



HAL
open science

Traitement des documents anciens à l'aide d'ontologie

Mickaël Coustaty, Norbert Tsope, Karell Bertet, Alain Bouju, Georges Louis

► To cite this version:

Mickaël Coustaty, Norbert Tsope, Karell Bertet, Alain Bouju, Georges Louis. Traitement des documents anciens à l'aide d'ontologie. Les Cahiers du numérique, 2013, 8, pp.91-118. <10.3166/LCN.8.3.91-118>. <hal-00915996>

HAL Id: hal-00915996

<https://hal.science/hal-00915996v1>

Submitted on 12 Dec 2013

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

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



HAL Authorization

TRAITEMENT DES DOCUMENTS ANCIENS À L'AIDE D'ONTOLOGIE

MICKAËL COUSTATY

NORBERT TSOPZE

KARELL BERTET

ALAIN BOUJU

GEORGES LOUIS

Cet article s'intéresse au traitement des documents anciens, et plus particulièrement à des images de lettrines (lettres décorées en début de paragraphe de livres du 16^e siècle). Cette approche s'appuie sur une démarche ontologique pour établir les liens entre les régions extraites par les algorithmes de traitement d'images d'une part et les éléments sémantiques d'autre part, afin d'aider les historiens dans leur interprétation et à les situer dans le temps. Nous avons combiné trois ontologies (thésaurus défini par les historiens, ontologie des traitements informatiques et ontologie spatiale) pour l'annotation de ces images ; ainsi nous avons donc défini des règles permettant d'annoter certaines régions comme étant la lettre, une partie du corps de personnage ou encore de caractériser une lettrine comme figurative.

1. Introduction

Depuis quinze ans, grâce aux possibilités offertes par le développement du numérique, des institutions culturelles, des établissements et organismes

d'enseignement et de recherche, des associations, ou quelques entreprises privées, numérisent le patrimoine écrit et graphique. Ces documents peuvent apparaître sous une myriade de forme (manuscrits, livres et imprimés de toute nature, cartes, estampes, photographies, partitions...).

Une fois les étapes de dématérialisation et de publication sur le web franchies, il convient de valoriser, d'exploiter et de communiquer aux usagers ces documents ou ensembles de documents auxquels est désormais attachée une décision de conservation définitive. Conserver impose plusieurs problèmes majeurs :

1) Stocker de manière pérenne l'information pour pouvoir la consulter même plusieurs dizaines d'années après.

2) Enrichir la base d'information de ces documents en venant extraire son contenu pour permettre une recherche non pas sur quelques mots-clés mais sur l'ensemble de l'information renfermée.

3) Permettre de visualiser et de naviguer dans ces corpus le plus facilement possible, tout en étant proche des attentes des utilisateurs. Il devient nécessaire de rendre cette information accessible à tous depuis n'importe où.

Le stockage pérenne des informations est un domaine très actif qui nécessite de se renouveler en permanence. Dans le cadre de ces travaux, nous nous intéressons plus particulièrement aux deux autres points. Ainsi, il convient de traiter cette grande masse d'information, de la structurer, pour proposer des outils de navigation et de recherche adaptés aux différents types de contenu, et capables d'interpréter les besoins des utilisateurs finaux (chercheurs en SHS, utilisateur lambda...). Structurer la masse d'information implique des outils informatiques capables d'extraire de l'information, de la structurer pour naviguer rapidement en son sein. Interpréter les besoins des utilisateurs se situe à la frontière entre informatique et SHS puisqu'il convient de proposer des outils capables d'aborder le problème de fossé sémantique, où différents mots peuvent être associés à des concepts similaires.

Dans le cas des images, cette tâche devient difficile puisqu'il faut pouvoir mettre en concurrence des données textuelles avec le contenu des images (les pixels). Afin de pallier ce problème, un domaine très actif en recherche (généralement connu sous le nom de CBIR, – *Content Based Image retrieval* – (Smeulders *et al.*, 2000 ; Jeon *et al.*, 2003 ; Liu *et al.*, 2007) vise à décrire et résumer le contenu des images en calculant des descripteurs sur elle-même. Ces descripteurs sont composés d'informations bas niveau extraites de l'image (couleur moyenne, nombre de formes contenues, taille des formes, etc). Rechercher une image consiste alors à retrouver celles contenant des informations bas niveau similaires. Les descripteurs présentent l'avantage de

faciliter la comparaison entre les images, mais ne permettent pas de comparaison directe avec une requête d'un utilisateur, c'est le problème du fossé sémantique. Dans cet article, nous nous intéressons au problème de fossé sémantique pour les images, et à l'utilisation des ontologies qui semblent être une solution prometteuse pour définir une sémantique explicite entre concepts visuels et langage naturel. Ainsi, deux ontologies sont définies : une associée aux experts de la Renaissance, une associée aux traitements des images de documents anciens et qui réutilise une ontologie spatiale. Ces ontologies sont peuplées à partir des connaissances des historiens et d'algorithmes appliqués sur les images, et sont enrichies à partir de règles d'inférences. Ces règles sont alors utilisées pour réduire le fossé sémantique, et proposer une mise en correspondance entre zones de l'image et mots-clés des historiens.

Il paraît donc nécessaire d'associer des mots du langage naturel aux images pour faciliter la recherche de l'utilisateur. Ainsi, cet article commence par une présentation détaillée du problème du fossé sémantique. Puis, dans un second temps, il présente un cas d'étude illustratif : les images de lettrines issues de documents anciens, la proposition des outils de navigation et de recherche d'images de la Renaissance. La troisième partie est consacrée à l'étude du cas, et enfin dans une quatrième partie, les caractéristiques propres à ce cas d'étude sont identifiées dans le but de positionner notre approche dans un cadre générique plus large d'aide à la réduction du fossé sémantique.

2. Fossé sémantique, ontologies et images

La numérisation massive des documents permet de les conserver. Les pages ainsi scannées peuvent être annotées facilement, mais cette annotation restera limitée à la page, pas au contenu. La gestion efficace de ces documents numérisés implique le développement de techniques d'indexation et d'annotation sémantique. D'une part, les algorithmes de traitement d'images permettent d'extraire la caractéristique bas niveau et, d'autre part, la perception humaine permet de donner une sémantique à des régions de l'image (haut niveau).

Dans la recherche des modèles capables de faire les corrélations entre les caractéristiques bas niveau et la sémantique de haut niveau, certaines approches se sont basées sur les « connaissances contextualisées » : on peut ainsi améliorer l'interprétation des images en construisant des hiérarchies sémantiques ou en intégrant des connaissances à priori. Ces connaissances permettent alors de situer l'image dans un contexte. Les connaissances contextuelles permettent d'obtenir une description plus fidèle, aident à raisonner et à améliorer l'annotation (Neumann *et al.*, 2008 ; Bannour *et al.*, 2012).

Il est important d'utiliser des méthodes explicites et formelles pour représenter les connaissances. Cela permet de prendre en compte la connaissance associée au contexte général, au contexte spécifique, et d'améliorer l'interprétation de l'image tout en s'affranchissant de l'implémentation utilisée. La connaissance issue du contexte général est considérée comme connaissance du domaine et la connaissance issue du contexte spécifique est généralement vue comme connaissance issue des images.

