, la méthode étant générique, il est possible, en changeant simplement la grammaire, de l'adapter à un nouveau type de documents. Ainsi, nous travaillons avec cette méthode sur la reconnaissance de partitions musicales, de tableaux, de formulaires et de formules mathématiques.

3.2.1 DMOS : une méthode générique de reconnaissance de documents structurés

Dans le cadre de la reconnaissance optique de documents, nous nous sommes fixé deux objectifs principaux :

- éviter à l'utilisateur une relecture de l'ensemble du document pour y détecter et corriger d'éventuelles erreurs résiduelles ;

- permettre au système de s'adapter rapidement à de nouveaux types de documents sans avoir à redévelopper une chaîne complète.

Le premier objectif nécessite de la part du système une forte fiabilité. Celle-ci peut s'obtenir, d'une part, en améliorant la qualité de la reconnaissance, notamment en résolvant les problèmes de segmentation et, d'autre part, en faisant détecter par le système lui-même les régions comportant des erreurs de reconnaissance. Dans ces deux cas, il est nécessaire d'utiliser la connaissance *a priori*, qui permet de régler certains problèmes de segmentation et de modéliser la *redondance*, ce qui autorise une détection d'erreurs.

Nous avons proposé, pour des documents à forte syntaxe dans lesquels des règles d'écriture peuvent être connues, une méthode baptisée DMOS (Description avec MODification de la Segmentation), constituée d'un formalisme grammatical de position EPF (*Enhanced Position Formalism*) permettant de modéliser la connaissance, et d'un analyseur associé autorisant une modification en cours d'analyse de la structure analysée. Cette modification permet d'introduire le contexte (niveau symbolique) dans la phase de segmentation (niveau numérique), afin d'améliorer la reconnaissance. En outre, cette méthode est chargée de faire appel à un classifieur pour reconnaître les symboles pouvant être assimilés à des caractères. Ainsi, des hypothèses de segmentation peuvent être produites grâce au contexte puis validées par le classifieur.

Le second objectif, c'est-à-dire la réalisation d'un système générique de reconnaissance, est atteint grâce à la méthode DMOS car elle offre l'avantage de séparer la connaissance (décrite sous la forme d'une grammaire) des outils de traitement [?], et de produire automatiquement l'analyseur par compilation de la grammaire. Cette décomposition facilite en effet largement la maîtrise de l'introduction de connaissances complexes et l'adaptation à différents types de documents.

La méthode DMOS a ainsi permis de développer des systèmes de reconnaissance de partitions musicales, de tableaux, de formulaires et de formules mathématiques, en définissant uniquement une description grammaticale du document.

3.2.2 Formalisme EPF

La grammaire d'un langage de programmation constitue une formalisation du langage en une dimension et est de ce fait insuffisante pour décrire les connaissances *a priori* sur un document bidimensionnel.

Nous proposons donc le formalisme EPF (*Enhanced Position Formalism*), qui permet de définir des positions relatives sans aucune limitation. Ainsi, pour définir le positionnement de deux objets graphiques A et B, il faut utiliser un opérateur de position AT de la manière suivante :

A AT(pos) B

Il signifie : on a le terminal ou non-terminal A et à la position pos relativement à A, on trouve le terminal ou non-terminal B.

Cependant, le sens de lecture de la grammaire impose une relation monodirectionnelle entre deux objets. Une fois B positionné par rapport à A, il est impossible de positionner en plus A par rapport à B. Ainsi la règle :

A AT(pos1) B AT(pos2) A

ne peut définir une relation bidirectionnelle entre A et B car le premier et le second A sont différents. En effet, chaque A dans la règle de grammaire représente une instance différente du terminal ou non-terminal A.

Pour pouvoir représenter la même instance de A, nous proposons de « sauvegarder » ($-->$) une instance d'un terminal ou d'un non-terminal. Cette « sauvegarde » de A offre alors la possibilité d'utiliser et de faire référence ($<--$) à A autant de fois que nécessaire :

```
(A ---> etiqRegleA) AT(pos1) B AT(pos2) (A <--- etiqRegleA)
```

Le terminal ou non-terminal ainsi « sauvegardé » peut bien entendu être utilisé dans une autre règle de grammaire. Le formalisme EPF permet donc de définir des relations complètement bidimensionnelles et bidirectionnelles sans aucune limitation.

3.2.3 Analyseur et gestion du bruit

Le formalisme EPF décrit ci-dessus permet de définir grammaticalement le document à reconnaître. De cette grammaire nous produisons automatiquement un analyseur qui possède des caractéristiques spécifiques à l'analyse de documents bidimensionnels. Ainsi, nous pouvons souligner les trois principales caractéristiques de l'analyseur que nous avons développé (à deux dimensions), par rapport à un analyseur classique (à une dimension) pour les langages formels :

- remise en cause de la structure analysée en cours d'analyse (pour effectuer des segmentations contextuelles) ;
- détection de l'élément suivant à analyser. En effet pour les analyseurs classiques, l'élément suivant est simplement celui qui est en tête de la chaîne analysée alors qu'en deux dimensions, l'élément suivant peut être n'importe où dans l'image, donc n'importe où dans la structure analysée. Afin de le trouver, l'analyseur que nous avons développé utilise les opérateurs de position du formalisme EPF pour définir une zone de l'image où l'élément suivant peut être trouvé. Ce sont les terminaux qui sont chargés de le trouver dans cette zone ;
- gestion correcte du bruit. Contrairement aux analyseurs classiques où la structure analysée est peu bruitée, en reconnaissance de documents il est nécessaire que l'analyseur soit capable de reconnaître le maximum d'informations dans un flux très bruité. Nous pouvons considérer que la gestion du bruit correspond à trouver l'élément suivant, malgré le bruit. Pour ce faire nous intervenons à deux niveaux sur l'analyseur :
 - pour un terminal, nous utilisons une pré-condition permettant de passer les éléments qui ne sont pas à reconnaître, soit le plus souvent du bruit. Elle permet de rechercher, dans la zone définie par les opérateurs de position, l'élément le plus proche possédant cette pré-condition ;
 - pour un non-terminal, l'analyse par défaut essaye chacune des règles sur l'élément courant jusqu'à ce que la première d'entre elles s'applique. Pour gérer le bruit, nous avons défini un opérateur qui effectue l'opération transposée, c'est-à-dire qui essaye une règle sur chacun des éléments de la structure jusqu'à ce qu'elle s'applique ou jusqu'à une condition d'arrêt soit atteinte :


```
FIND(règle) UNTIL(condition-d'arrêt).
```

En conclusion, nous sommes arrivés, grâce à la définition et l'implantation du formalisme EPF et de l'analyseur associé (gérant le bruit tout en modifiant la structure analysée en cours d'analyse), à concevoir un système générique de reconnaissance de documents structurés.

3.3 Reconnaissance des formes par systèmes d'inférence floue

Mots clés : modélisation explicite des connaissances, systèmes d'inférence floue, apprentissage, classification.

Résumé : *Dans le domaine de la reconnaissance de forme, la modélisation de systèmes complexes engendre généralement l'interaction de plusieurs processus interdépendants. Face à la complexité d'interaction des différents maillons qui composent un tel système, il est important de pouvoir maîtriser chacun des concepts mis en jeu. Dans cet objectif, les systèmes d'inférence floue (SIF) permettent la conception de systèmes de reconnaissance interprétables basés notamment sur une modélisation explicite des connaissances.*

Pour la plupart des problèmes réels de reconnaissance de forme, la conception automatique d'une modélisation explicite constitue un problème particulièrement complexe par suite de la grande variabilité observée tant à l'intérieur de chaque classe de formes (variabilité intra-classes) qu'entre classes de formes différentes (variabilité inter-classes). La plupart des approches s'appuient sur la capacité d'apprentissage automatique des méthodes de type *stochastique* ou *connexionniste* en se basant respectivement, soit sur une modélisation de l'étendue de la variabilité des formes, soit sur une discrimination directe des classes. Ces approches conduisent souvent à des systèmes de reconnaissance *opaques* (type boîte noire) pour lesquels il est très difficile d'avoir une interprétation des processus de décision mis en œuvre. C'est pourquoi, même si ces approches permettent la réalisation de systèmes affichant assez rapidement des performances intéressantes, ces systèmes se révèlent peu évolutifs, il est donc difficile d'améliorer leurs performances.

Dans l'optique de pouvoir réaliser des systèmes interprétables, nous avons mis au point une méthodologie de modélisation basée sur la génération automatique de *systèmes d'inférence floue* [?]. Plus précisément, les règles floues utilisées répondent aux hypothèses suivantes :

- les entrées $x = (x_1, \dots, x_n)$ sont non floues et définies dans l'espace à n dimensions appelé espace des attributs ou encore espace des entrées ;
- les sous-ensembles flous de sortie B_{ik} sont des singletons $\{b_{ik}\}$;
- les conclusions des règles ($Y_k = b_{ik}$) sont donc précises : Y_k représente le degré d'appartenance de la forme présentée en entrée, vis-à-vis de la règle R_i et de la classe k .

Considérons maintenant N_r règles floues formant un SIF permettant de caractériser ou de classer les formes considérées dans C classes. L'expression linguistique d'une règle floue associée à la caractérisation d'une classe k se formalise d'une manière générale de la façon suivante :

R_i : **SI** x_1 est M_{i1} **et** x_2 est M_{i2} **et** ... **et** x_n est M_{in}
ALORS la donnée appartient à la classe k **et** non aux autres classes.

Ce qui se traduit dans le formalisme classique des règles floues par :

R_i : **SI** x_1 est M_{i1} **et** x_2 est M_{i2} **et** ... **et** x_n est M_{in}
ALORS $Y_1 = \delta_{u1}$ **et** $Y_2 = \delta_{u2}$ **et** ... **et** $Y_u = \delta_{uu}$ **et** ... **et** $Y_c = \delta_{uc}$,
 où δ_{ij} désigne le symbole de Kronecker.

Chaque condition (x_j est M_{ij}) est interprétée comme le degré d'appartenance de l'observation x_j au sous-ensemble flou M_{ij} , c'est-à-dire $\mu_{M_{ij}}(x_j)$. L'inférence floue est alors définie de la manière suivante :

$$\mu_{B'_k}(y) = \perp_{i=1}^{N_r} I(\beta_i, \mu_{B_{ik}}(y)) \text{ avec } \beta_i = \mathbf{T}_{j=1}^n \mu_{M_{ij}}(x_j),$$

où \mathbf{T} est une T -norme symbolisée par le *et* utilisée dans la partie prémisse des règles, β_i représente la valeur d'activation de la règle R_i , I est l'implication floue [BM95], \perp représente l'opérateur d'agrégation et N_r est le nombre total de règles.

