

Copia.me

Ya se trate de código, literatura o artículos académicos, a nadie le gusta que la gente se copie (bueno, al menos no sin citar al autor). Así que a la gente de la consultora *2Diseños*, que estaba sin proyectos, se les ocurrió crear un servicio para detectarlas (y ganar un poco de plata, de paso): *Copia.me*.



El problema

Copia.me será un sistema en el que los usuarios subirán documentos de distintos tipos con el fin de detectar si estos son una copia o un plagio. Por ahora soportaremos textos, programas y partituras pero se espera incorporar más. También se contará con revisores freelancer que se encargarán de analizar manualmente los documentos para darles un análisis más inteligentes.

Servicios

El sistema contará con dos grandes funcionalidades (llamadas *servicios*):

- *Búsqueda de plagios*: ideado para contenidistas. Se cargará un documento original y una lista de documentos consideradas potenciales copias. *Copia.me* deberá determinar si alguno de ellos es un plagio del original.
- *Búsqueda de copias*: enfocado en docentes. Se cargará un lote de documentos y *Copia.me* detectará cuales son copias entre sí.

Calidades de servicio

Los usuarios de *Copia.me* podrán contratar tres niveles de calidad de servicio (de diferente costo):

- **Bronce**: sólo realiza detección automática, usando diferentes algoritmos, dependiendo del tipo de contenido:
 - Los documentos legales, literarios y académicos se comparan usando **Distancia de Levenshtein**
 - Los fragmentos de código se analizan usando **Detección de clones basado en árboles sintácticos**
 - Las partituras musicales se analizan usando **Acoustic fingerprint**
- **Plata**: además de la detección automática, envía un cierto porcentaje (configurable por el usuario) de los documentos a analizar contra revisores freelancers. Cada revisor recibe un email con cada par de documentos asignados.
- **Oro**: además de la detección automática, envía todos los documentos a los revisores, y además hace una revisión cruzada: un cierto porcentaje de los mismos (nuevamente, configurable por el usuario) los envía a más de un revisor.

Obviamente, contar con un gran equipo de revisores freelancers no es tarea fácil, así vamos a asumir algunas cosas:

1. De cada freelancer sabemos cuántos documentos puede revisar por máximo cada mes, y su email.
2. Siempre habrá revisores disponibles. **Bonus**: si no hubiera suficientes, avisar por mail a quienes administren el sistema.
3. Los revisores ingresarán al sistema y cargarán los resultados de sus análisis, siempre en el plazo máximo de 1 día.

4. Para la comunicación con los revisores, ya contamos con un componente que nos permite enviar mails.

Y por otro lado, los algoritmos automáticos de detección de copias son costosos, así que los ejecutaremos en horas particulares del día en que el procesamiento es más barato (por ejemplo, a la madrugada).

Clasificación

Finalmente, a partir de los análisis automáticos y manuales, Copia.me tiene que poder clasificar a los documentos de la siguiente forma:

- no revisado: se da cuando (aún) no fue analizado de ninguna forma;
- original: se da cuando todos los análisis lo marcaron como original;
- dudoso: se da cuando algunos análisis lo marcaron como copia, y otros como original;
- copia: se da cuando todos los análisis lo marcaron como copia.

En cualquier momento el usuario podrá consultar el estado de revisión de cada documento. **Bonus:** poder saber cuando el servicio se terminó, es decir, cuando todas los análisis se concretaron para todos los documentos.

Se pide:

Comunicar una solución usando prosa, diagramas y código o pseudocódigo:

1. Cargar un servicio con todos sus datos necesarios, de forma programática. Por ejemplo:

Dar de alta un servicio de búsqueda de plagios, nivel Plata con revisión manual del 50%, para detección de plagios, para un cierto documento original y otros 35 que son potenciales copias.

2. Realizar el análisis automático de los documentos a ciertas horas del día (aún no definimos cuáles).
3. Notificar a los revisores. Por ejemplo:

Servicio #4530: Maria tiene que que comparar el documento A contra el B , Juan el D contra el E y Ana el F contra el G.

Servicio #4531: Maria tiene que compara el documento M contra el N y el P contra el O. Juan y Ana no tienen que comparar nada porque ya completaron su cuota de revisiones mensuales.

4. Consultar el estado de cada documento de un servicio. Por ejemplo:

*El documento A inicialmente está **no revisado**, luego el algoritmo automático lo considera una copia y pasa a ser **copia**, pero luego Ana lo marca como original y pasa a ser **dudoso**.*

5. Consultar el precio de un servicio, el cual depende la cantidad de documentos involucrados y la calidad de servicio.