L'approche de Newell (1982) dans le cadre de l'ingénierie de connaissances, suggère de séparer le niveau « connaissance » du niveau « symbole » lors de la conception. Dans cette façon de modéliser, le contenu de la connaissance est traité au niveau connaissance alors que la structuration des données est faite au niveau symbole. Dans la modélisation, on distingue les connaissances descriptives, les connaissances opératoires, les connaissances stratégiques et les connaissances systèmes.

Les connaissances de domaine peuvent être représentées par des ontologies. Les ontologies, introduites dans le cadre du web sémantique, fournissent un cadre formel et explicite d'une conceptualisation partagée (Gruber, 1995). D'après cette définition, les ontologies capturent la connaissance pertinente d'un domaine, fournissent une compréhension commune des connaissances de ce domaine, déterminent le vocabulaire reconnu du domaine, donnent la définition explicite de ce vocabulaire et des relations entre ces termes, et tout cela sous forme de modèles formels.

Pour le cas particulier des images de documents anciens, l'annotation de leur contenu impose d'abord de structurer et modéliser la connaissance du domaine des historiens, et celles issues du domaine du traitement d'images, pour ensuite faire le pont entre ces deux blocs de connaissances.

Diverses approches statistiques basées sur l'apprentissage automatique sont proposées dans la littérature pour annoter de manière automatique les régions d'une image : SVM, arbres de décision, réseaux de neurones artificiels, réseaux bayésiens. Ces approches (Zhang *et al.*, 2012) apprennent des fonctions de correspondances entre les caractéristiques et les classes des régions. Bien que décrivant assez bien le contenu visuel des images, les méthodes statistiques ne modélisent pas correctement le sens de l'image tel que perçu par les humains parce que la sémantique y est limitée aux résultats de l'apprentissage de la fonction liant les caractéristiques de bas niveau à des concepts de haut niveau. Ces performances dépendent aussi du nombre de classes apprises.

Outre les méthodes statistiques, certains travaux (Kompatsiaris, 2008) se proposent d'utiliser les concepts (du domaine) pour annoter les images : annotation libre où aucun vocabulaire n'est prédéfini, annotation par des mots-

clefs où un ensemble de mots (ou de concepts) est proposé à l'utilisateur et annotation par une ontologie où un ensemble de mots ainsi que des relations entre ces mots sont fournis à l'utilisateur. L'utilisation d'une ontologie vise différents objectifs : description unifiée des caractéristiques de l'image, description visuelle des relations entre les caractéristiques (lignes, région...), utilisation des informations contextuelles et enfin rapprochement du niveau visuel et du niveau sémantique.

Notons aussi que dans le but de formalisation des connaissances, Neumann (Neumann *et al.*, 2008) propose de modéliser les scènes à l'aide de logiques de descriptions. L'apport principal des logiques de descriptions est d'éviter des erreurs de modélisation quand les connaissances et l'inférence sont construites de manière intuitive.

Un exemple d'ontologie de traitement d'image a été proposé par Clouard *et al.* (2010). Elle comporte 279 concepts, 42 rôles et 192 restrictions. Les concepts y sont classés par niveau : physique, perception, sémantique, tâche et contrainte. Mais les connaissances du domaine n'y sont pas représentées, et le contexte de l'image et les connaissances de l'utilisateur ne sont pas prises en compte.

Nos travaux se positionnent dans cet objectif de réduction du fossé sémantique. Nous avons pour cela choisi un contexte où la sémantique des images est clairement établie, celui des images de lettrines. Nous combinons dans ce cadre une ontologie de domaine représentant la connaissance des historiens, une ontologie des relations spatiales et une ontologie de traitement d'images pour structurer les caractéristiques extraites des lettrines.

3. Cas d'étude : ontologie des lettrines

Nous avons défini une représentation ontologique des images de lettrines. Cette ontologie offre une représentation standardisée de données hétérogènes complexes, qui intègrent à la fois des caractéristiques numériques extraites de l'image ou de certaines de ses régions d'intérêt, une information spatiale entre les régions, et la sémantique du domaine des historiens. Cette ontologie nous a également permis de réduire le fossé sémantique entre description pixellaire et description sémantique d'une image grâce à un mécanisme d'annotation automatique de certaines images ou de certaines régions de l'image à l'aide de propriétés déduites de l'ontologie. Un extrait de cette ontologie des lettrines est donné par la figure 1. Cette ontologie est accessible sur <http://www.navidomass.univ-lr.fr>.

Dans la suite, nous utilisons la légende présentée dans la figure 2 pour décrire les différentes ontologies. Ainsi, les classes sont représentées dans les figures par les symboles de la figure 2a (par exemple la classe « Image ») et les relations entre elles par les flèches de la figure 2b : trait continu par la relation « is a » et trait discontinu pour les autres relations (par exemple, les relations « hasLetter »).

Nous avons défini cette ontologie en plusieurs parties, illustrées par la fleur ontologique (figure 3), chaque étape est décrite plus en détail dans les sections suivantes.

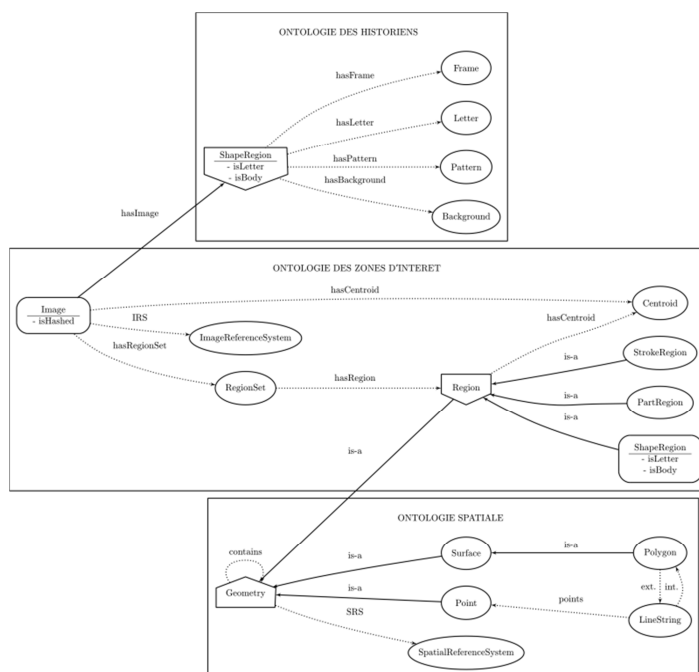


Figure 1. Extrait de l'ontologie des lettrines

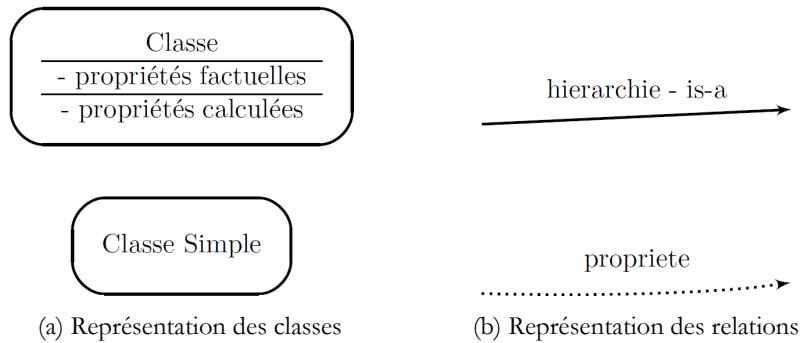


Figure 2. Description du formalisme utilisé pour décrire les ontologies

Ontologie des historiens (*historian knowledge*) (section 3.1) : la connaissance des historiens est décrite par un thesaurus (Jimenes, 2008) que nous avons transcrit en une ontologie possédant la classe Lettrine comme classe principale. Un fonds de 4 288 images de lettrines, annotées à la main par les historiens, est utilisé pour peupler cette ontologie.