Le résultat obtenu en sortie du SIF mesure alors le degré d'appartenance de la forme présentée en entrée, relativement à chacune des classes.

Les systèmes d'inférence floue permettent de concilier une modélisation de nature numérique qui caractérise la plupart des problèmes réels de reconnaissance de forme, avec une modélisation robuste et qualitative des connaissances extraites.

La génération automatique des SIF repose sur une analyse non supervisée des données d'apprentissage. Pour ce faire, nous proposons d'utiliser les concepts récents de la classification possibiliste [?] afin d'extraire et de qualifier un ensemble de propriétés pertinentes. La classification non supervisée de type possibiliste permet de caractériser de manière précise les propriétés intrinsèques de chaque prototype modélisé. Ces propriétés sont alors représentées par des sous-ensembles flous M_{ij} , directement par projection sur les différents axes de l'espace de représentation. Ces différents sous-ensembles flous sont ensuite rassemblés pour former des règles floues dites prototypes qui vont modéliser de manière "lisible" ces propriétés pertinentes.

L'approche méthodologique proposée a été mise en œuvre pour la conception d'un système complet de reconnaissance en-ligne d'écriture manuscrite.

3.4 Modèle et système de perception et d'interprétation (SPI)

Mots clés : perception, interprétation, cycles perceptifs, réseaux de neurones.

Résumé : *Les problèmes de perception et d'interprétation de signaux et d'images, considérés comme des sensations auditives ou visuelles, sont actuellement un enjeu de recherche important. Il s'agit de doter la machine de nouvelles capacités analogues à celles de l'être humain. Ces capacités reposent sur la possibilité d'organiser et d'intégrer, de manière dynamique, des données situées à plusieurs niveaux de représentation de connaissances en se fondant sur des critères d'organisation perceptive et des critères de cohérence.*

Les méthodes classiques de reconnaissance de formes s'appliquent uniquement à la reconnaissance d'objets isolés à partir de leur représentation sous forme vectorielle ou structurelle. Cette reconnaissance est un processus purement *statique*. Cependant, dès le début du siècle, le *Gestaltisme* avait notamment mis en évidence le fait que la perception humaine est influencée par les objets qui se trouvent dans le voisinage immédiat de l'objet à reconnaître, c'est-à-dire par son *contexte* spatial. De plus, les recherches en sciences cognitives ont montré que la perception et l'interprétation visuelle chez

[BM95] B. BOUCHON-MEUNIER, *La logique floue et ses applications*, Addison-Wesley, 1995.

l'Homme repose sur un processus de construction *dynamique* d'une représentation mentale [Kan79], [EK00]. Ce processus consiste à organiser et à intégrer dynamiquement des données et des connaissances situées à plusieurs niveaux de représentation, en se fondant sur des critères d'organisation perceptive et des critères de cohérence globale des différentes interprétations possibles. Mais ce n'est que depuis quelques années que les problèmes de perception et d'interprétation de signaux et d'images, par ordinateur, sont devenus un enjeu de recherche important [SB94], [BS99]. Il s'agit de doter la machine de nouvelles capacités analogues à celles de l'être humain.

Nous avons étudié, au sein du projet Imadoc, une nouvelle méthode de reconnaissance de formes *dynamique et multi-contextuelle* ainsi qu'un *système de perception et d'interprétation* de formes en contexte (SPI) [10].

Le *modèle générique* développé repose sur une organisation modulaire et hiérarchique nommée *colonne*, constituée d'un empilement de N niveaux de représentation. Chacun des niveaux comporte des interactions intra-niveau et inter-niveaux avec les niveaux voisins inférieurs et supérieurs (cf. figure 3). L'interaction entre plusieurs niveaux de représentation, l'entrée de données contextuelles sur plusieurs niveaux et la fusion de données en contexte se font d'une part, par un mode de représentation unifié sous forme de couches de neurones et, d'autre part, par une succession de *cycles perceptifs* qui modifient de façon itérative les activations des neurones de chacune des couches. Il s'agit donc d'un réseau *dynamique et récurrent*. De plus, compte-tenu de son architecture, ce système est intrinsèquement parallèle et réparti.

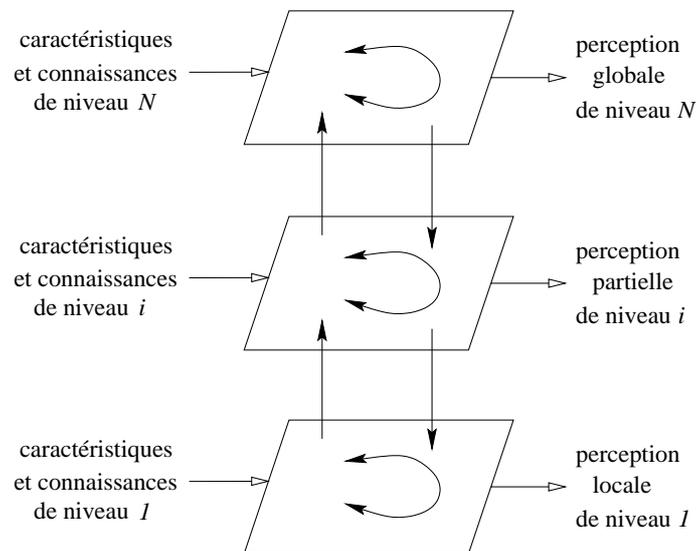


FIG. 3 – *Modèle générique d'une colonne de perception et d'interprétation.*

-
- [Kan79] G. KANIZA, *Computing Perceptual Organization in Computer Vision*, New York: Praeger, 1979.
[EK00] M. W. EYSENCK, M. KEANE, *Cognitive Psychology, A Student's Handbook*, Psychology Press, 2000.
[SB94] S. SARKAR, K. L. BOYER, *Computing Perceptual Organization in Computer Vision*, World Scientific, 1994.
[BS99] K. L. BOYER, S. SARKAR (éditeurs), *Perceptual Organization in Computer Vision*, Special Issue of *Computer Vision and Image Understanding*, vol. 76, n° 1, Academic Press Ed., 1999.

Chaque hypothèse de caractéristique issue de primitives, de sous-formes ou de formes est modélisée par un nœud-hypothèse dans un graphe valué ou dans un réseau par un neurone avec une valeur d'activation associée.

L'introduction de connaissances contextuelles dans les différents niveaux de représentation se fait en trois temps:

- création de nœuds représentatifs des connaissances disponibles *a priori*,
- introduction d'activations initiales pour ces nœuds,
- création de liens activateurs entre les nœuds modélisant des connaissances compatibles entre elles et de liens inhibiteurs entre les nœuds représentant des informations incompatibles entre elles.

Les données issues de la forme placée à l'entrée du système (sensations visuelles) sont hiérarchisées et entrées dans le système (cf. figure 3) à des niveaux différents selon leurs natures respectives. Elles sont perçues et interprétées en fonction de leurs contextes respectifs. À chacun des niveaux de représentation correspond donc une interprétation possible sur la sortie correspondante.

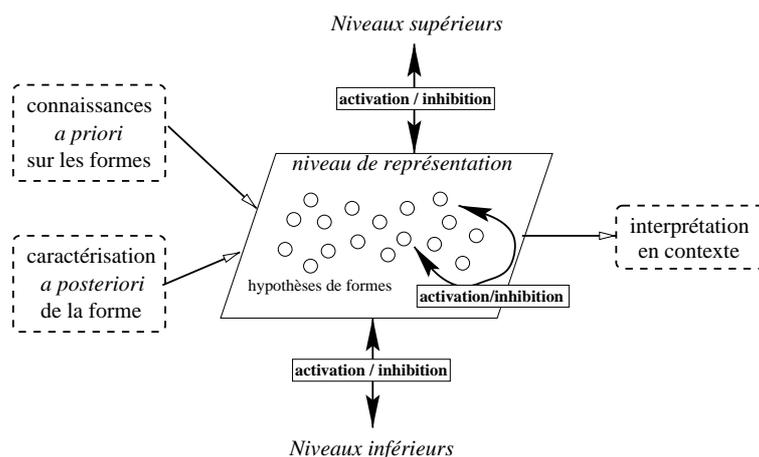


FIG. 4 – Schéma d'un niveau de représentation.

À chaque niveau (cf. figure 4), le fonctionnement du système comporte trois phases successives :

- l'extraction et l'entrée des informations issues de la forme à percevoir (*sensations visuelles*) relatives à chaque niveau de représentation,
- la fusion de ces informations avec les connaissances *a priori* (contextes intra et inter-niveaux),
- une interprétation en sortie du niveau de représentation considéré.

Les cycles perceptifs consistent en une succession de trois processus qui se répètent :

- un processus de relaxation à l'intérieur de chaque niveau qui augmente les activations latérales des informations compatibles entre elles et qui accroît les inhibitions latérales des informations incompatibles entre elles (mécanisme de perception en contexte),

- un processus ascendant qui consiste en une augmentation des valeurs d'activation ou d'inhibition associées aux nœuds du graphe d'un des niveaux de représentation en fonction de celles de la couche de niveau inférieur (mécanisme d'induction),
- un processus descendant qui propage les informations des niveaux les plus élevés vers les niveaux inférieurs également selon un processus d'activation et d'inhibition (mécanisme de déduction).

Ces mécanismes d'activation, d'inhibition et de fusion de connaissances servent à gérer la cohérence de l'ensemble des informations disponibles, chacune d'elle étant placée en contexte. Il s'agit dès lors d'une interprétation multi-contextuelle. Ceci se traduit, *in-fine*, par l'émergence d'une interprétation. Cette interprétation apparaît alors comme une *évidence* qui résulte de la *fusion cohérente* de l'ensemble des connaissances disponibles. L'interprétation est partielle si elle correspond à un niveau de représentation intermédiaire et elle est globale si elle intervient au niveau le plus élevé de la colonne.

Le fonctionnement du système possède trois types de critères d'arrêt. S'il apparaît, au niveau de représentation le plus élevé dans la colonne, un neurone dont l'activation est très supérieure à celle de tous les autres, il correspond à l'interprétation finale de la forme observée. Si aucune valeur d'activation ne dépasse les autres, la forme observée est rejetée comme étant inconnue. Si après un nombre de cycles prédéterminé il existe, dans un niveau de représentation donné, un neurone dont l'activation est très supérieure à celle de tous les autres neurones, celui-ci correspond à une interprétation partielle. Ceci permet notamment au système de généraliser et de proposer une interprétation probable d'une forme qui n'a pas été apprise, mais dont tous les éléments constitutifs sont cohérents entre eux.

