

## MULTIMEDIA



# Búsqueda por contenido multimedia

En la actualidad nos encontramos en medio de un diluvio de datos digitales multimedia. Recientemente se ha reportado [Quora11] que cada día se suben aproximadamente 200 millones de fotos a Facebook. En el caso del video, se estima que para el año 2016 se transmitirán a través de Internet en un mes el equivalente a seis millones de años de video, y que en cada segundo se transmitirá un promedio de 1.2 millones de minutos de video. Asimismo, se estima que para el mismo año el tráfico de video en Internet corresponderá al 55% del tráfico total en la Red [Cisco12]. Todo este diluvio digital está siendo impulsado por la alta disponibilidad de aparatos productores de contenido multimedia (como cámaras digitales y de video, smartphones o tablets) y la necesidad de compartir dicha información, facilitado por el boom de las redes sociales.

Las aplicaciones derivadas de la producción y uso de datos multimedia son muy variadas, por ejemplo en áreas como la biometría, la bioinformática, la industria manufacturera, la industria del entretenimiento, los servicios financieros, la medicina, la química, la biología, y un sinnúmero de otras áreas. Las problemáticas asociadas a Multimedia son asimismo muy amplias: cómo acceder a colecciones masivas de datos multimedia, cómo publicarlos y distribuirlos, cómo interactuar con ellos, cómo garantizar su autenticidad, cómo sacarles el máximo provecho en nuestro quehacer diario y en las distintas tareas productivas, etc. Surge entonces la necesidad de realizar investigación en una amplia gama de dominios distintos para resolver todas las problemáticas mencionadas.



### **Benjamin Bustos**

*Profesor Asistente DCC Universidad de Chile. Doctor en Ciencias Naturales de la Universidad de Konstanz, Alemania (2006); Magíster en Computación (2002) e Ingeniero Civil en Computación, Universidad de Chile (2001). Áreas de investigación: Bases de Datos Multimedia, Búsqueda por Similitud e Indexamiento de Bases de Datos Espaciales y Métricas. [bebustos@dcc.uchile.cl](mailto:bebustos@dcc.uchile.cl)*

## TIPOS DE DATOS MULTIMEDIA

Los datos multimedia se pueden dividir en distintos tipos. Dentro de los más comunes se encuentran:

- **Imágenes.** Conceptualmente es una matriz donde se almacena la información del color representado en cada píxel. Existen distintos formatos de archivos de imagen, algunos de ellos comprimen la información con pérdida (por ejemplo, jpg) y otros sin pérdida (por ejemplo, png).
- **Video.** Es una secuencia de *frames*, en donde cada *frame* representa una imagen en un instante de tiempo. Para dar la sensación de continuidad, un video tiene entre 25 y 30 *frames* por segundo.
- **Modelo 3D.** Son versiones digitales de objetos que pueden ser reales (autos, tazas, árboles, barcos, etc.) o ficticios (creados por un diseñador). Por lo general se representan utilizando una malla de triángulos, aunque existen otras representaciones posibles (por ejemplo, nubes de puntos en el espacio 3D o los denominados *surfels*, que son discos que cubren la superficie del modelo).
- **Serie temporal.** Es una secuencia de valores, provenientes de alguna fuente, medidos a lo largo del tiempo. Cualquier dato que pueda obtenerse en intervalos regulares de tiempo puede ser representado usando una serie temporal.
- **Texto.** También se puede considerar al texto plano del documento como información multimedia, por ejemplo cuando un documento no tiene estructura o ésta no se utiliza para describir su contenido.