Ontologie des régions d'intérêt (section 3.2) : cette ontologie, qui possède la classe Image comme classe principale, est composée de régions d'intérêt extraites à la fois de la couche forme et de la couche texture d'une image de lettrine, ainsi que de leurs caractéristiques bas niveau. Cette ontologie est peuplée par 909 images de lettrines décrites par 5 588 régions de forme et par 451 711 régions de traits. Nous avons enrichi cette ontologie des régions avec un mécanisme de partitionnement de l'image et par des relations spatiales, dans le but de pouvoir localiser chacune des régions selon qu'elle se situe au centre ou sur un bord.

Ontologie des lettrines (section 3.3) : l'ontologie des historiens fournit une représentation haut niveau des lettrines, alors que l'ontologie des régions d'intérêt en est une représentation bas niveau. Nous les avons combinées en une seule ontologie des lettrines en reliant les classes principales des deux ontologies (classes Lettrine et Image), permettant ainsi de représenter et d'interroger de façon homogène l'ensemble des données décrivant les lettrines. Nous avons ensuite cherché à réduire le fossé sémantique entre le niveau sémantique (haut niveau) et le bas niveau par un enrichissement de l'ontologie reposant sur le mécanisme d'inférence.

Inférence sur les régions d'intérêt de la couche forme (section 3.4) : au niveau de la couche forme, nous avons ajouté des descriptions logiques de propriétés permettant d'identifier et d'annoter certaines régions forme comme

la lettre de la lettrine (propriété *isLetter*), ou encore une partie d'un personnage du fond de la lettrine (propriété *isBody*).

Inférence sur les régions d'intérêt de la couche texture (section 3.5) : en ce qui concerne la couche texture, c'est l'ensemble des régions constitutives d'une image qui fournit une information pertinente. Nous avons utilisé les caractéristiques de l'ensemble des régions extraites pour déterminer si l'image est une lettrine à fond hachuré ou non (propriété *isHashed*).

Nous avons utilisé PROTEGE 3.4.4 (Knublauch *et al.*, 2004), éditeur d'ontologies, pour construire cette ontologie, et l'API JENA (Jena, 2011), pour la peupler. Nous avons également utilisé l'outil DLV1, base de données déductive, qui permet de définir les règles d'inférence et de les tester en ligne de commandes. Basé sur la programmation logique déductive, DLV est simple à utiliser, et offre une facilité d'expression des requêtes, avec des temps de réponse relativement faibles. Nous utilisons également SEWELIS – *SEmantic WEb Logical Information System* (Ferré, 2009) pour une recherche d'information dans l'ontologie construite à la fois par requête et par navigation, et qui laisse envisager une meilleure interaction avec l'utilisateur

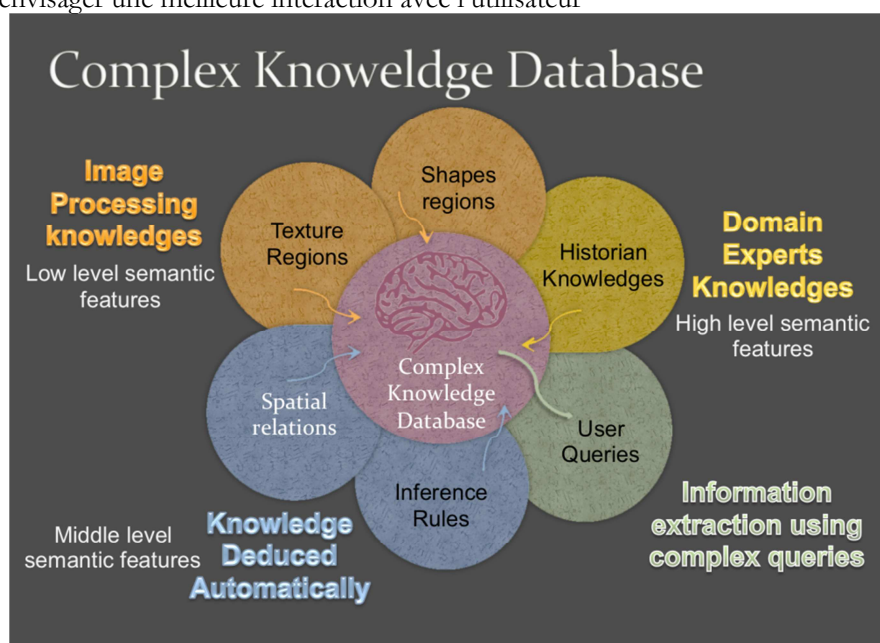


Figure 3. Fleur ontologique des lettrines

3.1. Ontologies des historiens

Les images de lettrines sont des images graphiques que l'on retrouve dans des documents anciens des XV^e et XVI^e siècles. La figure 4 en donne un exemple. Il s'agit d'une lettre décorée, au début des paragraphes des livres de cette époque, obtenue à l'aide de tampon en bois, sculpté à la main. Composée principalement d'une lettre en son centre, une lettrine se caractérise également par un arrière-plan qui peut être décoratif, ou bien représenter des scènes sociales de l'époque (scènes figuratives). Les nuances et les ombres étaient obtenues à l'aide de traits parallèles.

L'étude des lettrines fait l'objet de travaux au centre d'étude supérieur de la renaissance (CESR) de Tours. Les lettrines représentent une source d'information exploitée par les historiens pour situer les documents dans le temps, ou encore pour étudier les scènes sociales des fonds figuratifs. En effet, les tampons étant utilisés plusieurs fois, leur usure permet de situer les documents les uns par rapport aux autres. De plus, ils portent souvent des caractéristiques propres à celui qui les a sculptés.

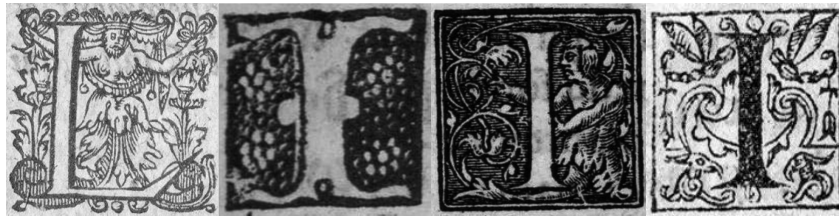


Figure 4. Exemples de lettrines

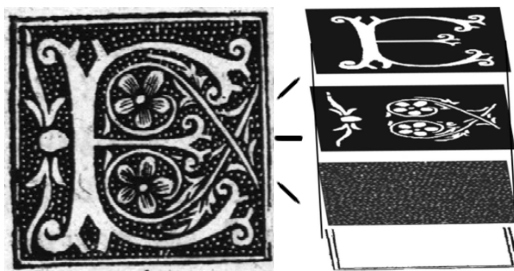


Figure 5. Décomposition en couches des historiens

Figure 6. Extrait de l'ontologie des historiens (partie T-Box)

Nous avons retranscrit ce thesaurus en une ontologie des historiens. La figure 6 décrit cette ontologie. Elle est constituée d'une classe principale *Lettrine* reliée aux classes *Letter*, *Background*, *Pattern* et *Frame* représentant chaque couche sémantique par des propriétés (*hasLetter*, *hasBackground*, *hasPattern*, *hasFrame*). La classe *Letter* est quant à elle reliée aux classes *ColorLetter*, *TypeFont* et *Alphabet* par des propriétés. La propriété *IdentificationLetter* permet de préciser de quelle lettre il s'agit.