Un mécanisme de *focalisation d'attention* est également mis en œuvre dans le système. Ce mécanisme consiste en une sélection dynamique des nœuds-hypothèses et des liens correspondants à partir des résultats d'une phase de pré-reconnaissance. Ceci conduit à ne considérer qu'un faisceau restreint dans l'ensemble du treillis d'hypothèses. Il s'agit donc d'un réseau *oncogénique*.

4 Domaines d'applications

Résumé : *Les interactions hommes-documents sont fréquentes : création, consultation, modification, extraction sélective d'information. Dans le cas de documents initialement créés sous forme papier une rétroconversion les mettant sous forme numérique rendra possible des traitements automatiques. Dans d'autres cas les interfaces de saisie ou de modification au clavier et à la souris peuvent se révéler peu ergonomiques. Plutôt que d'adopter une présentation par type de traitements, nous avons préféré faire apparaître les types de documents sur lesquels Imadoc a déjà une expérience. Cette classification nous paraît plus claire pour des utilisateurs potentiels et met mieux en évidence les caractéristiques de chaque type de documents.*

4.1 Partitions musicales

Mots clés : Partitions musicales, édition, Braille, bibliothèques numériques, bases de données musicales, saisie orientée stylo.

Résumé : *Les applications de la reconnaissance de partitions musicales sont importantes pour le développement de l'informatique musicale. Dans le cadre de la rétro-conversion de partitions, elles interviennent principalement dans le monde de l'édition de partitions et dans la constitution de bases musicales pour les bibliothèques numériques. La reconnaissance en ligne (utilisant un stylo numérique) permet de simplifier largement la saisie de nouvelles partitions.*

4.1.1 Rétro-conversion

La reconnaissance optique de partitions musicales est un domaine d'étude particulièrement intéressant car très représentatif des difficultés généralement rencontrées en reconnaissance de documents structurés.

En outre, ses applications directes sont relativement nombreuses. Elles s'inscrivent dans le cadre du développement de l'informatique musicale, qui manipule généralement l'information musicale à l'aide de la notation classique : la partition. Cependant, cette notation relativement complexe est difficile à saisir sous une forme numérique. De la même manière que la reconnaissance optique de caractères (*OCR*) pour les traitements de texte, la reconnaissance optique de partitions musicales permet d'éviter cette phase de saisie lente et fastidieuse.

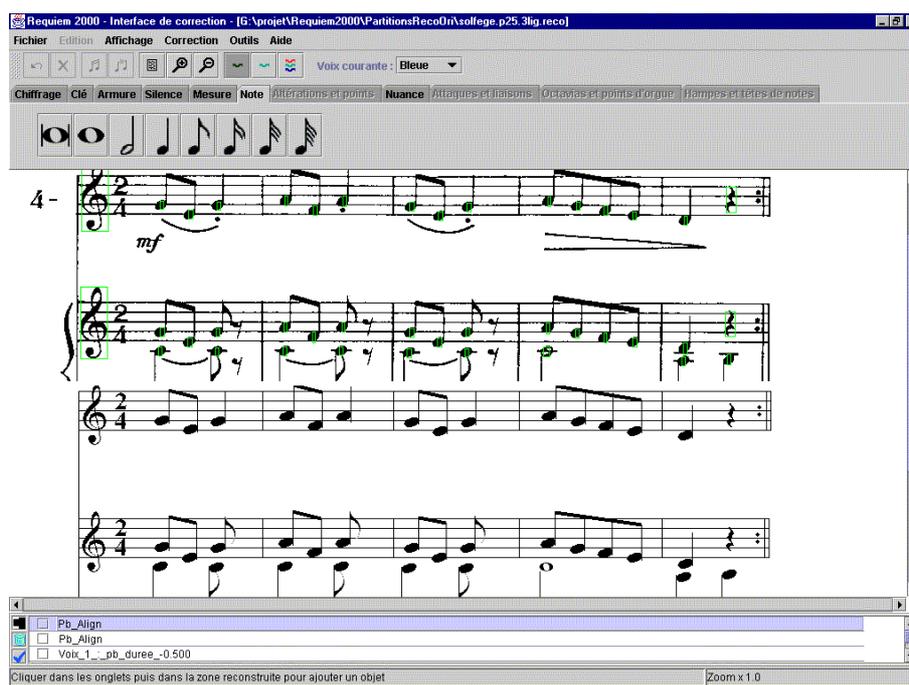


FIG. 5 – Partition reconstruite (les portées du bas) et éditable (structure pivot), résultat de la reconnaissance à partir d'une image binaire (les portées du bas).

Il est ainsi possible, grâce à un fichier au format MIDI (*Musical Interface for Digital Instruments*) généré par la phase de reconnaissance, de produire par exemple des documents sonores de très faible encombrement pour utilisation sur le Web.

Les musicologues ont également un besoin important en bases de données qui leur sont nécessaires pour analyser le style et la structure musicale (statistiques), indexer des bases de thèmes musicaux. . .

Lorsque le document analysé est une partition imprimée, les applications ayant trait à l'édition peuvent être par exemple :

- réaliser une nouvelle édition d'une édition ancienne (avec une mise en page différente), qui ne serait peut-être pas éditée pour des raisons de coûts de saisie trop élevés ;
- produire, à partir d'une partition d'orchestre, les partitions pour chaque instrumentiste et inversement ;
- extraire chacune des voix d'une partition à portées polyphoniques, afin de générer une partition avec une seule voix par portée (lorsque par exemple, chaque voix correspond à un chanteur) ;
- adapter une partition pour un autre instrument ;
- traduire une partition en Braille.

Afin d'utiliser la méthode DMOS (section 3.2) pour la reconnaissance de ce type de document structuré, nous avons défini, en utilisant le formalisme de position EPF, une grammaire complète de la notation musicale avec l'aide de Bernard Rétif, professeur, conseiller aux études au Conservatoire de musique de Région, à Caen. Cette grammaire permet de définir un système de reconnaissance de partitions musicales qui produit une représentation pivot (figure 5) servant de point de départ aux applications présentées ci-dessus.

4.1.2 Saisie de partitions musicales

L'ergonomie des logiciels d'édition de partitions musicales est relativement compliquée, au point que de nombreux musiciens abandonnent leur utilisation car la saisie d'une nouvelle partition est beaucoup trop longue. Les musiciens ont l'habitude d'écrire au stylo leurs partitions. Nous proposons donc une nouvelle interface de saisie de partitions musicales qui utilise un stylo numérique. Celle-ci s'appuie sur l'interface générique orientée stylo développée dans le projet Imadoc. Le premier prototype permet une saisie rapide et naturelle de partitions musicales.

4.2 Formules mathématiques

Mots clés : Formules mathématiques, édition, Braille, bibliothèques numériques, saisie orientée stylo, assistants numériques (PDA).

Résumé : *Les applications de la reconnaissance de formules mathématiques sont nombreuses pour l'édition d'ouvrages scientifiques et la constitution de bases de données scientifiques. L'utilisation d'interfaces orientées stylo offrent une possibilité de saisie rapide et intuitive de formules mathématiques.*

4.2.1 Rétro-conversion

La reconnaissance de formules mathématiques permet une phase automatique de rétro-conversion de documents qui n'existent que sous forme papier, ce qui évite une étape longue et fastidieuse de saisie manuelle. Les applications de cette rétro-conversion sont nombreuses, notamment dans le monde de

l'édition. Ainsi de nombreuses éditions anciennes de livres de mathématiques, de physique... , pourraient être rééditées. Dans le cadre des bibliothèques numériques, la constitution de bases de documents contenant des formules mathématiques (comme les articles scientifiques, les dépôts de brevets...) serait facilitée par la rétro-conversion et permettrait d'effectuer différents types d'indexation.

Comme pour l'écriture musicale, il est possible grâce à cette rétro-conversion, de produire des formules mathématiques ou des articles scientifiques directement en Braille.

4.2.2 Saisie de formules mathématiques

La saisie de nouvelles formules mathématiques est une opération relativement lourde. Ainsi les interfaces existantes dans les traitements de texte sont difficiles d'emploi, nécessitent un apprentissage et engendrent un temps de saisie relativement long.

Nous pouvons, en travaillant sur la reconnaissance de formules mathématiques manuscrites, permettre à un utilisateur d'écrire sur une feuille de papier sa formule et la produire automatiquement sous la même forme que celle obtenue par l'intermédiaire d'une saisie au clavier et à la souris.

Une autre possibilité est d'utiliser l'interface générique orientée stylo développée dans le projet Imadoc. L'utilisateur se sert alors d'un stylo numérique pour écrire directement la formule mathématique à l'intérieur d'un traitement de texte. L'intérêt de cette interface est de permettre une saisie parfaitement intuitive, tout en validant l'écriture reconnue grâce à un dialogue entre l'utilisateur et la machine.

Cette interface pourrait être embarquée sur des assistants numériques pour produire ainsi des calculatrices scientifiques complètes.

4.3 Tableaux et formulaires

Mots clés : Tableaux, formulaires, documents anciens, archives, bibliothèques numériques, indexation, gestion électronique de documents (GED).

Résumé : *Les applications de la reconnaissance de structures tabulaires (tableaux et formulaires) sont très larges. Il est par exemple possible de traiter automatiquement des documents anciens ou récents, à l'aide de modèles génériques ou spécialisés afin d'effectuer une indexation automatique ou une reconnaissance de l'organisation structurelle d'un tableau. Ces applications sont très importantes pour permettre un accès électronique à des documents pour l'instant archivés et très difficiles d'accès.*

Après les partitions musicales et les formules mathématiques, nous avons utilisé la méthode DMOS (section 3.2) pour construire des systèmes de reconnaissance de structures tabulaires comme les tableaux ou les formulaires.

Il est ainsi possible de définir des modèles génériques de tableaux complexes afin de reconnaître l'organisation intrinsèque d'un tableau, et ce quelle que soit sa forme. Grâce à la souplesse de la méthode DMOS, il est également possible de définir des modèles plus ou moins spécialisés qui pourront permettre une reconnaissance malgré une mauvaise qualité du document d'origine.

