

FindImageDuplicates en 10 minutos al día

tecnologías usadas

- findimagedupes
- bash
- mustache
- bootstrap
- CodeIgniter

objetivo y método

se pretende buscar las imágenes iguales o muy iguales de todas las disponibles en la plataforma 10 minutos al día para eliminar esos duplicados.

al no poder instalar el programa `findimagedupes` en la plataforma de Dreamhost, optamos por:

- hacer una copia de todas las fotos en local con la herramienta `rsync` (para agilizar las actualizaciones)
- buscar duplicados (o similares muy similares) con la herramienta `findimagedupes`
- del fichero TXT generado por `findimagedupes`, generar un fichero YML con un formato específico para poder usar con `mustache`
- usar `mustache` para generar a partir de los datos de ficheros duplicados un HTML donde poder visualizar las diferentes imágenes detectadas como duplicadas y eliminar cuando convenga estas imágenes
- para la eliminación, hemos desarrollado en el controlador y modelo de la plataforma 10 minutos al día unas funciones específicas que se encargan de borrar de la BDD y del FS el fichero correspondiente
- queda por desarrollar la parte AJAX del proyecto, para que la eliminación sea asíncrona y así poder ir eliminando diferentes imágenes sin perder la situación.

uso

- al hacer un UPLOAD de imágenes, habrá que hacer un RSYNC con el repositorio local
 - las imágenes borradas de otras cribas desaparecerán con el RSYNC
- ejecutar el procedimiento de búsqueda de duplicados
 - hay que hacer que las imágenes que han sido borradas desaparezcan de la BDD de `findimagedupes`
 - <http://www.jhnc.org/findimagedupes/manpage.html>
 - `-purge + -f=`
- hacer la comprobación y eliminar para que no se envíen los duplicados detectados

mejoras/problemas

- ¿qué pasa con las fotos que no son duplicados pero que `findimagedupes` detecta como tal?
- ¿qué pasa con las fotos existentes pero que no constan en la BDD? automatizar eliminado ¿cómo?
- automatizar proceso de `rsync` + `findimagedupes` + `mustache`
- petición AJAX tropieza con protección CodeIgniter
 - «No 'Access-Control-Allow-Origin' header is present on the requested resource. Origin 'null' is therefore not allowed access. The response had HTTP status code 500.»

Last update: 24/10/2017 misproyectos:findimagedupes10minutos https://miguelangel.torresegea.es/wiki/misproyectos:findimagedupes10minutos?rev=1508886468
16:07

From:
<https://miguelangel.torresegea.es/wiki/> - **miguel angel torres egea**



Permanent link:
<https://miguelangel.torresegea.es/wiki/misproyectos:findimagedupes10minutos?rev=1508886468>

Last update: **24/10/2017 16:07**