

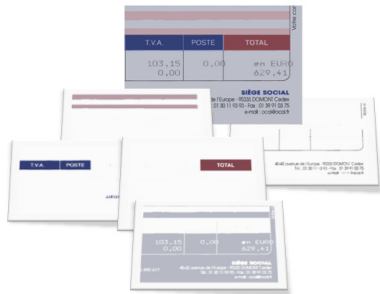
Description

PLEIAD est une plateforme de mutualisation des études et réalisations faites au sein du groupe Documents de l'équipe Imagine. Elle propose un socle commun de méthodes accessibles aux chercheurs et doctorants et s'inscrit dans une réelle démarche d'amélioration incrémentale. Elle propose des solutions logicielles issues des grands projets de dématérialisation, valorisation et d'exploitation des documents numérisés auxquels l'équipe Documents d'IMAGINE a fortement contribué. Cette plateforme repose en grande partie sur la constitution de la bibliothèque de méthodes LibCRN dédiée au traitement des documents et disponible en téléchargement sur SourceForge sous licence LGPL. Elle est écrite en C++ et fonctionne sous Windows, Linux, Mac OS X et Android.

Annotation en plans couleurs

Pixl-DOD (depuis 2011)

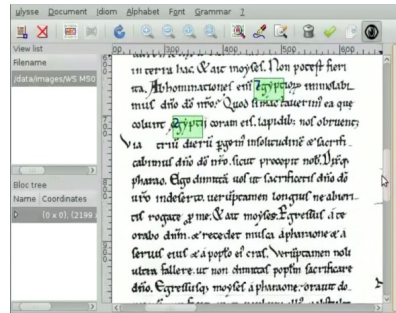
Les fonctionnalités développées dans ce module portent sur les problématiques de modélisation colorimétrique des documents, de correction colorimétrique, travaillant sur des espaces couleurs adaptés. Elles permettent de traiter des documents administratifs couleur de structure complexe. Il est possible via cette interface de réaliser une séparation en plans couleur et d'extraire toute surcharge colorée moyennant un apprentissage léger.



Recherche par similarité de formes et de mots

Ulysse (depuis 2008)

Cet outil de recherche d'occurrences de mots clés dans des textes anciens a été conçu pour compenser les insuffisances des outils OCR sur des textes multilingues ou pour lesquels il n'est pas possible de réaliser de reconnaissance performante.

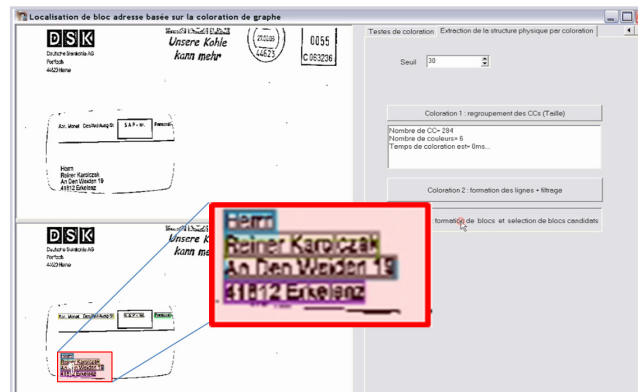


Une sélection par l'exemple (en image ou en texte) permet de lancer une recherche sur une base d'images et de trier les résultats selon leur similarité.

Repérage temps réel de zones d'intérêt importantes

CESA-Vinci

Cet outil à vocation plus industrielle permet de sélectionner en temps réel des zones d'intérêt dans les images (bloc adresse par exemple) afin d'en faciliter la lecture (grâce au ciblage rapide). Il permet également de réaliser du tri automatique de documents selon leur contenu.



Contact

pleiad@liris.cnrs.fr



*Plateforme
d'Enrichissement,
d'Indexation et
d'Analyse de
Documents*



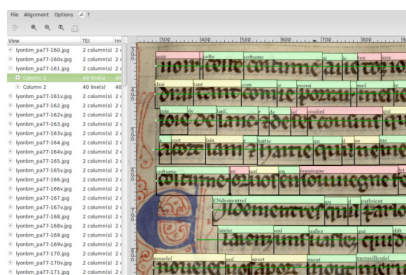
Objectifs

Le traitement manuel des grandes masses d'images est devenu impossible et la multiplication des fonds numériques due aux très nombreux projets de numérisation de ces dernières années (Google, Europeana, Gallica ...) pose aux usagers des problèmes d'accès au contenu. Il est devenu nécessaire de trouver des approches efficaces et innovantes pour caractériser, traiter et analyser ces images de documents de façon automatique. C'est dans ce contexte que s'inscrit la plateforme Document PLEIAD. Cette plateforme est une vitrine qui présente des différentes actions et projets en cours autour de cette activité d'Analyse et de Reconnaissance de Documents. Des projets phares sont menés en collaboration avec des partenaires académiques en Sciences Humaines et Sociales et des partenaires industriels et ont nourri la plate-forme. C'est le cas des travaux qui concernent les documents du patrimoine, autour de projets ANR et de collaborations à orientation plus industrielles concernant plus directement le traitement d'images de documents d'entreprises ou administratifs.

Alignement image-transcription

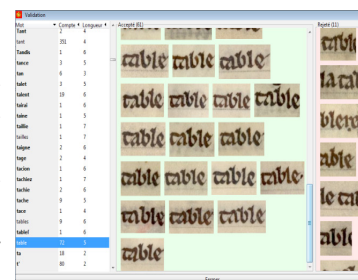
Oriflamms (depuis 2012)

Dans le cadre du projet ANR Oriflamms, un outil interactif d'alignement entre l'image des textes manuscrits et leur transcription encodée en XML a été réalisé.



Cette interface produit un alignement automatique qu'il est possible de modifier manuellement en déplaçant les extrémités des limites entre mots.

Un second outil de visualisation des occurrences de mots identiques a été conçu comme interface de validation de l'alignement. Il permet de réaliser des corrections de façon interactive par l'expert SHS.



Dématérialisation des manuscrits d'auteurs

Bouvard (depuis 2008)

Cet outil a été conçu comme soutien aux réflexions menées entre le LIRIS et l'Institut des Sciences de l'Homme (ISH) pour l'accompagnement du travail des experts SHS pour l'accès, la navigation et l'enrichissement d'œuvres littéraires rares. Il contient des fonctionnalités d'accès aux contenus par fragments textuels dont l'extraction peut être automatique ou manuelle en interaction avec l'expert (selon un encodage SVG). Il offre des fonctionnalités de « bureau virtuel » pour le classement des fragments et leur restitution à tout moment.

Fonctionnalités

- Alignement image-transcription
- Dématérialisation des manuscrits d'auteurs
- Recherche par similarité de styles
- Annotation en plans couleurs
- Recherche par similarité de formes et de mots
- Repérage temps réel de zones d'intérêt importantes

Recherche par similarité de styles

Graphem (depuis 2008)

Dans le cadre du projet ANR Graphem, un outil de recherche par similarité de styles d'écriture a été produit afin d'aider le classement réalisé par les experts paléographes et les aider à découvrir les indices visuels ayant conduit à faire évoluer l'écriture à travers le Moyen-âge.

