

Problema a resolver

Ver fotos de productos es muy útil tanto para los usuarios como para los vendedores de estos pero las búsquedas de fotos en internet suelen dar muchos resultados redundantes o inútiles.

Se han hecho muchos estudios y hay muchos métodos para la búsqueda de imágenes en la red pero se ha hecho muy poco sobre búsqueda de imágenes orientada a productos.

La búsqueda de imágenes de productos tiene sus propias propiedades. Suele haber imágenes de gran calidad asociadas a estos productos creadas por las empresas que los venden pero hay pocas y no dan toda la información que el usuario necesita, por ejemplo en Amazon solo suele haber unas 5 imágenes por producto. En internet hay muchas más fotos de estos productos que podemos encontrar con el buscador de imágenes de Google pero los resultados pueden no servir porque se indexan mediante información en texto.

Sería útil un método que pueda recoger imágenes desde varios ángulos y con una descripción para meterlo en una base de datos. La diversidad y la relevancia son importantes para la experiencia de usuario en la búsqueda.

Aspectos novedosos

En este trabajo se combinan los buscadores de Google y Amazon para encontrar imágenes relevantes de un producto. Se encontraran imágenes de alta calidad en Amazon y se añadirán algunas mas encontradas en Google. Si no se encuentra un producto en Amazon se buscará en Google

La búsqueda se hace mediante un algoritmo propio que visualiza los productos. Es un algoritmo generalizado y se puede utilizar en aplicaciones similares.

Al contrario que con métodos anteriores aquí se estima la relevancia del ranking y explora mediante el algoritmo creado.

Como se resuelve el problema

Siempre se ha utilizado la propagación de afinidad (AP) convencional la cual no puede controlar todas las características que se necesitan.

Se aplica la AP al ranking de las imágenes. Heurísticamente se convierten en puntuaciones de relevancia o confianza.

Aquí se aplicarán cuatro métodos diferentes y se compararán los resultados de estos.

- Mejores imágenes de Google.
- Combinación imágenes de Amazon y las mejores de Google.
- Imágenes de Amazon y Google con el algoritmo AP.
- AP con ranking de Google.

Como se valida la propuesta y experimentos

Primero se utiliza Google para hacer una estimación de relevancia con 100 productos. El experimento se hace con 300 productos de las secciones de electrónica y muebles de Amazon. Estos productos tienen intersección con ninguno de la estimación.

Tras las pruebas se observa que las mejores búsquedas son las del método amazon+google_top. En segundo lugar queda amazon+google_cap.

Al grupo de fotos de Amazon, de 1 a 7 se le quitan los duplicados de forma simple.

Para Google se utilizan 200 imágenes y se quitan las repetidas con un método complejo. En este método se utiliza un parámetro alfa.

Para ajustar el parámetro se hacen 30 búsquedas y se generan resultados con diferentes valores. 5 usuarios observan los resultados y eligen las mejores imágenes en términos de visualización para un producto y se compara la aproximación con los cuatro métodos.

Una vez terminada la aproximación se comparan los métodos en términos de relevancia y diversidad, se marca la relevancia de la imagen según la votación de 3 usuarios.

Se evalúa la diversidad de los ejemplares. Si hay mayor diversidad se cubren mas apariencias visuales del producto. Se definen niveles de similaridad, si estos son bajos es que hay mucha y se observa que los valores mas bajos son los de amazon+google_cap dando muchas más diversidad por lo que se elige este método para el experimento final.

Ahora 15 usuarios hacen la búsqueda de estos 300 productos y se compara con el método seleccionado.

Conclusión

Después de todas las pruebas se demuestra que el mejor método es el que utiliza amazon+google_cap ya que los resultados tienen diversidad y relevancia.

Se propone para un futuro el uso de mas motores de búsqueda