Le système de reconnaissance peut donc travailler aussi bien sur des documents récents de bonne qualité que sur des documents anciens altérés, dont la structure est irrégulière. Les structures tabulaires

peuvent de plus être noyées dans du texte ou placées n'importe où dans le document. Les documents à traiter peuvent aussi bien être dans un flux homogène qu'hétérogène de format de documents. Ces nombreuses capacités offrent un très large spectre d'applications :

- traitement automatique de bons de commande, bons de livraison. . .
- indexation automatique de documents de toutes sortes et provenances :
 - formulaires anciens (section 6.2),
 - cartouches de plans,
 - documents d'archives ou de bibliothèques. . .
- classement automatique en fonction du contenu ;
- entrée automatique de documents dans un système de gestion électronique (GED) ;
- accès électronique possible à des documents qui sans indexation automatique resteraient très difficiles d'accès dans des archives. . .

Un autre intérêt de la méthode DMOS est que pour des documents le nécessitant, l'adaptation du système se fait rapidement, car il n'est nécessaire d'intervenir que sur la définition d'une nouvelle grammaire.

Les grammaires définies à l'aide du formalisme EPF de la méthode DMOS peuvent bien entendu être associées pour, par exemple, définir un modèle de tableaux de formules mathématiques.

4.4 Documents anciens

Mots clés : documents anciens, TEI, XML, comparaison de documents, variantes, numérisation.

Résumé : *La numérisation des « documents anciens » — qu'il s'agisse de manuscrits médiévaux, de brouillons d'écrivains récents mais aussi des plans cadastraux napoléoniens ou des registres de baptêmes — est un sujet très porteur. Mais les techniques employées, en aval de la simple numérisation des dits documents, restent encore très simplistes. Le suivi de normes (comme celles de la TEI ou du projet Master), mais aussi des techniques hypertextes et l'expérimentation, sont fondamentales pour tout ce secteur des sciences humaines.*

Les historiens des textes se sont intéressés très tôt à l'informatique. Bases de données, puis outils d'édition ont vite permis de diffuser et de publier des corpus de textes anciens. Les notions d'hypertexte puis les outils de numérisation permettent dès à présent de disposer de nombreuses bases de manuscrits. Mais cela correspond-il aux besoins des chercheurs en philologie, en histoire des textes, en génétique ? La réponse est non, car les outils proposés sont plus orientés vers la diffusion des informations que vers leur exploitation.

La problématique repose sur plusieurs concepts [?] :

- savoir scanner des manuscrits (pouvant parfois être de plusieurs mètres-carrés, comme les plans cadastraux ou les affiches),
- être capable d'en extraire les informations topologiques (cadres de formulaires, portées musicales, vignette illustrée marquant un début de chapitre dans un parolier du XVI^e siècle, etc.), puis les informations textuelles (numéro de registre, nom d'un conscrit, etc.) ou musicales (notes d'un psautier, etc.),

- structurer ces informations pour y accéder (structures à la TEI, XML, liens hypertextes, etc.) et y faire du *data mining*,
- faire un lien entre une partie d'image (par exemple d'un mot d'un manuscrit) et du texte (le mot correspondant dans la version « diplomatique » ou éditée), la pose de ces liens devant bien sûr être aussi automatique que possible et compatible avec les nouveaux standards comme *Xlink*, *Xpointer*, etc.
- détecter, puis montrer les variantes d'une même œuvre (édition synoptique des Évangiles par exemple), etc.

Le projet Imadoc a depuis longtemps travaillé sur ces problèmes et participe à diverses actions de recherche sur ce sujet : projet Philectre ^[dAR99] (philologie électronique, projet du GIS cognition), conventions avec la société Archimaine, les Archives départementales, le Ministère de la culture.

4.5 Écriture manuscrite

Mots clés : écriture manuscrite, en-ligne, hors-ligne.

Résumé : *Aujourd'hui, les taux de reconnaissance obtenus sur l'écriture manuscrite en-ligne et hors-ligne laissent entrevoir à court terme la possibilité de très nombreuses applications commerciales et industrielles.*

Pour envisager des applications « utilisables » par le plus grand nombre, qui sont donc économiquement intéressantes, il faut parvenir à des taux de reconnaissance pour l'écriture manuscrite très proches de ceux de l'être humain. Face à cet enjeu économique important, les premières études sur la reconnaissance d'écriture ont débuté très rapidement, sans avoir forcément conscience de la complexité des mécanismes intervenant dans la lecture, car elle constitue pour l'homme une activité courante et naturelle. Ces premières études se sont très vite heurtées à cette complexité et ont engendré la conception de systèmes de reconnaissance d'écriture manuscrite aux performances médiocres. Ces premiers systèmes, trop rapidement commercialisés, ont alors eu tendance à discréditer les applications envisagées pour intégrer l'écriture comme une nouvelle modalité d'interaction possible entre l'homme et la machine. Depuis quelques années, les études menées sur la reconnaissance d'écriture connaissent un nouvel essor avec l'apparition de nouvelles méthodologies pour faire face aux problèmes les plus complexes de la reconnaissance des formes. Dans ce sens, nous avons cherché à développer un système de reconnaissance d'écriture manuscrite omni-scripteurs dont le caractère évolutif et les performances déjà obtenues permettent dès à présent d'envisager la conception de ce type d'application.

En *reconnaissance hors-ligne*, les besoins industriels sont aujourd'hui considérables, et par conséquent l'attente de systèmes de reconnaissance robustes et fiables est très grande. L'étendue des applications est vaste ; elle recouvre notamment les problèmes de traitement automatique du montant des chèques, ou des adresses postales, mais aussi, d'une manière générale, le traitement de tout type de document papier, notamment les bordereaux de livraison, les feuilles de sécurité sociale, ou encore les télécopies.

Il existe déjà de nombreuses applications industrielles qui traitent de manière partielle ces différents problèmes. L'apport d'une reconnaissance fiable, même limitée à 30% des cas (rejet direct d'une

[dAR99] G. DE VENTABERT, J. ANDRÉ, H. RICHY *et alii*, « Réflexion sur la numérisation des documents anciens — expériences du projet Philectre », *Document numérique* 3, 1-2, 1999, p. 57–74.

grande partie des documents) peut être d'une très grande utilité dans les domaines d'application dans lesquels les quantités de documents à traiter sont considérables. C'est pourquoi, l'effort des systèmes de reconnaissance hors-ligne porte aujourd'hui sur la garantie d'une grande fiabilité des résultats en cas de traitement automatique effectif pour éviter tout risque d'erreur (même si pour cela il faut rejeter a priori un grand nombre de cas et donc les traiter manuellement). Dans cette optique, nos recherches ont pour objectif de concevoir un système de reconnaissance hors-ligne basé sur les concepts de la modélisation en-ligne de l'écriture que nous avons mis au point (cf section 2.2.3). Ce système devrait permettre, de par son caractère interprétable, de concilier de bons taux de reconnaissance tout en garantissant une grande fiabilité des résultats.

4.6 Interfaces homme-machine orientées stylo

Mots clés : interfaces orientées stylo, communication homme-machine, éditeur orienté stylo.

Résumé : *La conception d'interfaces homme-machine orientées stylo vise à établir de nouvelles modalités d'interaction entre l'homme et la machine qui sont plus conviviales et plus naturelles.*

Dans ce cadre, notre objectif est de développer des interfaces permettant d'intégrer de manière homogène la reconnaissance aussi bien du geste que de l'écriture manuscrite, ou encore d'autres symboles graphiques, afin de pouvoir utiliser une modalité de communication directe avec l'ordinateur par l'intermédiaire d'un papier électronique et d'un stylet.

L'étendue des applications orientées stylo est très vaste : agenda électronique, téléphone portable, CAO, édition électronique, système embarqué, système de contrôle aérien, etc.

Dans ce contexte, nous travaillons dans deux domaines d'applications : le développement d'une interface générique pour la communication homme-machine orientée stylo et l'édition interactive de documents électroniques. Nous avons conçu et réalisé plusieurs prototypes de démonstration (édition interactive de partitions musicales, saisie de graphes dessinés à main levée, ...) qui ont mis en évidence la convivialité de ces nouvelles modalités d'interaction homme-machine.

5 Logiciels

Participants :

Éric Anquetil, Jean Camillerapp, Bertrand Couïasnon, Ivan Leplumey, Hélène Richy.

Résumé : *L'équipe a développé et maintient un ensemble d'outils qui permettent d'évaluer rapidement la faisabilité du traitement de nouveaux types de documents et ainsi de répondre à des demandes industrielles.*

Ces outils se répartissent entre différents niveaux :

- Traitement d'images :
 - binarisation adaptative et extraction du tracé,
 - détection des segments quasi rectilignes qui se croisent ou qui se chevauchent ;

- Reconnaissance :
 - reconnaissance en ligne de lettres et de mots cursifs manuscrits tracés sur une tablette graphique, logiciel RESIF, déposé à l'APP sous le numéro IDDN.FR.001.490009.00.SP.2000.000.00000,
 - reconnaissance de symboles imprimés extraits d'images de document ;
- Reconnaissance de la structure :
 - de partitions musicales,
 - de formules mathématiques,
 - de documents comportant des tableaux ;
- Comparaison de documents :
 - comparaison du contenu de très gros textes,
 - comparaison de partitions musicales à partir d'un corpus en notation abc ;
- Interface stylo :
 - un noyau générique de développement d'interface stylo,
 - une interface sous Amaya² de correction manuscrite de documents HTML.

6 Résultats nouveaux

6.1 Reconnaissance de formules mathématiques

Mots clés : Reconnaissance de documents, connaissance a priori, formules mathématiques, hors-ligne.

Participant : Bertrand Coüasnon.

Résumé : *Afin de valider l'aspect générique de la méthode DMOS (section 3.2), nous avons développé un analyseur hors-ligne d'une partie de la notation mathématique imprimée. Grâce à cette méthode, nous avons pu diminuer le nombre de symboles à faire reconnaître par les classifieurs, en utilisant des règles grammaticales (dans le formalisme EPF) pour en décrire certains. Cette possibilité est particulièrement intéressante car la notation mathématique utilise couramment de 150 à 250 symboles. Le fait d'en reconnaître une partie grâce à des règles de description, permet de réduire le nombre de classes traitées par le classifieur. De plus elle permet de résoudre les symboles de dimension variable comme $\sqrt{b^2 - 4ac}$. Nous essayons également d'aborder certains problèmes de sur et sous-segmentation. En outre, la généricité de la méthode DMOS a permis un développement très rapide de ce système.*

². Amaya est un environnement auteur pour le web diffusé par le consortium W3C depuis l'adresse <http://www.w3.org/Amaya>

6.1.1 Principe

Le système de reconnaissance implantant la méthode DMOS (section 3.2) a été initialement conçu pour la reconnaissance de partitions musicales (section 4.1.1). Après un long développement afin de généraliser cette méthode, nous testons maintenant son utilisation sur différents types de documents. Nous présentons nos premiers résultats sur les formules mathématiques [13].