## ÁREAS DE INVESTIGACIÓN

La investigación en Multimedia abarca distintos ámbitos de la Ciencia de la Computación, entre los cuales se cuentan Computación Gráfica, Procesamiento de Señales (imágenes, audio, etc.), Recuperación de Información, Bases de Datos, Interacción

Humano Computador, Sistemas Distribuidos, etc. Debido a esto, existen distintas áreas de especialización en Multimedia. Por ejemplo, la conferencia ACM Multimedia este año 2012 solicitó contribuciones en las siguientes áreas [ACM12]:

1. **Media Content Analysis and Processing.** Esta área se enfoca en el estudio de métodos para comprender y procesar información multimedia, por ejemplo algoritmos de extracción de características, segmentación y detección de objetos, algoritmos de aprendizaje de máquina para el análisis de contenido, etc.
2. **Multimedia Activity and Event Understanding.** Se enfoca en detectar eventos o actividades (información semántica de alto nivel) a partir del contenido multimedia.
3. **Multimedia Search and Retrieval.** Se enfoca en la búsqueda y recuperación de información a partir del contenido multimedia, para resolver tareas como búsqueda o detección de objetos, detección de copias, detección de "casi duplicados", etc., en grandes volúmenes de datos.
4. **Mobile and Location-Based Media.** Se enfoca en la investigación multimedia (análisis, búsqueda, interfaces, despliegue de contenido, etc.) para ambientes móviles.
5. **Social Media.** Se enfoca en estudiar cómo combinar la información generada en las redes sociales (tags, enlaces, información asociada a comunidades, etc.) con los datos multimedia asociados a dicha información.
6. **Multimedia Systems and Middleware.** Se enfoca en la investigación de sistemas multimedia en sus diferentes componentes, por ejemplo sistemas operativos, arquitecturas distribuidas, o representación de medios continuos o dependientes del tiempo, tanto a nivel de software como de hardware.
7. **Media Transport and Sharing.** Se enfoca en los métodos de transmisión o formas para compartir datos multimedia, por ejemplo datos producidos por sensores en tiempo real o datos transmitidos en vivo.

8. **Multimedia Security and Forensics.** Comprende temas como la seguridad, privacidad, y autenticación de contenido multimedia.

9. **Multimedia Authoring, Production and Consumption.** Estudia el ciclo de vida del material multimedia (creación, distribución, almacenamiento, acceso, etc.).

10. **Multimedia Interaction and Applications.** Estudia la interacción de los usuarios con datos multimedia en ámbitos como la educación, la salud, las relaciones sociales, etc., y en cómo hacer que dicha interacción humano-multimedia sea lo más natural posible.

11. **Multimedia Art, Entertainment and Culture.** Se enfoca en el estudio del uso de multimedia para la conservación del patrimonio cultural, para la creación artística y para el entretenimiento en general.

En este artículo voy a profundizar en particular en el área de búsqueda y recuperación de información multimedia, específicamente en métodos de búsqueda por similitud basados en el contenido multimedia.

