

PROTOTYPE GALLICA IMAGES

Réalisation d'un prototype de moteur de recherches d'images à partir de données colorimétriques ou de proportions.

Aucune méta-donnée bibliographique (notice, tdm, ocr, légendes, ...) ou de production n'est utilisée dans la définition des similitudes.

Construction d'un premier index de plus de 620 000 images

Pour la pré-sélection des images :

Un batch calcule des statistiques sur chaque canal de couleur

Un seul canal pour les images en niveau de gris

Trois canaux pour les images couleurs (rouge, vert, bleu)

Les images sont ensuite regroupées par intervalle dans l'application Web

5 statistiques calculées par l'outil ImageMagick à partir de la distribution de chaque couleur dans l'image :

Moyenne (Mean)

Ecart-type (Standard deviation)

Asymétrie (Skewness)

https://fr.wikipedia.org/wiki/Asym%C3%A9trie_%28statistiques%29

Courbure (Kurtosis)

<https://fr.wikipedia.org/wiki/Kurtosis>

Entropie (Entropy)

Pour le tri des images :

Un batch calcule l'empreinte FCTH (fuzzy color and texture histogram) dont l'algorithme est fourni dans le projet Lucene/LIRE

<http://www.lire-project.net/>

L'empreinte FCTH caractérise chaque image par un tableau de 192 octets

Les images présélectionnées sont ensuite triées selon l'algorithme de calcul de distance Tanimoto dans l'application Web

Cela permet :

un système qui à partir d'une image Gallica propose d'autres images Gallica potentiellement similaires

un système qui à partir d'un fichier importé ou adressé par son URL propose des images Gallica potentiellement similaires (c'est-à-dire un moteur de recherche inversée d'images)

une API capable d'interroger le fonds d'images Gallica selon : des intervalles de couleurs, des proportions (largeur/hauteur), une palette (noir et blanc, ou couleur)

En complément :

une mise en avant des collections spécialisées
(estampes, photographies, affiches, monnaies et médailles, arts du spectacle, cartes et plans, ...)

une nouvelle forme de navigation

une valorisation particulière pour les cartes Gallica :
affichage simultané avec une cartographie
contemporaine