Nous avons défini, à l'aide du formalisme EPF (section 3.2.2), une grammaire des formules mathématiques en nous fixant certaines contraintes :

- reconnaissance d'une partie de la notation mathématique correspondant aux expressions arithmétiques de base (avec indice et exposant), aux expressions trigonométriques, aux sommes, produits, racine carrée et intégrale ;
- reconnaissance d'une seule formule, bien que la reconnaissance de plusieurs formules dans le même document soit envisageable ;
- reconnaissance de l'écriture imprimée, bien que notre description soit conçue dans l'optique de pouvoir être généralisée à l'écriture manuscrite.

6.1.2 Description grammaticale

La description grammaticale que nous avons définie suit le sens de lecture naturel d'une formule. Elle s'effectue en quatre grandes étapes :

- détection du début de la formule : la description dans le formalisme suit un ordre d'analyse gauche droit. Il nous faut donc pouvoir détecter le début de la formule. On commencera alors l'analyse de l'expression à partir du symbole débutant la formule ;
- gestion des alignements, indices et exposants : étant donné deux symboles, il faut être capable de préciser si le deuxième est en exposant, en indice ou dans l'alignement. Ceci est effectué grâce à une description des positionnements relatifs de deux symboles consécutifs ;
- intégration de la récursivité : la notation mathématique est une notation très récursive. Lors de la recherche d'une expression en indice, en exposant, dans les limites d'une sommation, dans les bornes d'une intégrale. . . , l'analyse se poursuit d'une manière récursive sur une partie de l'image ;
- réduction du nombre de classes de symboles grâce à la description grammaticale de leur forme :
 - ceci est possible pour les symboles principalement composés de segments. Ils ne sont donc plus reconnus par les classifieurs, mais par l'analyse grammaticale. Dans la partie de la notation considérée, on comptabilise 64 symboles, dont 14 sont reconnus par des règles de description. Dans la documentation \LaTeX , on constate que pour certains opérateurs, on pourrait en reconnaître plus des $2/3$ grâce aux segments ;
 - cette utilisation des segments par des règles de description permet également de résoudre, par l'intermédiaire des propriétés de la méthode DMOS, certains problèmes de sur et sous-segmentation liés à ces symboles.

6.1.3 Résultats

Nous n'avons pas encore pu calculer des taux de reconnaissance, car les classifieurs ne sont pas encore intégrés à l'analyseur (les composantes connexes sont étiquetées à la main). Cependant notre

analyseur a été mis au point sur une soixantaine de formules (parfaitement reconnues) numérisées à 300 dpi. Quelques exemples de celles-ci sont présentés dans la figure 6.

$$\frac{4x^2 + 3x + 4}{5x^3 + 4x + 2} \quad \sum_{0 \leq i \leq n} i = \frac{n \times (n + 1)}{2}$$

$$|a + b| \leq |a| + |b| \quad \cos^2(x) + \sin^2(x) = 1$$

$$\prod_{i=0}^n \frac{2}{3} - \frac{\int_0^{\frac{\pi}{2}} y^2 + 3dy}{\frac{2}{3} + 56} - 1 + 2 \times \sum_{i=0}^n i^2 / \frac{1 + \frac{1+4}{3+x+yz}}{2xy + \frac{5+6}{4}} \times e^{\frac{1}{3}}_{a+\frac{1}{3}}$$

FIG. 6 – Exemples de formules reconnues

Il nous reste, après intégration des classifieurs, à valider la chaîne sur un ensemble de test important. Il faudra également augmenter le vocabulaire reconnu, puis adapter l'analyseur à l'écriture manuscrite. Enfin, il sera possible d'étendre la description pour qu'elle puisse reconnaître plusieurs formules dans un document, ou bien, des formules mélangées à du texte.

Ce travail montre, dès maintenant, l'intérêt de la genericité de la méthode DMOS pour pouvoir définir rapidement un système de reconnaissance adapté à un nouveau type de document en modifiant simplement la grammaire.

6.2 Reconnaissances de la structure des registres matricules

Mots clés : Reconnaissance de documents, connaissance a priori, structures tabulaires.

Participants : Bertrand Couïasnon, Jean Camillerapp, Ivan Leplumey.

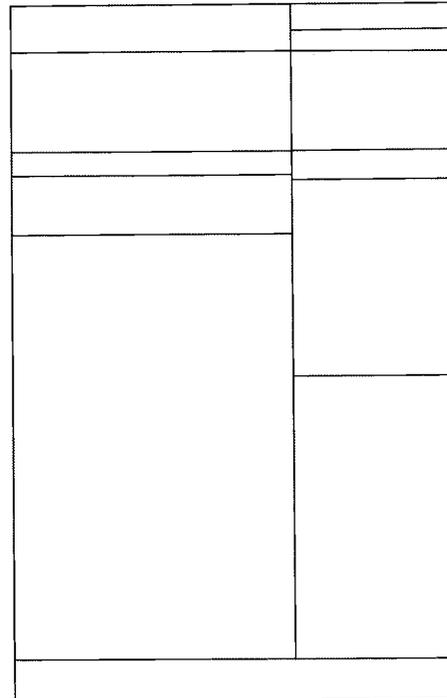
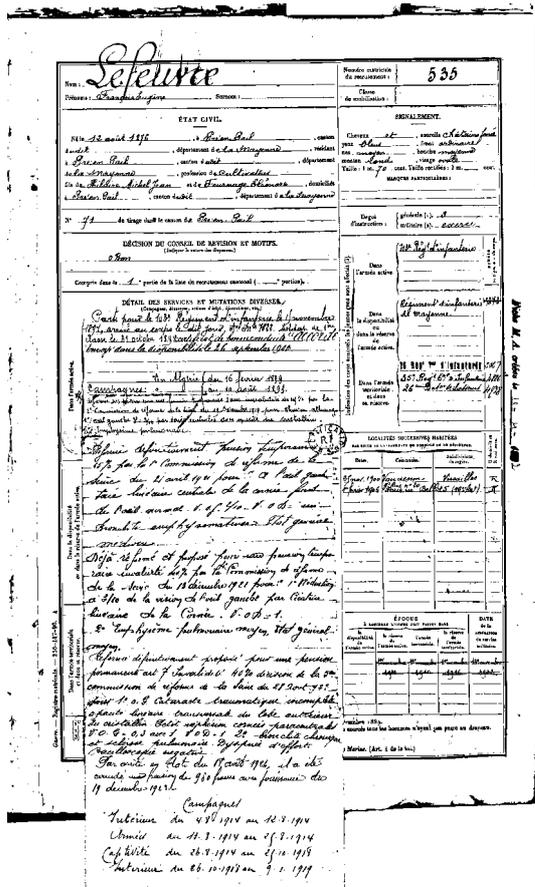
Résumé : *La numérisation massive des archives induit des problèmes d'accès aux documents. La reconnaissance automatique de leur structure permet de mettre à la disposition du public les parties de documents qui, en fonction des délais légaux, ne sont plus protégées, alors que le document dans son ensemble l'est encore. Cette reconnaissance permet également de localiser des régions sur lesquelles il faudra appliquer des traitements particuliers afin de réaliser une indexation automatique.*

Ce traitement, qui s'appuie sur la méthode DMOS, constitue un nouvel exemple de la genericité de celle-ci.

Les services des Archives départementales entament une procédure de numérisation de leurs documents. Cette dématérialisation du document en facilite la conservation et diminue les risques d'endommagement ou de destruction. Si l'interprétation automatique du contenu de certains de ces documents ne paraît pas actuellement possible, une indexation partielle est tout à fait envisageable.

Dans le cadre des conventions que nous avons avec la société Archimaine, les Archives départementales de la Mayenne et le Ministère de la culture, nous avons abordé le traitement de registres matricules militaires du XIX^e siècle.

Ces documents présentent une structure assez stable au fil des années, même si les dimensions des cases ont évoluées. De plus, la nécessité de noter plus de renseignements que n'en prévoyaient initialement les concepteurs des imprimés, (en particulier en raison des services des soldats pendant la guerre de 14-18) a conduit les scripteurs à mettre en place de petits volets de papier qui viennent perturber l'aspect visuel des documents.



(a) Images initiales après binarisation adaptative. Noter la présence d'un volet.

(b) Structures localisées dans l'image.

FIG. 7 – Structuration d'une page d'un registre matricule

De notre point de vue, ces documents constituent un point de convergence et de valorisation de nos recherches antérieures :

- les recherches faites sur l'extraction du tracé pour la reconnaissance de l'écriture manuscrite,
- la localisation des filets séparant les cases par la méthode de Kalman,
- la validation de la généricité de la méthode DMOS (section 3.2).

La localisation précise et automatique des différents cases dans de tels documents présente certains intérêts :

- la durée de non divulgation des informations n'est pas la même selon les cases. La partie *Etat Civil* est protégée pendant 100 ans, tandis que les parties qui contiennent des renseignements médicaux doivent rester confidentielles pendant 150 ans ;
- la transmission d'une page complète, même comprimée, demande des ressources réseaux importantes. La restriction aux régions d'intérêt peut apporter un gain important ;
- le repérage des régions pour lesquelles la reconnaissance de l'écriture manuscrite permettra de réaliser une indexation automatique.

La structure des formulaires a donc été traduite sous la forme d'une grammaire décrivant les positions respectives des différentes cases, mais sans introduire un positionnement précis des filets séparant ces cases. Cependant la présence des volets et la faible qualité de certains documents nous ont conduit à étendre la méthode DMOS pour qu'elle puisse travailler en présence de bruit et qu'elle se satisfasse d'une reconnaissance partielle. De plus, dans une optique de traitement semi-automatique des documents trop difficiles, la méthode a été étendue pour qu'elle explicite les raisons de l'échec de la reconnaissance.

Pour valider notre approche, nous avons à notre disposition 10 registres, ce qui représente plus de 5000 images. Ils étaient répartis sur la période 1878-1900 et constituent un échantillonnage représentatif de ce type de documents. Parmi ces registres trois possèdent beaucoup de volets et de collages. Nous avons mis au point la grammaire et ajusté quelques paramètres sur quelques images issues de 4 registres, puis nous avons validé la méthode sur l'ensemble des images des 10 registres. Pour cette validation nous avons fait des statistiques sur les dimensions des cases dans un même registre et nous avons repéré les images pour lesquelles la hauteur varie de plus d'un millimètre par rapport à la dimension moyenne constatée dans chaque registre. Nous obtenons des taux de reconnaissance de la structure de l'ordre de 98% avec 2% de rejet et aucune erreur [12].