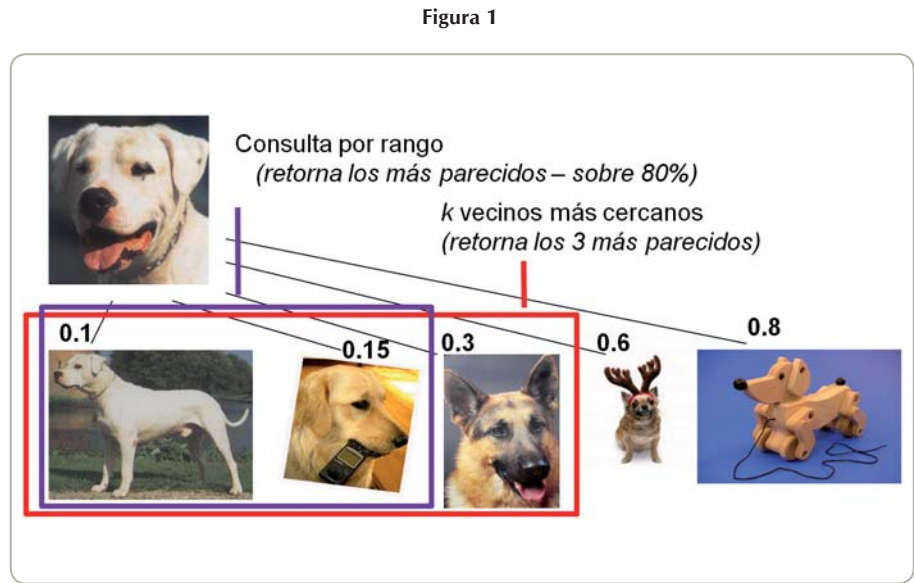
## BÚSQUEDA BASADA EN CONTENIDO EN COLECCIONES DE DATOS MULTIMEDIA

Ya se ha expuesto previamente que la producción de material multimedia ha crecido en forma muy importante en la última década, por la amplia disponibilidad de aparatos de digitalización y que se prevé que el intercambio de datos multimedia en Internet seguirá creciendo en los próximos años. Toda esta vasta cantidad de datos requiere entre otros, de métodos de búsqueda para poder encontrar información relevante. El problema que surge entonces es cómo poder realizar las búsquedas en forma eficiente (rápidamente) y eficaz (obteniendo resultados de buena calidad)

Los buscadores actuales en la Web de información multimedia, por ejemplo Google Images (images.google.com) están basados

en consultas textuales, es decir, el usuario debe ingresar palabras claves con la cuales buscar la información multimedia requerida. Esto ha sido un desarrollo natural a partir de toda la tecnología ya existente utilizada para buscar información textual en la Web. El problema es que los resultados que se pueden obtener a través de estas búsquedas textuales no siempre son satisfactorios, dada la complejidad de los datos multimedia. En efecto, en general los datos multimedia se almacenan directamente como “datos crudos” (la salida directa generada por los algoritmos de digitalización) y es común que no estén asociados a las anotaciones textuales que requerirían los motores tradicionales de búsqueda en la Web. Pregúntese usted mismo, ¿a cuántas de las últimas fotos que ha subido a su red social favorita le ha asociado anotaciones (tags)? Más aún, incluso si se diera el trabajo de anotar todas sus fotos, dichas anotaciones son subjetivas (salvo que sean metadatos que se extraigan automáticamente, como fecha de producción, posición geográfica, autor, etc.) y no necesariamente describen en forma completa lo que aparece en ellas. Por último, es tan vasta la cantidad de información multimedia que se genera constantemente, que sólo para aplicaciones muy acotadas sería posible agregar anotaciones a estos datos en forma manual.

Una forma alternativa para realizar una búsqueda en colecciones multimedia se basa en usar el contenido digital mismo de los datos multimedia para realizarla. Por ejemplo, en el caso de imágenes existen variadas técnicas para analizar en forma automática los colores, bordes, texturas y formas que aparecen en ella, y utilizando directamente esta información se puede realizar la búsqueda. Es decir, en vez de colocar una o varias palabras como términos de consulta (por ejemplo, escribir “torre Eiffel” en el buscador), uno coloca una imagen de consulta (por ejemplo, una foto de la torre Eiffel), la cual es analizada en su contenido y luego se realiza la búsqueda con dicha información. Esta forma de consultar se conoce como “búsqueda por ejemplo” (*query-by-example*) y es la base de los métodos de búsqueda basados en contenido.



Por supuesto, dependiendo del tipo de dato multimedia con el que se desee trabajar y el problema específico en el que se desee desarrollar la aplicación, serán distintas las formas de analizar el contenido digital para poder realizar las búsquedas. No es lo mismo implementar un buscador de fotos en la Web que implementar un método de comparación de radiografías, ya que las características importantes a analizar pueden ser completamente distintas, a pesar de que ambas son imágenes. Una manera de formalizar y generalizar este proceso es utilizando el concepto de espacios métricos, que se detalla a continuación.

## BÚSQUEDA POR SIMILITUD EN ESPACIOS MÉTRICOS

En general, no tiene mucho sentido realizar búsquedas exactas en colecciones multimedia, dado que sólo en el caso de copias digitales el contenido de dos objetos multimedia será igual (bit a bit). Es por esto que en la búsqueda en datos multimedia es fundamental el concepto de búsqueda por similitud.