Nous avons ensuite peuplé cette ontologie par 4 288 images de lettrines annotées à la main par les historiens.

3.2. Ontologies des régions d'intérêt

Les images de lettrines sont des images particulièrement difficiles à traiter. En effet, elles se sont détériorées avec le temps – jaunissement du papier, pages abîmées, tachées – et les techniques développées doivent être robustes à ces détériorations. Par ailleurs, il s'agit d'images binaires à base de traits sur lesquelles les approches texture classiques d'analyse ne peuvent pas s'appliquer. C'est pourquoi une approche spécifique d'extraction d'information à partir d'images à base de traits a été proposée dans Coustaty (2011).

Tout d'abord, l'image est décomposée en plusieurs couches selon une décomposition décrite dans Dubois *et al.*, (2008), résultat d'une série de projections (Hamidi *et al.*, 2010). Cette décomposition est particulièrement pertinente pour l'analyse de lettrines. En effet, on peut observer que la géométrie d'une image se retrouve principalement sur sa couche forme, alors que la couche texture permet d'isoler les traits qui la composent. Nous avons extrait des régions d'intérêt à partir de chacune de ces deux couches. Un exemple de décomposition est donné dans la figure 7.



(a) Image d'origine (b) Couche forme (c) Couche texture

Figure 7. Décomposition en couches d'une lettrine : résultats obtenus par le traitement proposé par Coustaty (2011)

3.2.1. Régions d'intérêt de la couche forme

La lettre, ou encore des éléments de scènes naturelles des lettrines à fond figuratif, correspondent à des formes de l'image. C'est pourquoi nous avons mis en place une méthode d'extraction adaptée à la couche forme, décrite dans (Coustaty *et al.*, 2011) qui permet d'en extraire certaines régions forme constitutives de l'image. Ces régions d'intérêt sont obtenues en trois étapes :

1) Tout d'abord, l'utilisation d'une loi de Zipf (Pareti *et al.*, 2006) a été utilisée pour sa robustesse aux variations de niveaux de gris, et pour son indépendance à la couleur des composantes. Elle permet de segmenter la couche forme en composantes connexes.

2) Ensuite, les composantes connexes dont l'aire est supérieure à 1 % de l'image sont retenues, les autres étant considérées moins pertinentes. Les composantes connexes ainsi retenues forment les régions d'intérêt de la couche forme.

3) Enfin, nous associons à ces régions d'intérêt des caractéristiques décrivant leur forme :

– L'excentricité Ecc de la région (*Excentricity*), donne une indication de forme. Elle est définie comme le rapport entre le rayon minimal r_m et le rayon maximal r_M de l'ellipse minimale englobant la région (Pratt, 2007) :

$$Ecc = \frac{r_M - r_m}{r_M + r_m}$$

– La moyenne en niveaux de gris (*GreyMean*), et leur écart type (*GreySTD*) permettent d'estimer la couleur de la région, ainsi que sa régularité.

– Le nombre d'Euler En de la région (*Euler*) permet d'en estimer la compacité. Introduit dans Pratt (2007), il se calcule en fonction du nombre de trous H que contient une forme et se réduit à $E_n = 1 - H$ pour une seule région.

Ces régions d'intérêt ne sont pas obtenues par un traitement classique de segmentation d'image en régions, mais par une méthode adaptée décrite dans (Coustaty, 2011).

3.2.2. Régions d'intérêt de la couche texture

La couche texture se compose essentiellement des traits de l'image. Dans une image de lettrines, un ensemble des traits aux caractéristiques visuelles

similaires (longueur, orientation, épaisseur, courbure) peuvent correspondre à un fond hachuré, ou encore à des zones d'ombres. Nous avons mis en place un traitement de la couche texture pour en extraire des zones constituées de traits similaires (Coustaty, 2011). Ces régions de traits sont obtenues en 4 étapes :

1) Nous appliquons tout d'abord une binarisation suivi d'un débruitage pour ne conserver que les traits dans l'image binarisée.

2) Une squelettisation nous permet ensuite d'extraire les traits qui peuvent être d'épaisseurs variables.

3) Nous avons ensuite défini des caractéristiques de traits prégnantes pour la vision humaine, à savoir longueur, épaisseur, orientation, homogénéité et courbure. Une classification non supervisée à partir de ces caractéristiques nous permet de créer des classes de traits similaires.

4) Enfin, des régions de traits sont obtenues en regroupant les traits aux caractéristiques similaires et qui sont voisins dans l'image. Des caractéristiques sont ensuite recalculées pour l'ensemble des traits de chaque région : le nombre de traits d'une région (*StrokesNumber*) ; la longueur des traits (*StrokesLength*) ; leur épaisseur (*StrokesWidth*) ; leur orientation (*StrokesOrientation*) ; leur homogénéité (*Strokes-Homogeneity*) ; et leur courbure (*StrokesCurvature*).

3.2.3. Ontologie enrichie des régions d'intérêt

L'ontologie des régions d'intérêt est constituée d'une classe principale *Image* (cf. figure 11). Des informations globales relatives à l'image sont introduites par les propriétés *ImageReferenceSystem*, *hasCentroid*, *hasLength* et *hasWidth*. Les couches forme et texture sont des images dérivées de l'image initiale, stockée dans la sous-classe *DerivedImage*, alors que l'image initiale est stockée dans la sous-classe *SourceImage*.

Les classes *ROI* (*Region Of Interest*) et *ROISet* permettent d'associer un ensemble de régions à une image. L'aire d'une région, ainsi que les coordonnées de son centre de gravité, permettant ainsi de la situer dans l'image, sont introduites par les propriétés *hasCentroid* et *hasAire*.

Pour distinguer les régions formes des régions traits, nous avons créé deux sous-classes de la classe *ROI*, à savoir les classes *ShapeROI* et *StrokesROI*. Chacune de ces classes est décrite par les caractéristiques des régions d'intérêt à l'aide de leurs propriétés.

Nous avons ensuite enrichi l'ontologie des régions d'intérêt par un mécanisme de partitionnement d'une image de lettrine : chaque image est partitionnée en neuf zones (voir figure 8), une zone correspond au centre de l'image, 4 zones correspondent aux différents bords, et les 4 dernières aux coins

de l'image. Les partitions sont des régions définies par la sous-classe *PartROI* de la classe *ROI*.

1	2	3
4	5	6
7	8	9

Figure 8. Partitionnement d'une image utilisée pour localiser les régions

Les relations spatiales sont alors utilisées pour positionner une région dans une zone. C'est sur les préconisations de la norme ISO sur la manipulation des informations spatiales (13249-3 :2002, 2002) que nous avons choisi d'utiliser l'algèbre RCC8 (Cohn *et al.*, 1997), adaptation des relations d'Allen (Allen, 1983) aux données à deux dimensions, qui propose huit types de relations spatiales entre deux objets spatiaux (*cf.* figure 9). Nous utilisons également le format SFS (*Simple Feature Specification*) pour représenter des objets spatiaux, standard dérivé des spécifications spatiales de SQL préconisées par le consortium OGC (*OpenGIS Consortium*) (Inc., 1999) (*cf.* figure 10).