6.3 Didacticiel d'aide à l'apprentissage de la lecture et de l'écriture

Participant : Éric Anquetil.

Mots clés : didacticiel, analyse de l'écriture manuscrite, interface orientée stylo, lecture, écriture.

Résumé : *L'objectif de ce projet est de réaliser un didacticiel d'aide à l'apprentissage de la lecture et de l'écriture manuscrite dans le cadre de l'enseignement pré-élémentaire et élémentaire. Ce didacticiel viendrait compléter les enseignements classiques associés à l'apprentissage de l'écriture en permettant notamment une évaluation interactive précise et personnalisée de chaque élève, ainsi que son suivi pédagogique en classe ou à son domicile.*

Le système de reconnaissance d'écriture manuscrite ^[Anq97] et l'interface orientée stylo [7] développés dans l'équipe nous ont conduit à envisager la conception d'un logiciel permettant aux élèves

[Anq97] E. ANQUETIL, *Modélisation et reconnaissance par la logique floue : application à la lecture automatique en-ligne de l'écriture manuscrite omni-scripteur*, thèse de doctorat, université de Rennes 1, janvier 1997.

du primaire de vérifier et de conforter la "qualité" du tracé des formes fondamentales des lettres qu'ils écrivent. De plus, ce logiciel facilite le contrôle du ductus (l'ordre des tracés et leurs sens).

Les différentes fonctionnalités offertes par le didacticiel peuvent être résumées en quatre points :

- donner la possibilité à l'enseignant de créer lui-même le support des différents exercices,
- permettre une présentation animée de différents tracés,
- permettre à l'élève de procéder à des auto-évaluations,
- permettre d'archiver les tracés et les réponses de chaque élève afin de pouvoir effectuer un suivi pédagogique.

La communication est basée sur une interface orientée stylo intégrant : la gestion des opérations d'édition à l'aide de gestes graphiques simples et intuitifs effectués à main levée, la saisie par écriture manuscrite de commandes et de mots, l'interprétation des tracés de symboles graphiques, une interaction directe avec les menus, boutons, etc. Ces différentes opérations d'interaction (pointage, sélection, déplacement, saisie, etc.) sont centralisées autour d'un même média de communication que tout le monde maîtrise, le stylo. Ainsi, un élève ne sachant ni lire ni écrire peut d'ores et déjà communiquer avec un ordinateur grâce à ce type d'interface.

En dehors du côté ergonomique lié à l'utilisation d'une tablette et d'un stylo, ce périphérique de communication permet en plus de recueillir des informations non visibles sur l'écriture qu'il serait difficile de déduire d'une écriture effectuée simplement sur une feuille de papier. On dispose notamment d'informations relatives à la dynamique du tracé (ordre, sens, vitesse) ainsi qu'à la pression du stylo exercée sur la tablette.

Le périphérique utilisé (stylo + tablette) couplé au système de reconnaissance d'écriture permet d'élaborer une expertise précise et automatique sur de nombreux concepts associés à l'écriture manuscrite. L'ensemble de ces résultats pourront être exploités afin d'élaborer les exercices du didacticiel pour l'aide à l'apprentissage de l'écriture.

Nous disposons aujourd'hui d'une première maquette de démonstration réalisée par Sébastien Olivier dans le cadre d'un stage qui devait se poursuivre par un DRT dans ce domaine. Elle permet d'illustrer un ensemble de possibilités offertes par le didacticiel. Des contacts ont été pris avec l'inspection académique d'Ille et Vilaine pour la conception et la validation pédagogique des exercices.

6.4 Modélisation du contexte par arbre de décision flou pour la reconnaissance de symboles graphiques

Mots clés : arbre de décision flou, système d'inférence floue, classification non supervisée, modélisation du contexte, apprentissage, classification.

Participants : Éric Anquetil, Nicolas Ragot.

Résumé : *Lorsque l'on traite le problème de la reconnaissance des symboles graphiques, il est souvent nécessaire de considérer non seulement les propriétés intrinsèques de la forme, mais aussi la position relative des éléments graphiques à reconnaître ainsi que les connaissances contextuelles associées au cadre de l'application. Dans ce but, nous proposons une approche de modélisation basée sur les arbres de décision flous.*

Lors de la reconnaissance de symboles graphiques, un classifieur peut ne pas parvenir à déterminer la classe associée à un tracé [4]. Ce dernier peut alors être rejeté (rejet de proximité). Le classifieur peut aussi être amené à « confondre » un symbole avec un autre, ce qui peut engendrer un rejet de confusion. Dans les deux cas, l'adjonction d'informations supplémentaires d'ordre contextuel peut permettre de résoudre le problème.

Dans les applications orientées stylo développées dans l'équipe [7] (éditeur de texte, éditeur de graphe et éditeur musical), il est possible d'associer un contexte relatif aux différents tracés, afin de faciliter la reconnaissance. Il peut s'agir de la position relative de deux symboles, de leur taille relative, de leur position dans le document, etc. Il est souhaitable de modéliser ce contexte par un procédé non empirique, automatique et générique. Une approche basée sur les arbres de décision flous a donc été envisagée, car ceux-ci permettent de gérer l'imprécision des données liée à la variabilité inter et intra-scripteur tout en conservant un résultat facilement interprétable et donc exploitable.

Lors de la phase d'apprentissage, un partitionnement en modalités floues de l'espace de représentation des données est d'abord cherché. Ensuite, les attributs permettant de séparer au mieux les classes à reconnaître sont déterminés en fonction de cette partition. Lors de la phase de généralisation, les informations issues de l'apprentissage sont exploitées pour classer de nouveaux symboles graphiques.

Dans l'approche proposée, le partitionnement flou est élaboré à partir d'une classification floue non supervisée de type C-Moyennes floues^[Bez81]. La projection des prototypes obtenus permet de déterminer les modalités qui sont modélisées sous forme de sous-ensembles flous. Pour l'exploitation de l'arbre, chaque chemin de la racine à une feuille est traduit sous forme de règle. L'ensemble de celles-ci forme un système d'inférence floue (cf. section 3.3) permettant de classer de nouveaux tracés.

Des expérimentations ont été menées sur des benchmarks classiques comme la base des Iris de Fischer et la base des formes d'ondes de Breiman et comparées avec un algorithme d'arbre de décision classique (le C4.5^[Qui93]) et un algorithme d'arbre de décision flou (le système Salammbô^[Mar98]). Les résultats sont encourageants puisque les taux de classification obtenus sont comparables, tout en produisant un arbre de taille relativement petite (cf. tableau 1).

Base	Algorithme	Taille de l'arbre	Taux de classification
Iris	C4.5	8.5	95.2%
	Salammbô	4	96%
	algorithme proposé	7.1	96%
Formes d'ondes	C4.5	44.6	72.7%
	Salammbô	66.9	78.2%
	algorithme proposé	27.6	75.67%

TAB. 1 – Résultats.

[Bez81] J. BEZDEK, *Pattern Recognition with fuzzy objective function algorithms*, Plenum Press, 1981.

[Qui93] J. QUINLAN, *C4.5 Programs for machine learning*, Morgan Kaufmann, 1993.

[Mar98] C. MARSALA, *Apprentissage inductif en présence de données imprécises : construction et utilisation d'arbres de décision flous*, thèse de doctorat, université de Paris 6, janvier 1998.

6.5 Comparaison de textes et recherche approchée de thèmes musicaux

Participants : Jacques André, Éric Anquetil, Laurent Pasquer, Hélène Richy.

Mots clés : comparaison de chaînes, génétique, XML, partitions musicales, TEI.

Résumé : *On adapte des algorithmes de comparaison de chaînes conçus pour l'étude du génome humain à l'analyse comparative de textes littéraires (recherche « floue » d'éléments communs à deux textes) et à la recherche de thèmes dans une partition musicale.*

Les algorithmes de comparaison de chaînes sont relativement anciens (Aho, Knuth, ...) [Ste94] et ont servi de base aux produits comme le `diff` d'Unix. Mais, ce sont les travaux sur le génome humain qui les ont remis d'actualité, un problème étant, par exemple, de trouver les plus longues séquences génétiques communes. Ces méthodes, notamment celle de Smith et Waterman, consistent à calculer la distance entre deux chaînes en considérant les opérations de base (omission, insertion, etc. de caractères) nécessaires pour passer de l'une à l'autre. On calcule récursivement une matrice de similarité en fonction d'un coût de substitution des caractères (c'est-à-dire d'une tolérance aux différences).

Ces méthodes peuvent s'appliquer à d'autres chaînes de caractères que les seules chaînes génétiques, par exemples à des textes littéraires voire à des langages codés ou à des partitions musicales.

6.5.1 Comparaison de textes

Pour les textes, on peut détecter des sous-chaînes qui ne sont pas rigoureusement identiques (mots en plus, en moins, différents, etc.), comme c'est le cas dans des versions successives d'une même œuvre. Cependant, ces œuvres peuvent parfois avoir duré des décennies, si ce n'est des siècles, et bien sûr l'orthographe a évolué. Il nous a alors paru indispensable de tenir compte de ces variations locales et de préférer l'utilisation de méthodes comme BLAST [AGM⁺90] (*Basic Local Alignment Search Tool*) que nous avons adaptées au traitement des textes par l'utilisation de dictionnaires et de la notion de point d'ancrage (permettant de détecter le départ de séquences communes à deux textes). Un logiciel, PAT, est en cours d'expérimentation afin de maîtriser le choix des paramètres (distance entre deux mots, etc.). Outre les applications en philologie (critique génétique notamment), ces méthodes peuvent utilement être appliquées à la recherche de plagiat.

Une autre voie, faisant encore l'objet d'expérimentations (recherche de textes caractéristiques, granularité, discrétisation de l'affichage, etc.) est la « visualisation » (au sens donné en *data mining*) de propriétés globales d'un gros document. En comparant un document mot à mot à lui-même, on peut montrer des particularités de vocabulaire. La figure 8 montre à gauche un gros extrait de l'*Évangile selon Saint-Luc* : le carré noir correspond à la « généalogie » où la densité des mots « fils de » (Joseph, fils d'Héli, fils de Matthat, ... , fils de Dieu) est très forte. Au centre, extrait d'un chansonnier du troubadour De Ventadour : les carrés géométriquement répartis correspondent à un changement de rythme marqué par des séquences de « e » (« et » en vieux provençal). À droite, un article scientifique français : le carré noir correspond à un programme d'informatique dont le vocabulaire anglais ne se retrouve pas ailleurs.

