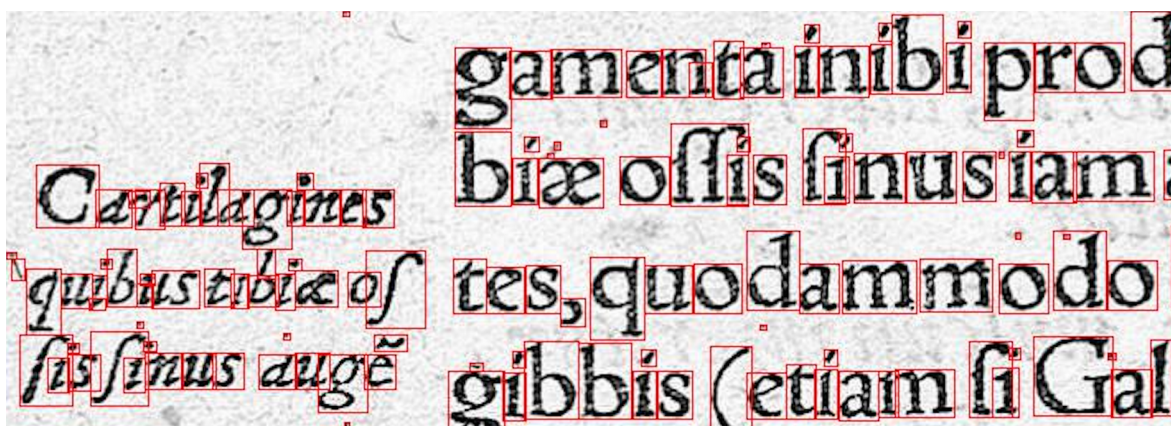




**Remise :** Dépôt sur Celene, ou envoyé à [brouard@univ-tours.fr](mailto:brouard@univ-tours.fr) (scan ou PDF)

La transcription de documents anciens consiste à annoter chaque élément d'une page de façon à permettre différents traitements tels que la reconstitution du texte sous une forme exploitable informatiquement, la comparaison de caractères ou de lettrines à travers différents documents, la recherche rapide d'un élément d'après des indications, ou d'après une image, l'extraction de caractères d'une même fonte, etc. L'image ci-dessous montre ce qu'on entend par élément dans une page (c'est une phase préliminaire de détection, différents « filtres » sont appliqués ensuite de façon à éliminer ce qui ne semble pas avoir d'intérêt et « reconstruire » des caractères parfois dégradés. Chaque élément encadré est potentiellement une *composante* (parfois c'est un « bruit » – une tache par exemple – qu'on ne mémorisera pas, parfois c'est une composante – le « C » en haut de la marge, ou le « g » débutant le paragraphe – et parfois la composante est fractionnée – c'est systématiquement le cas pour les « i » avec leur point) :



Sous l'hypothèse qu'on dispose d'un logiciel capable de traiter une image afin d'en extraire les coordonnées des différentes composantes, voici quelques éléments concernant le système d'informations qu'on veut mettre en service :

Le but est de conserver des informations sur des ouvrages : titre, personnes connues (auteurs, imprimeur, éditeur) et lieux associés, nombre de pages, année d'édition, commentaires... Chaque page de l'ouvrage est numérisée et conservée sous la forme d'un fichier image. Ce sont ces fichiers qui sont analysés pour en tirer les informations sur les composantes : chaque composante a des coordonnées (abscisses x et ordonnées y du coin haut gauche et bas droite du rectangle englobant la composante), un type et un commentaire (ce dernier est optionnel). Elle peut aussi avoir des mots-clés, faire partie d'une police de caractères, et elle peut être reliée à une autre composante, avec un score de similarité (un

nombre réel qui est d'autant plus grand que les 2 composantes se ressemblent, ce qui permet de retrouver par exemple des composantes d'après le fait qu'elles se ressemblent). Une lettrine ou un caractère sont des exemples de types de composantes.

### Questions relatives au sujet

- 1 ► Proposez un MCD
- 2 ► Donnez le MLD associé

### Interrogation de la base

- 3 ► Quels sont les titres de ouvrages dont l'auteur est « Andreas Vesalis » ?
- 4 ► Quelle est la liste des fichiers image, ordonnée par n° de page, de l'ouvrage n° 28 ?
- 5 ► Quelle est la liste des composantes dont les coordonnées y se situent entre 188 et 224, dans la page n° 54 de l'ouvrage n°28, classée par x croissants (le x du coin gauche) ?
- 6 ► Lister l'ensemble des n° de page et les coordonnées des composantes de type lettrine présentes dans l'ouvrage n°28, classé par n° de page.
- 7 ► Quel est l'ensemble des mots-clés associés aux lettrines et à l'auteur «Joachim Du Bellay » : mot-clé, et nombre de fois où il est rencontré
- 8 ► Quelle est la liste des noms de polices de caractères rencontrées dans l'ouvrage n°143 ?
- 9 ► Quels sont les ouvrage (titre, année) dont des composantes de type « lettrine » sont associées au mots-clés « église » ?
- 10 ► Quelles sont les années d'édition et les titres des ouvrages possédant des composantes dont le score de similarité avec la composante n° 2345660 est au moins égal à 0,88 ?

◇ ◇ ◇    Fin du sujet    ◇ ◇ ◇