La búsqueda por similitud consiste en encontrar objetos que sean “parecidos” o

“relevantes” dentro de una colección de datos. En el caso de definir una consulta multimedia usando “búsqueda por ejemplo”, el objetivo consiste en encontrar todos los objetos de una colección multimedia que sean parecidos a la consulta. La similitud entre objetos se define a través de una función que retorna un valor real positivo. Por comodidad, usualmente dicha función representa la disimilitud entre los dos objetos: si la función retorna un valor pequeño es porque los objetos se parecen, si retorna un valor alto es porque son distintos. La Figura 1 muestra un ejemplo de búsqueda por similitud en imágenes y los dos tipos básicos de consulta: por rango y vecinos más cercanos.

El enfoque clásico para realizar búsquedas por similitud basadas en contenido consiste en representar un objetivo multimedia a través de un descriptor. Este descriptor se construye a partir de características o atributos extraídos del objeto multimedia, por ejemplo en el caso de las fotos se puede analizar el color, las formas, las texturas, etc., que aparecen en ellas. Usualmente, las características extraídas son atributos numéricos con el cual se forma un vector característico del objeto multimedia. La Figura 2 muestra un ejemplo del proceso de extracción de características.

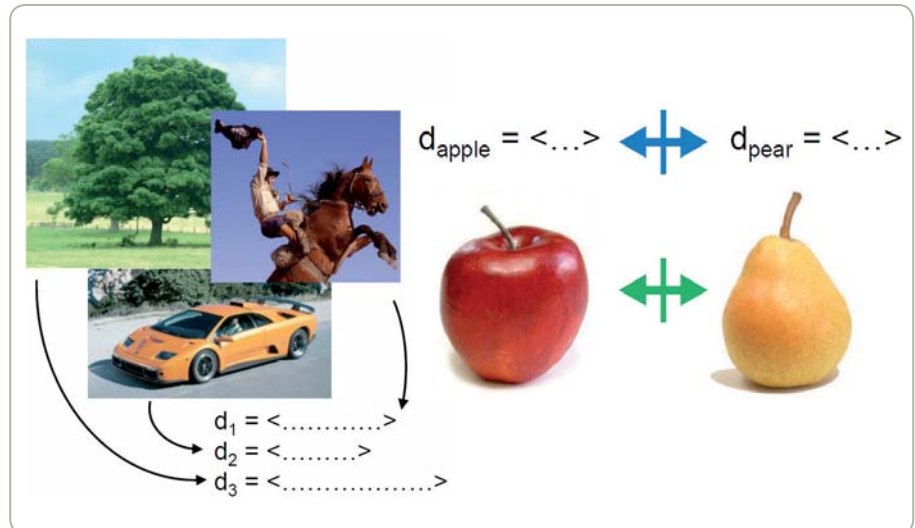
La función de extracción de características se define de forma que si dos objetos multimedia son considerados similares por un ser humano (esto es totalmente dependiente del tipo de dato multimedia y del contexto o aplicación en el cual se desea medir la similitud), entonces sus vectores característicos correspondientes estarán cercanos de acuerdo a alguna función de distancia (la función que calcula la disimilitud). Lo usual es definir la función de distancia de tal forma que cumpla con las propiedades de una métrica (positividad estricta, reflexividad, simetría, y la desigualdad triangular), por lo que la colección de datos junto con la función de distancia forman un espacio métrico. Finalmente, el descriptor de cada objeto multimedia se puede almacenar en una estructura de datos ad hoc (un índice) para luego poder realizar las búsquedas en forma eficiente. En resumen, el problema original de búsqueda se transforma desde la comparación directa entre objetos multimedia a la búsqueda de objetos cercanos en un espacio métrico (ver Figura 3).

Para realizar una búsqueda, se le debe aplicar al objeto de consulta el mismo proceso de extracción de características, y luego se realiza la consulta especificada (por rango o  $k$  vecinos más cercanos). Si bien la consulta se puede responder comparando secuencialmente todo objeto de la colección con la consulta, el uso de un índice puede ayudar a mejorar la eficiencia de la búsqueda. Además, se debe considerar qué tan adecuada es la función de distancia utilizada, así como las características que se están extrayendo para realizar la búsqueda, para que ésta tenga sentido y retorne resultados relevantes.