[Ste94] G. A. STEPHEN, *String Searching Algorithms*, World Scientific Publishing Co., 1994.

[AGM⁺90] S. ALTSCHUL, W. GISH, W. MILLER, E. MYERS, D. LIPMAN, « Basic Local Alignment Search Tool », *Journal of Molecular Biology* 215, 1990, p. 403-410.

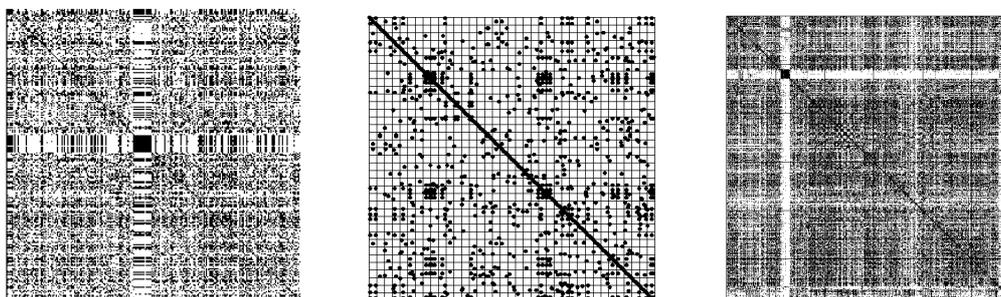


FIG. 8 – Résultats d’auto-comparaison de documents.

6.5.2 Recherche de thèmes musicaux

Les partitions musicales sont codables informatiquement à l’aide de langages textuels (comme abc, MusicML, ...) et peuvent donc être assimilées à du texte. Une base de données de partitions peut alors être consultée par ces algorithmes de recherche pour, par exemple, localiser un thème musical donné ou détecter le plagiat. Pour cela, nous utilisons une base de donnée codée de façon spécifique permettant de dissocier la hauteur des notes du rythme. La saisie du thème à rechercher est faite sur une tablette graphique (voir section 4.1.2). La figure 9 montre (en haut à gauche) un thème saisi et les partitions de la base où on retrouve cette même séquence, mais avec un rythme différent et une hauteur différente. Seul *Au clair de la lune* a un pourcentage d’adéquation (80%) intéressant.

Une chaîne complète de traitement est en cours d’expérimentation.

6.6 Histoire de la typographie

Participant : Jacques André.

Mots clés : histoire des sciences, caractères, approximation de courbes.

Résumé : *La typographie ancienne nécessite une vision historique pour l’analyse de documents anciens.*

Pour un scientifique, faire un minimum d’histoire des sciences est souvent non seulement un droit mais souvent un devoir. Ceci est encore plus vrai lorsque cette recherche historique vient appuyer les recherches d’un projet. Dans le cadre d’Imadoc, nous sommes amenés à étudier des documents anciens. Il est donc tout normal d’étudier aussi la façon dont ces documents sont composés. Au niveau des caractères³, plusieurs champs sont en cours d’étude :

- Lors de la Renaissance, divers modèles mathématiques (basés sur « la règle et le compas ») ont été définis pour décrire les capitales. On montre comment des cercles peuvent avoir approché les ellipses des contours intérieurs de lettres comme le O [6].
- Les manuscrits officiels (actes d’état civil, registres des armées, etc.) utilisent la bâtarde pour écrire les entrées de rubriques (noms de personnes, etc.). Avant de les reconnaître, il convient d’abord d’étudier l’histoire de ces caractères, leurs formes, leur évolution.

3. Et de façon plus générale des signes, comme nous l’avons fait l’an dernier pour les signes de correction typographique.

FIG. 9 – Recherche d'un thème musical dans une base de donnée

- Même aujourd'hui, la notion de caractère a beaucoup évolué depuis l'Ascii à Unicode en passant par Latin-1. Il convient donc d'analyser cette évolution tant historique que normative [3] et de réfléchir aux recherches universitaires pouvant mener au pluri-linguisme [5].

6.7 Lecture automatique de l'écriture manuscrite

Participants : Laurent Pasquer, Guy Lorette, Éric Anquetil.

Mots clés : Modélisation, reconnaissance, interprétation, écriture manuscrite, lecture automatique.

Résumé : *Nous avons expérimenté une nouvelle approche perceptive de lecture automatique d'écriture manuscrite.*

Nous avons mené de nouvelles expérimentations de lecture automatique de l'écriture manuscrite en utilisant le modèle SPI (*Système de Perception et d'Interprétation*) précédemment développé [9] [10] pour construire un système de reconnaissance de mots manuscrits (cf. figure 10).

Ces expérimentations ont donné de bons résultats [1] résumés sur la figure 11, puisque l'apprentissage a été effectué à partir de lettres isolées tandis que la reconnaissance a porté sur des mots cursifs écrits par sept utilisateurs différents dont le système ne connaissait pas a priori le style d'écriture.

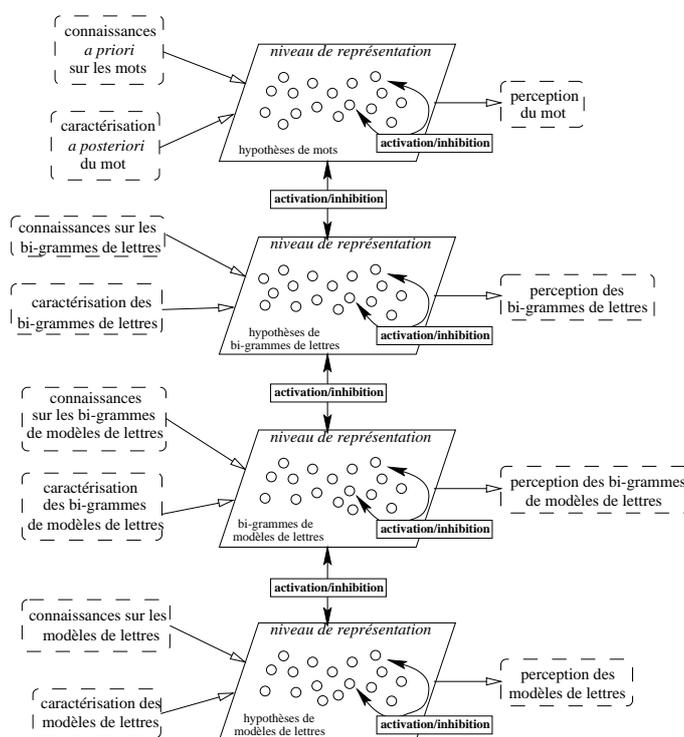


FIG. 10 – Système itératif d’interprétation multi-contextuelle et de lecture automatique de mots manuscrits.

Taille du lexique utilisé	Réponse correcte en première position	parmi les 2 premières positions	parmi les 3 premières positions
L_{1128}	94 %	96 %	97 %
L_{6915}	84 %	88 %	90 %

FIG. 11 – Résultats de l’expérimentation dans un contexte interactif (pas de rejet) du système de lecture de mots manuscrits.

Ces résultats correspondent à l’utilisation du système de reconnaissance dans un contexte *interactif*, c’est-à-dire que le système ne rejette aucun mot mais préfère toujours donner une réponse. L’interaction est ainsi plus forte entre l’utilisateur et le logiciel : plutôt que de refuser de reconnaître un mot, le système répond par une liste ordonnée de mots les plus proches de celui écrit.

En comparant les taux de réussite correspondant à la bonne réponse en première position (le mot lu correctement est en tête de liste) ou parmi les trois premières réponses, on s’aperçoit que l’utilisateur ne doit rechercher le mot qu’il a écrit au delà des trois premiers proposés par le système que dans 3 % (resp. 10 %) des cas lorsque le lexique utilisé comporte 1128 mots (respectivement 6915 mots).

6.8 Interface-Utilisateur-Document-Interactif

Participants : Hélène Richy, Guy Lorette.

Mots clés : Interface utilisateur document, interface stylo, document interactif, papier électronique, correction typographique,.

Résumé : *L'utilisation d'une interface-stylo en tant qu'interface-utilisateur-document-interactif permet d'améliorer nettement l'ergonomie de l'interaction homme-document (IHD). Ce type d'interaction suppose une modélisation du geste graphique et du contexte d'édition dans lequel le geste est effectué. Les recherches de l'équipe Imadoc portent principalement sur cette modélisation du geste graphique et de son contexte ainsi que sur la mise au point d'une interface générique de correction de documents électroniques.*

L'édition électronique permet de fabriquer des documents numériques, de les corriger, de les afficher sur un écran et d'en produire une version papier. Néanmoins, la phase de relecture et de correction reste en grande partie manuelle et fastidieuse. Pour faciliter et automatiser ce travail de correction, nous avons intégré à un système d'édition interactive une interface stylo capable de reconnaître les symboles de correction et de les interpréter.

L'utilisation d'une interface stylo offre de nombreux avantages : elle nécessite peu d'apprentissage de la part de l'utilisateur, elle permet d'indiquer de manière ergonomique et conviviale : le type de correction à effectuer et l'endroit où celle-ci doit être faite. Elle permet également d'entrer les arguments de la commande de correction soit en pointant dans des menus, soit en écrivant de manière naturelle la partie de texte corrigée.

L'an passé, un premier système de correction utilisant une tablette et un stylo électronique avait été développé dans l'environnement Unix. Ce système permet d'éditer interactivement des pages *Web* avec l'éditeur *Amaya* [8].

La reconnaissance de la forme géométrique du symbole de correction ne suffit pas à elle seule pour effectuer la correction d'un document électronique. Encore faut-il qu'ensuite l'ordre de correction soit correctement interprété. Il faut pour cela que le symbole soit positionné correctement et que la portion de texte concernée soit correctement détectée. L'imprécision de certains tracés entraîne parfois des corrections non désirées ou l'impossibilité d'associer à un symbole de correction la portion de texte normalement attendue.

Cette année, les travaux ont porté sur l'amélioration du système précédent. La reconnaissance de la forme géométrique des symboles a été rendue plus fiable et la localisation des symboles plus précise par l'utilisation de méthodes statistiques (droite de régression, histogramme de projections) qui permettent d'obtenir, à partir des tracés réels, des tracés idéalisés. Un petit nombre de points caractéristiques remarquables déduits de ces tracés idéalisés (appelés : *points de localisation*) permet de faire le lien avec le contexte *logique* de la correction (caractère, mot, ligne). L'utilisation de ces méthodes a été testée au travers d'un programme de simulation doté d'une interface graphique.