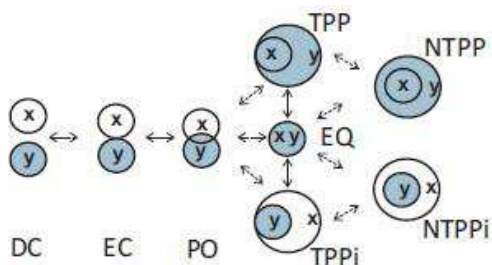


Figure 9. Relations spatiales de l'algèbre RCC-8
(Image extraite du site wikipedia.fr)

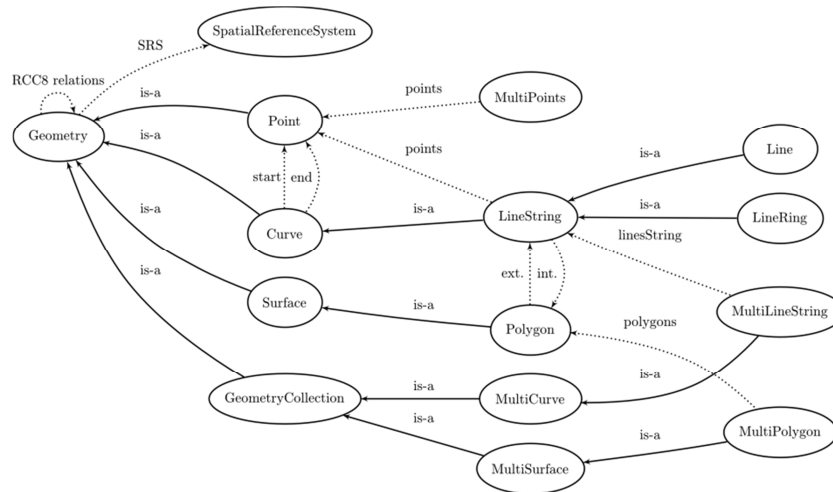


Figure 10. Ontologie spatiale du standard SFS de l'OGC

Ainsi, la classe *ROI* et ses sous-classes héritent de la classe abstraite *Geometry* du standard SFS. Les relations spatiales RCC8 sont des propriétés calculées permettant de comparer entre elles des régions d'intérêt. En particulier, elles permettent de situer des régions *ROIShape* et *ROIStrokes* dans une des neuf partitions *ROIPart*. Notons ici que les relations spatiales entre une région d'intérêt et chacune des neuf partitions peuvent également se calculer lors de la phase d'extraction des régions, mais au détriment de l'aspect normatif. Nous avons sélectionné aléatoirement un jeu de test de 909 images de lettrines à partir desquelles 5 588 régions ont été extraites de la couche forme, et 451 711 régions ont été extraites de la couche texture. L'ontologie des régions d'intérêt est peuplée par ces images et ces régions.

3.3. Ontologie des lettrines

Les trois ontologies portent chacune une information pour une image de lettrine : l'ontologie des historiens fournit une représentation sémantique des lettrines, alors que l'ontologie des régions d'intérêt en est une représentation bas niveau. C'est pourquoi nous les avons combinées en une seule ontologie des lettrines en liant la classe principale *Lettrine* de l'ontologie des historiens à la classe principale *Image* de l'ontologie des régions d'intérêt par la propriété *hasImage*. Cette ontologie regroupe l'ensemble des classes et propriétés des deux ontologies ainsi combinées.

Par conséquent, chaque image de l'ontologie des régions d'intérêt est décrite à la fois par ses régions d'intérêt extraites (et leurs caractéristiques bas niveau), ainsi que par les annotations manuelles réalisées par les historiens. Notons cependant qu'une partie seulement des images annotées de l'ontologie métier sont présentes dans l'ontologie des régions d'intérêt. Par conséquent, certaines lettres de l'ontologie des historiens ne sont pas décrites par des régions d'intérêt.

L'intégration au sein d'une même ontologie de l'ensemble des données décrivant une image de lettre permet d'envisager une réduction du fossé sémantique entre les caractéristiques bas niveau de l'ontologie des régions d'intérêt, et leurs annotations manuelles décrites par l'ontologie des historiens.

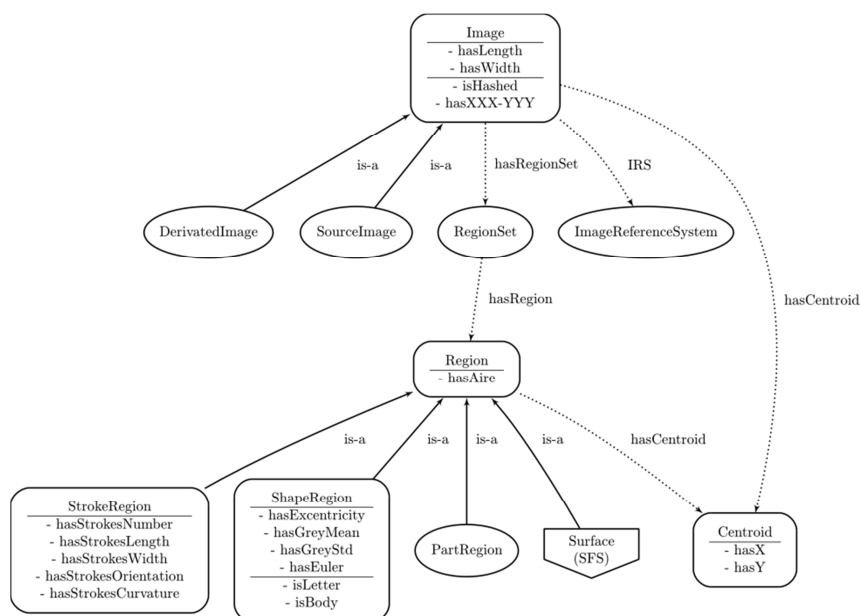


Figure 11. Ontologie enrichie des régions d'intérêt

Notre contribution à une réduction de ce fossé sémantique repose sur un enrichissement de l'ontologie par de nouvelles propriétés permettant d'annoter sémantiquement des images, ou des régions d'intérêt. Il s'agit de propriétés calculées, définies par une formule ou une règle logique, qui peuvent intégrer l'ensemble des données décrivant les images. Ces propriétés sont indiquées comme calculées sur le schéma de la figure 11.

3.4. Inférence sur les régions d'intérêt de la couche forme

La lettre, ou encore des éléments de scènes naturelles des lettrines à fond figuratif, correspondent à des formes de l'image. Plus précisément, la lettre correspond à une grande région d'intérêt de la couche forme située au centre de l'image et contenant peu de trous. C'est pourquoi nous avons cherché à identifier la région d'intérêt de la couche forme correspondant à la lettre à l'aide d'une propriété calculée *isLetter*. Il s'agit là d'un mécanisme d'annotation sémantique de certaines régions au sens image.

Par ailleurs, on observe que des parties de personnages des lettrines figuratives correspondent à de grandes régions d'intérêt de la couche forme que nous avons cherchées à identifier par la propriété calculée *isBody*. Il s'agit de régions suffisamment grandes, qui ne touchent pas les bords de l'image et qui ne correspondent pas à la lettre.