Au delà de la correction typographique traditionnelle, les travaux en cours permettent d'envisager d'autres applications telles que la correction de tableaux, de graphiques, de formules mathématiques,

de documents multimedia ou d'hypertextes, la correction de mise en page ainsi que le développement d'interfaces-utilisateur-document-interactif entièrement manuscrites.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Convention de recherche avec le Ministère de la Culture

Participants : Bertrand Coüasnon, Jean Camillerapp, Ivan Leplumey.

Les travaux sur les registres matricules qui sont décrits dans la section 6.2 ont été financés par une convention de recherche avec le Ministère de la Culture :

- Numéro de la convention : INSA 0013,
- Intitulé : Rapport d'expertise sur les possibilités de traitements automatiques de documents numérisés,
- Période : avril à décembre 2000.

Ces travaux ont donné lieu à des conventions complémentaires, sans incidence financière, avec les Archives départementales de la Mayenne qui ont fait numériser ces registres et la Société Archimaine de Laval qui a réalisé cette numérisation afin de préciser les conditions d'utilisation et de mises à disposition des données.

8 Actions régionales, nationales et internationales

8.1 Actions régionales

Les travaux sur le traitement des registres matricules résultent de l'intérêt et du soutien manifestés par les archives départementales de la Mayenne, et par la société Archimaine de Laval, spécialiste en technique de numérisation de grands formats (par exemple le cadastre napoléonien) et de microformes (comme les microfilms des Mormons).

Dans le cadre de la reconnaissance de l'écriture manuscrite, l'Ecole Polytechnique de l'Université de Nantes a mis à notre disposition sa base de données IRONOFF. En contrepartie, nous avons mis à la disposition de la société Vision Object qui travaille avec cette école dans ce domaine notre base de caractères en ligne manuscrits. Cette base a également fait l'objet d'un dépôt à la bibliothèque nationale.

Jean Camillerapp fait partie des coordinateurs d'expertise des conventions CIFRE auprès de l'ANVAR.

8.2 Actions internationales

Guy Lorette est responsable des relations internationales de l'IFSIC, membre de la CAI (commission des affaires internationales) de l'Université de Rennes 1 ainsi que le représentant français au *Governing Board* de l'IAPR (*International Association of Pattern Recognition*) et membre du *Constitutions and Bylaws Committee* de cette association.

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

9.1.1 Comité éditorial de journaux

Jacques André est :

- co-rédacteur en chef de la revue *Document numérique* des éditions Hermes, qui aura publié en 2000 :
 - « Gestion des documents et gestion des connaissances » (vol. 3, n° 3-4, 350 pages)
 - « Indexation et documents » (vol. 4, n° 1-2, en préparation).
- rédacteur en chef de la *Lettre GUTenberg* (jusqu'en juin 2000) et des *Cahiers GUTenberg* ; ces derniers ont notamment publié :
 - les actes de la conférence *GUT'2000* sur XML (n° 35-36, 220 pages)
 - un numéro courant (n° 37 – sous presse) ;
- responsable de la rubrique « documents numériques » de l'encyclopédie *Techniques de l'ingénieur* ;
- membre de l'*editorial board* de la *Typography series* de Addison-Wesley.

Guy Lorette est membre du comité de rédaction des revues :

- *Pattern Recognition*,
- *IJDAR—International Journal of Document Analysis and Recognition*.

9.1.2 Comités de programme de conférences

Jacques André a été ou est membre du comité de programme de : GUT2000 (Toulouse), DDEP00 (Munich, 2000), DDEP2002, CIDE'00 (Lyon, 2000), CIDE'01 (Toulouse, 2001), CFDE (Tunis, 2002).

Guy Lorette a été ou est membre des comités de programme de : RFIA'2000 (Paris, 2000), CIFED'2000 (Lyon, 2000), ACIDCA'2000 (Monastir, M2000), IWFHR'7 (Amsterdam, 2000), ICMI'2000 (Pékin, 2000), ASTI'2001 (Paris, 2001), IWFHR'8 (Niagara on the Lake, 2002).

Jean Camillerapp a été membre du comité de programme de RFIA'2000 (Paris, fév. 2000).

9.1.3 GdR

Laurent Pasquer et Guy Lorette participent au groupe de travail « 5.2-Écrit », thème 5 : Communication, du GDR-PRC I³ (Information, Interaction, Intelligence).

Jacques André et Hélène Richy participent au groupe de travail « 7.1-Documents Multi-Médias », thème 7 : Systèmes I³ du GDR I³.

9.1.4 Sociétés savantes

Jacques André est membre du conseil d'administration de l'association GUTenberg.

Jean Camillerapp et Guy Lorette participent aux activités de l'association GRCE : «Groupe de Recherche en Communication Écrite ». Guy Lorette est vice-président GRCE en charge des relations internationales.

Guy Lorette participe aux activités de l'association AFRIF : « Association Française de Reconnaissance et d'Interprétation de Formes ». Guy Lorette est membre du CA et représentant à l'IAPR.

9.1.5 Listes de discussion

Jacques André est responsable de la liste *typographie* hébergée à l'Irisa (sympa@irisa.fr) et archivée sous l'adresse : <https://www.irisa.fr/wws/info/typographie>; et de la liste *GTTS: Groupe de Travail sur la Typographie des Sciences* hébergée à l'Irisa (sympa@irisa.fr) et archivée sous l'adresse : <https://www.irisa.fr/wws/info/gtts>.

9.2 Enseignement universitaire

L'équipe étant en grande partie constituée par des enseignants chercheurs, ceux-ci sont très impliqués dans les activités d'enseignement, mais comme la plupart ne sont pas rattachées à ce thème de recherche, elles ne sont pas citées ici.

DEA d'informatique, IFSIC, participation de Guy Lorette (cours perception reconnaissance et interprétation d'images).

DESS Méthodes informatiques et technologies de l'information et de la communication (DESS Mitic), IFSIC, responsabilité de Hélène Richy, préparation et suivi du dossier de création de cette nouvelle filière spécialisée dans la conception d'interfaces et les nouvelles technologies, définition des programmes, organisation et répartition des enseignements (ouverture prévue en septembre 2001).

Participation d'Éric Anquetil et de Guy Lorette à la 3^e année de la filière "Systèmes d'information statistique" de l'ENSAI (cours : reconnaissance à l'aide de réseaux de neurones, reconnaissance de l'écrit manuscrit).

Stages effectués au sein du projet :

- Éric Flageul, Pascal Garcia et Nicolas Ragot (DEA D'INFORMATIQUE, IFSIC).
- Nicolas Socheleau et Benoît Dien (stage d'été, INSA).
- Sébastien Ollivier (stage d'IUP Miage en vue de poursuivre en DRT).

9.3 Participation à des colloques, séminaires, invitations

Jacques André a organisé la journée Irisatech sur « La galaxie XML » le 7 avril 2000.

Jacques André a donné une conférence invitée : « De Pacioli à Bézier, 5 siècles de mathématiques pour la typographie » au Congrès IREM, Rennes, juin 2000 ; exposé repris dans le cadre de « L'année des mathématiques », Université de Rennes 1, juin 2000.

Jacques André a donné une conférence invitée : « Recherche universitaire pouvant aider au pluri-linguisme sur Internet », colloque Lexipraxis, Paris, décembre 2000.

10 Bibliographie

Thèses et habilitations à diriger des recherches

- [1] L. PASQUER, *Conception d'un modèle d'interprétation multi-contextuelle, application à la reconnaissance en-ligne d'écriture manuscrite*, thèse de doctorat, université de Rennes 1, janvier 2000.

Articles et chapitres de livre

- [2] J. ANDRÉ, P. LAURENT, « Publications scientifiques électroniques: quoi et comment? », *Actes GUT2000: Cahiers GUTenberg*, 35-36, mai 2000, p. 5–13.
- [3] J. ANDRÉ, « Normes et codage des caractères », in: *Traité informatique*, Techniques de l'Ingénieur, décembre 2000, ch. H 7028, p. 1–24.
- [4] ÉRIC ANQUETIL, B. COÛASNON, F. DAMBREVILLE, « A Symbol Classifier able to Reject Wrong Shapes for Document Recognition Systems », in: *Graphics Recognition, Recent Advances*, A. K. Chhabra et D. Dori (éditeurs), *Lecture Notes in Computer Science, 1941*, Springer, 2000, p. 209–218.

Communications à des congrès, colloques, etc.

- [5] J. ANDRÉ, « De Babel Web à Orléans: plurilinguisme, Internet et recherche universitaire », in: *Actes du congrès LexiPraxis 2000*, p. 27–28, Paris, décembre 2000.
- [6] J. ANDRÉ, « De Pacioli à Bézier: 5 siècles de mathématiques pour la typographie », in: *Actes du congrès IREM*, J.-P. Escofier (éditeur), Rennes, juin 2000. à paraître.
- [7] E. ANQUETIL, « Interfaces orientées stylo et reconnaissance dynamique de tracés dessinés à main levée », in: *Neurosciences et sciences de l'ingénieur (NSI)*, p. 173–176, Dinard, septembre 2000.
- [8] J. AZÉ, H. RICHY, G. LORETTE, « Interface-stylo de correction en ligne de documents électroniques: application aux pages web », in: *Actes du colloque CIFED' 2000*, H. Emptoz, N. Vincent (éditeurs), Presses polytechniques et universitaires romandes, p. 261–269, Lyon, juillet 2000.
- [9] L. PASQUER, E. ANQUETIL, G. LORETTE, « Système de perception et d'interprétation (SPI) multi-contextuelle pour la reconnaissance en-ligne d'écriture manuscrite », in: *Neurosciences et sciences de l'ingénieur (NSI)*, p. 169–172, Dinard, septembre 2000.
- [10] L. PASQUER, E. ANQUETIL, G. LORETTE, « Système itératif d'interprétation multicontextuelle pour la lecture d'écriture manuscrite », in: *RFIA'2000*, p. 347–356, Paris, février 2000.
- [11] H. RICHY, « Métadonnées pour les documents sur l'Internet », in: *Document Électronique Dynamique (Actes du Colloque CIDE'2000)*, M. Gaio, E. Turpin (éditeurs), Europa production, p. 141–151, Lyon, juillet 2000.

Rapports de recherche et publications internes

- [12] B. COÛASNON, J. CAMILLERAPP, I. LEPLUMEY, « Traitement automatique de documents d'archive numérisés », *rapport de convention*, Ministère de la Culture, décembre 2000.

Divers

- [13] P. GARCIA, B. COÛASNON, « Reconnaissance de formules mathématiques », *JED'2000*, Lyon, juillet 2000.