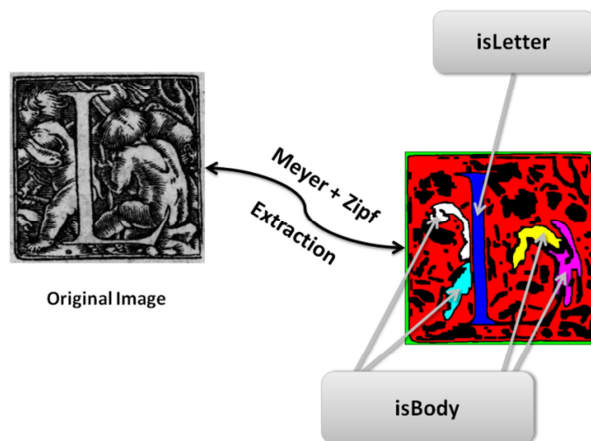


Figure 12. Extraction de formes d'une lettrine

Ces deux propriétés (cf. figure 12) reposent sur les caractéristiques bas-niveau des régions d'intérêt de la couche forme, ainsi que sur un partitionnement spatial de l'image. La propriété *isBody* s'exprime également en fonction de la propriété *isLetter* (une région lettre n'est pas une partie d'un personnage), et d'une information sémantique (seules des régions des lettrines figuratives sont à considérer).

Propriété isLetter : description et expérimentation

Cette propriété calculée, définie pour la classe *ShapeROI*, indique si une région est identifiée comme la lettre d'une lettrine. Elle s'exprime en fonction de quatre propriétés qui intègrent des caractéristiques des régions de la couche forme, mais également une information spatiale avec le partitionnement de l'image.

1) Région située au centre de l'image (*i.e.* contenue dans la partition au centre) : il s'agit de vérifier que la région est contenue dans la partition centrale (propriété spatiale *contains*), et ne chevauche par les régions du bord. Comme l'image est partitionnée suivant la figure 8, cette propriété permet de retrouver la lettre se trouvant dans la partition centrale (partition 5).

2) Région avec peu de trous : le nombre d'Euler, associé à chaque région de la couche forme (propriété *hasEuler* de l'ontologie des régions), est ici utilisé. Après quelques expérimentations, nous avons positionné le nombre d'Euler entre -2 et +2 pour caractériser une région avec peu de trous.

3) contenant le centre de l'image de la lettrine dans son rectangle englobant : cette propriété permet de préciser que le centre de l'image (propriété *hasCentroid* de l'ontologie des régions) doit appartenir au rectangle englobant de la région choisie, avec une marge de 15 pixels.

4) région d'aire maximale : dans l'ontologie des régions, l'aire est indiquée pour chaque région (propriété *hasArea* de l'ontologie des régions). Parmi toutes les régions vérifiant les propriétés précédentes, celle d'aire maximale est retenue et annotée par la propriété *isLetter*.

La propriété *isLetter* a été appliquée sur l'ensemble des 909 lettrines peuplant l'ontologie des régions. Après vérification manuelle, 816 régions s'avèrent être correctement annotées, et 103 ne le sont pas, d'où un taux d'erreur de 11 %. Cependant, cette règle ne garantit pas l'existence d'une région annotée pour chaque lettrine.

Il serait ici possible de comparer la forme de la région avec celle attendue de la lettre, information présente dans l'ontologie des historiens (propriété *hasIdentification-Letter*). Une telle extension permettrait de détecter les cas où la lettre se décompose en plusieurs régions, et ainsi réitérer le calcul de la propriété *isLetter* jusqu'à ce que toutes les régions de la lettre soient détectées.

Propriété isBody : description et expérimentation

Cette propriété, définie pour la classe *ShapeROI*, indique qu'une région est identifiée comme la partie d'un personnage du fond d'une lettrine. Elle s'exprime en fonction de cinq propriétés qui intègrent des caractéristiques des

régions forme, une information spatiale avec le partitionnement, ainsi qu'une information de l'ontologie du domaine :

1) Région d'une lettrine figurative : seules les régions des lettrines figuratives (propriété *hasPattern* de l'ontologie des historiens) sont considérées.

2) Région située au centre de l'image : comme pour la propriété *isLetter*, cette propriété se teste à l'aide de la propriété spatiale *contains*.

3) Région avec peu de trous : comme pour la propriété *isLetter*, il s'agit de tester le nombre d'Euler (propriété *hasEuler* de l'ontologie des régions) associé à chaque région dans l'ontologie des régions d'intérêt.

4) Région de couleur gris-clair : la caractéristique colorimétrique d'une région (propriété *hasGreyMean* de l'ontologie des régions), doit être supérieure à 90 pour indiquer une couleur gris-clair. Le niveau de gris va de 0 (noir) à 255 (blanc le plus intense).

5) Région qui n'est pas une lettre : parmi toutes les régions vérifiant les propriétés précédentes, seules celles qui ne vérifient pas la propriété *isLetter* sont annotées par *isBody*.

La validation des régions annotées est un processus manuel. C'est pourquoi nous avons sélectionné aléatoirement pour la règle *isBody* un sous-ensemble de 45 images de lettrines. Cet ensemble est constitué de 27 lettrines à fond décoratif, 18 lettrines à fond figuratif, à partir desquelles 112 régions forme au total ont été extraites. La propriété *isBody* a permis d'annoter correctement des régions forme de 17 des 18 images figuratives, d'où un taux d'erreur de l'ordre de 2 %.

La figure 15 représente la lettrine figurative dont les régions ne sont pas correctement annotées. On peut y observer que la région jaune¹ correspondant à une partie de personnage n'est pas annotée car son nombre d'Euler est trop grand. En revanche, la région blanche est annotée alors qu'elle ne correspond pas à une partie de personnage. L'utilisation d'un partitionnement plus complexe de l'image, ou encore d'autres descripteurs statistiques des régions, pourrait permettre d'annoter correctement les régions de ce type.

¹ Les couleurs ne sont pas visibles sur cette version imprimée

3.5. Inférence sur les régions d'intérêt de la couche texture

Un mécanisme d'inférence sur les régions d'intérêt de la couche texture a également été mis en place dans le but de réduire le fossé sémantique entre l'ontologie des régions d'intérêt et l'ontologie des historiens.

On observe qu'une lettrine hachurée contient des régions de traits dont la superficie totale est suffisamment grande pour couvrir une bonne partie de l'image. Ce sont donc les caractéristiques de l'ensemble des régions de traits qui portent une information pertinente au niveau de l'image elle-même pour déterminer si elle est ou non hachurée. Nous avons cherché à identifier les lettrines hachurées à l'aide de la propriété calculée *isHashed* d'une image.

Pour cela, nous avons tout d'abord enrichi la description d'une image avec des propriétés décrivant l'ensemble des régions de traits qui la composent.

La propriété *isHashed* permet de déterminer si une lettrine possède un fond hachuré à partir de caractéristiques bas niveau, et se vérifie en la comparant avec la propriété *hasBackground* de l'ontologie des historiens. C'est pourquoi nous avons pu utiliser une approche classique de classification supervisée (à savoir un arbre de décision C4.5) pour déterminer les seuils des différentes caractéristiques permettant de décider si la lettrine est hachurée ou non.

La figure 13 présente l'arbre de décision simplifié construit à cet effet à l'aide de l'algorithme C4.5, où :

SV : sont des étiquettes des branches de l'arbre conduisant à la classe *isHashed* et sont décrits par : S est un opérateur de comparaison qui peut être <, ≤, > ou ≥ et V est une valeur seuil de l'attribut de comparaison.

YX : sont des nœuds de l'arbre et X correspond à chacune des 6 propriétés spécifiques des régions de traits stockées dans la classe *StrokesROI*, à savoir *hasStrokes-Number*, *hasStrokesLenght*, *hasStrokesWirdth*, *hasStrokesOrientation* et *hasStrokesHomogeneity* ; Y correspond à la valeur minimale, maximale, moyenne, ou à l'écart-type (*Min*, *Max*, *Avg* et *Std*) de chacune des 6 caractéristiques de l'ensemble des régions de traits constitutives de l'image.

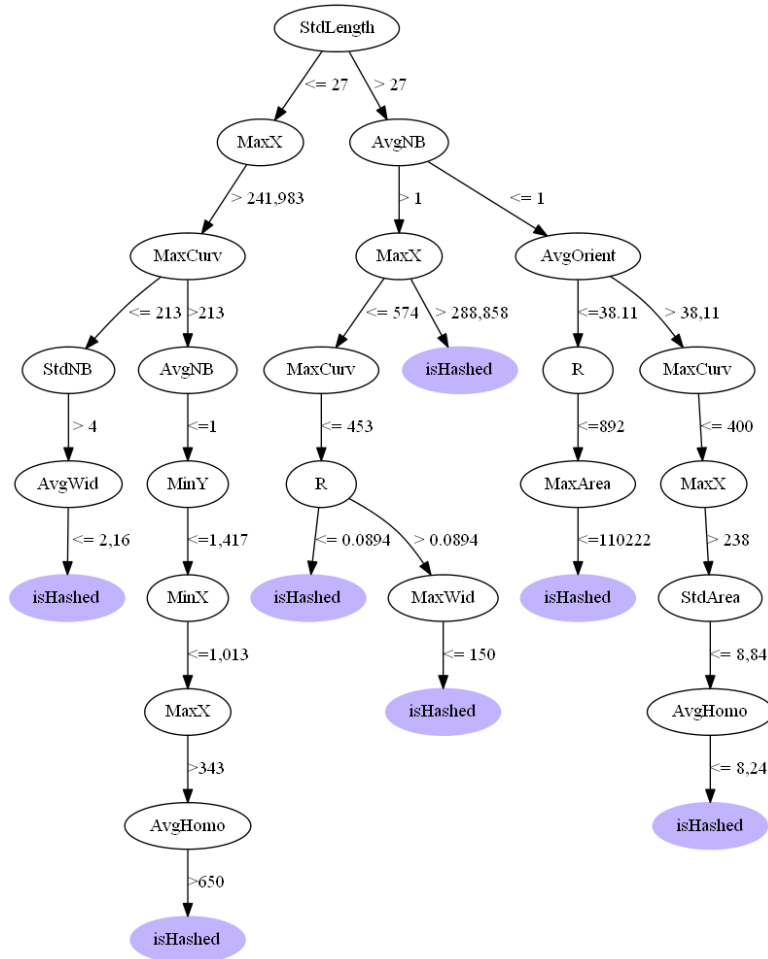


Figure 13. Arbre de décision simplifié permettant de trouver les règles

(*isHashed*) : comparaison des valeurs minimale, maximale, moyenne, ou d'écart type des caractéristiques des régions de traits par rapport à des seuils (7 feuilles) données par l'arbre de décision

Cette règle a été appliquée sur l'ensemble des 909 lettres peuplant l'ontologie des régions parmi lesquelles 140 sont hachurées. 126 images sont annotées automatiquement. Une vérification automatique est alors possible en interrogeant l'ontologie métier (des historiens) peuplée manuellement : 123 sont

effectivement hachurées. D'où un rappel de $\frac{123}{140} = 87,8 \%$ et une précision de $\frac{123}{126} = 97,8 \%$.

4. Limites et discussions

Le problème de recherche d'images de documents anciens par leur contenu impose deux contraintes majeures. Premièrement, il faut concevoir un système qui puisse modéliser et structurer les connaissances issues du domaine des historiens (concepts sémantiques), et celles issues du domaine du traitement des images (caractéristiques bas niveau). De plus, il faut que ce système puisse répondre au problème de fossé sémantique, et faire le pont entre les caractéristiques bas niveau et les concepts sémantiques utilisés pour l'interrogation par les historiens. Nous avons proposé d'utiliser le formalisme des ontologies, pour palier à ce problème.

Dans la construction d'une application qui, comme la nôtre, tend à combiner des traitements d'images (dits de bas niveau) et des traitements symboliques (dits de haut niveau), on se trouve confronté à une grande masse de données hétérogènes. Au départ de ces données, des règles sont appliquées pour identifier certaines de ces régions comme des composants sémantiques du document de base (cette région est la lettre dans la lettrine, cette région est une partie d'un corps...). Toutes ces données sont alors rassemblées de manière cohérente au sein d'une ontologie globale pour permettre de nouvelles déductions, et de rechercher une information hétérogène. Bien que tout ceci soit implémenté et fonctionnel, le problème du fossé sémantique ne peut cependant pas être limité à cette vision simpliste.

Afin d'illustrer la complexité du problème, nous proposons d'étudier deux exemples. Ceux-ci mettent en exergue la complexité de la tâche, et afin de faciliter leur visualisation, nous présentons le résultat final d'un enchaînement de traitement d'images suivi de déductions sur les résultats. De telles images sont produites automatiquement à partir des données enregistrées, et ont été utilisées pour confronter nos résultats à une vérité terrain.

Dans le premier exemple (*cf.* figure 14), chaque image reprend la lettrine de départ, puis, en surimpression, la zone qui a été identifiée comme une lettre. La première décrit une identification réussie, et la deuxième correspond à un échec partiel.

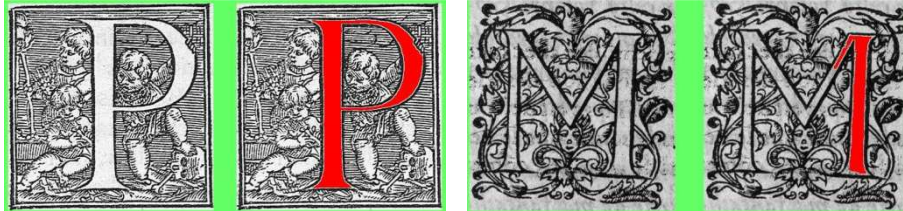


Figure 14. Résultat final d'un enchaînement de traitement d'images

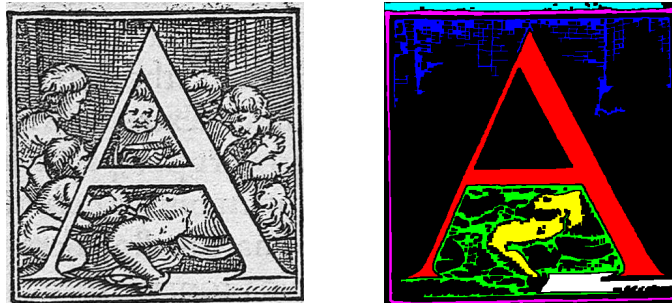


Figure 15. L'image où des zones ne sont pas correctement identifiées comme partie de personnage par *isBody*

Le second exemple (*cf.* figure 15) représente l'image de la lettrine figurative (au motif figuratif) dont les zones ne sont pas correctement annotées par la propriété *is-Body*. Dans la seconde image apparaît en surimpression, l'ensemble des régions extraites de la couche forme. La zone rouge correspondant à la lettre sera correctement annotée par la propriété *isLetter*. En revanche, la zone jaune, qui correspond à une partie d'un personnage, ne sera pas annotée par la propriété *isBody* car son nombre d'Euler est trop grand. La zone blanche sera quant à elle annotée alors qu'elle ne correspond pas à une partie d'un personnage.

Notre approche utilise des règles pour détecter des zones portant un intérêt sémantique. Comme nous l'avons présenté dans cet article, ces règles reposent sur un partitionnement et des caractéristiques relativement simples. Nous envisageons donc d'utiliser un partitionnement plus complexe de l'image, ou encore d'autres descripteurs statistiques des zones, pour permettre une meilleure annotation des zones de ce type. De même une gestion plus souple de la classification (avec l'utilisation de rejet, de degré de confiance, classifieur

flou...) pourrait permettre de proposer à l'utilisateur des régions candidates, mais pour lesquelles le système ne peut prendre de décisions fermes.

Cependant, rien ne garantit que l'utilisation de caractéristiques, et de classifieurs, permette d'obtenir des résultats plus pertinents du point de vue de l'utilisateur. Ceci provient du fait que notre approche n'intègre la sémantique qu'au niveau de l'analyse des résultats de traitements des images, sans prendre en compte la sémantique des formes qui sont extraites. Cette voie nous paraît intéressante pour les années à venir, afin de piloter les différentes phases de traitements de l'information à tous les niveaux (bas niveau et haut niveau).

5. Conclusion et perspectives

Dématérialiser et publier des contenus culturels en ligne est un enjeu actuel, exacerbé avec le développement du numérique. Cependant, le problème de la valorisation et de l'exploitation de ces contenus est non résolu. Il pose la question de l'extraction automatique du contenu, pour traiter la masse de documents, et la question de structurer cette information pour être facilement accessible. Facilement accessible implique la définition d'outils utilisant un langage proche des utilisateurs finaux (experts ou non).

Ainsi, cet article présente des travaux qui visent à réduire le fossé sémantique entre le vocabulaire utilisé par des historiens de la Renaissance et celui du traitement d'images dans le but de proposer un moteur de recherche. Le problème du fossé sémantique est un vaste problème qui nécessite la modélisation des connaissances du domaine d'application et du domaine d'expertise : le traitement d'image dans notre cas.

Afin de palier à ce problème, nous avons utilisé des ontologies pour définir une sémantique explicite entre concepts visuels et langage naturel, et surmonter le problème de la représentation implicite et cachée d'une connaissance visuelle. De plus, ces ontologies sont utilisées comme une référence commune pour échanger des contenus sémantiques. Elles ont été peuplées par des données issues du domaine d'application et des traitements des images d'une part, et enrichies par des règles d'inférences d'autre part. Ces règles sont utilisées pour créer un lien entre les mots-clés historiens et des régions des images, et ainsi réduire le fossé sémantique. Les résultats obtenus sur notre base nous permettent de valider ce modèle, et d'observer une évaluation très correcte.

En perspective de ces travaux, nous travaillons actuellement à la généralisation de ce modèle pour l'appliquer à des images de documents culturels récents, comme les bandes dessinées. De plus, il pourrait être intéressant de proposer une génération automatique de règles d'inférences à

partir de bouclage de pertinence (l'utilisateur pourrait ainsi valider plusieurs images qui serviraient d'apprentissage à la règle). Enfin, bien que ce modèle propose des résultats très intéressants, il faudrait réfléchir à son optimisation pour l'étendre à de grandes bases de données qui ne sont actuellement pas directement intégrables via des ontologies.

Bibliographie

- 13249-3 :2002 I., Information technology-Database languages ; SQL Multimedia and Application Packages ; Part 3 : Spatial, 2002.
- Allen J.F. (1983). Maintaining knowledge about temporal intervals. In *Communications of the ACM*, 26 November 1983, ACM Press. p. 832-843.
- Bannour H., Hudelot C., (2012, juin). Combinaison d'information visuelle, conceptuelle, et contextuelle pour la construction automatique de hiérarchies sémantiques adaptées à l'annotation d'images. *Actes de la conférence RFLA 2012, Reconnaissance des Formes et Intelligence Artificielle*, p. 460-469.
- Clouard R., Renouf A., Marinette R. (2010). An Ontology-Based Model for Representing Image Processing Application Objectives. *International Journal of Pattern-Recognition and Artificial Intelligence*, vol. 24, p. 1181-1208.
- Cohn A., Bennet B., Gooday J., Gotts N. (1997). Representing and reasoning with qualitative spatial relations about regions. *Spatial and temporal reasoning*, Kluwer, p. 97-134.
- Coustaty M. (2011). *Contribution à l'analyse complexe de documents anciens. Application aux letrines*. PhD thesis, Université de La Rochelle, France.
- Coustaty M., Pareti R., Vincent N., Ogier J.-M. (2011). Towards historical document indexing : extraction of drop cap letters. *IJDAR*, vol. 14, n° 3, p. 243-254.
- Dubois S., Lugiez M., Péteri R., Ménard M. (2008). Adding a Noise Component To A Color Decomposition Model For Improving Color Texture Extraction. *Proc. of the 4th European Conference on Colour in Graphics, Imaging, and Vision (CGIV'08)*, p. 394-398.
- Ferré S. (2009). *Navigating the Semantic Web with Logical Information Systems*. Rapport Recherche Interne.
- Gruber T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, vol. 43, p. 907-928.
- Hamidi A.E., Ménard M., Lugiez M., Ghannam C., (2010). Weighted and extended total variation for image restoration and decomposition. *Pattern Recognition*, vol. 43, n° 4, p. 1564-1576.

- Inc. O.G.C. (1999). *OpenGIS Simple Features Specification for SQL*. OpenGIS Project Document 99-049.
- Jena (2011). *Jena - A Semantic Web Framework for Java*. <http://jena.sourceforge.net/>
- Jeon J., Lavrenko V., Manmatha R. (2003). Automatic image annotation and retrieval using crossmedia relevance models. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, SIGIR '03*, ACM, New York, NY, USA, p. 119-126.
- Jimenes R. (2008). *Les Bibliothèques Virtuelles Humanistes et l'étude du matériel typographique*. Technical report, Centre d'Étude Supérieur de la Renaissance,
- Knublauch H., Ferguson R., Noy N., Musen M. (2004). The protégé OWL plugin : An open development environment for Semantic Web applications. *International Semantic Web Conference (ISWC)*, vol. 3298, p. 229-243.
- Kompatsiaris Y., Hobson P. (ed.) 2008. *Semantic Multimedia and Ontologies*, vol. 1, Theory and Applications, Springer.
- Liu Y., Zhang D., Lu G., MaW.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, vol. 40, n° 1, p. 262-282.
- Neumann B., Möller R., (2008). On scene interpretation with description logics. *Image Vision Comput.*, vol. 26, January, p. 82-101.
- Newell A., (1982). The Knowledge Level. *Artif. Intell.*, vol. 18, n° 1, p. 87-127.
- Pareti R., Vincent N. (2006). Ancient Initial Letters Indexing. *Proc. of the 18th International Conference on Pattern Recognition (ICPR'08)*, IEEE Computer Society, Hong Kong, China, p. 756-759.
- PrattW. (2007, February). *Digital Image Processing : PIKS Scientific Inside*. 4^e ed, Wiley-Interscience.
- Smeulders A.W.M., Worring M., Santini S., Gupta A., Jain R. (2000). Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, p. 1349-1380.
- Zhang D., Islam M.M., Lu G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, vol. 45, p. 346-362.