## EFICIENCIA Y EFICACIA

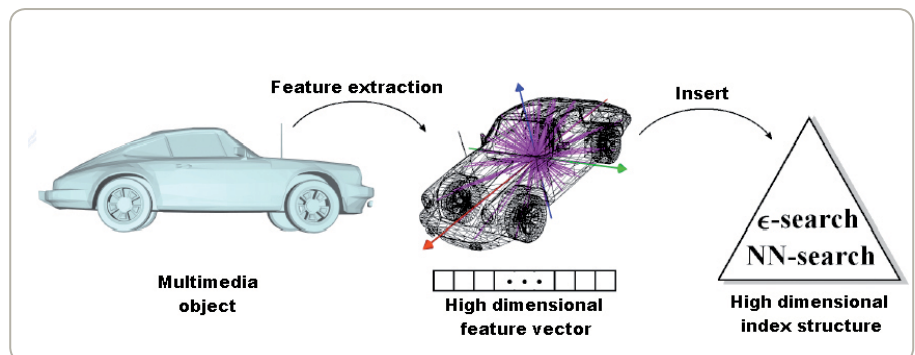
La eficiencia de la búsqueda por similitud corresponde al tiempo de procesamiento requerido para obtener la respuesta. Si uno requiere buscar en colecciones de datos masivas, es imprescindible considerar índices que permitan realizar las búsquedas en forma eficiente. Cuando la función de disimilitud que se utiliza es una métrica, es

Figura 2



Extracción de características en imágenes. El descriptor obtenido para una manzana debiera ser distinto al descriptor de la pera.

Figura 3



Proceso de extracción de características e indexamiento.

posible indexarla utilizando índices métricos [Zezula06], los cuales hacen uso de la propiedad de la desigualdad triangular para descartar objetos o grupos completos de objetos de la colección durante la búsqueda. Si bien los índices métricos pueden mejorar la eficiencia de la búsqueda, no siempre es fácil determinar cuál es la mejor técnica de indexamiento en cada aplicación particular y, peor aún, la colección de datos puede ser intrínsecamente “difícil de indexar”, problema conocido como la “maldición de la dimensión”.

Por otra parte, la eficacia de un sistema de búsqueda en colecciones multimedia se define como su capacidad para encontrar objetos relevantes a la consulta especificada y, al mismo tiempo, para descartar aquellos objetos que no sean relevantes. Definir qué es relevante y qué no lo es en forma genérica no tiene mucho sentido, ya que esto depende mucho del tipo de dato multimedia, del contexto en el que se realiza la búsqueda y la aplicación en la cual se utilizarán los resultados. Para datos y aplicaciones específicas se han creado

Los buscadores actuales en la Web de información multimedia, por ejemplo Google Images están basados en consultas textuales, es decir, el usuario debe ingresar palabras claves con la cuales buscar la información multimedia requerida (...) El problema es que los resultados que se pueden obtener a través de estas búsquedas textuales no siempre son satisfactorios, dada la complejidad de los datos multimedia.

coleccion de referencia, que se componen de un conjunto de objetos multimedia, de un conjunto de consultas y de un conjunto de respuestas a dichas consultas. Es decir, se conoce de antemano qué es lo que el sistema de búsqueda debería retornar para cada consulta definida en la colección. De esta forma se puede medir qué tan parecida es la respuesta del sistema de búsqueda con el ideal definido en la colección de referencia: mientras más parecido sea, más eficaz es el sistema de búsqueda.

Definir una colección de referencia puede ser un trabajo tedioso, ya que idealmente es una persona o un grupo de personas la que define el conjunto de consultas y sus respuestas (lo que implica tener que revisar la colección de datos “manualmente”). También se han utilizado las anotaciones asociadas a los objetos multimedia para clasificarlos y dividirlos en grupos. La ventaja de disponer de colecciones de referencia genéricas y disponibles para toda la comunidad es que permite comparar la eficacia de diferentes sistemas o métodos de búsquedas y así saber cuál retorna los resultados de mejor calidad.

## DESAFÍOS A FUTURO

Existen muchos problemas no resueltos en el área de búsqueda en colecciones multimedia, y por supuesto en toda la diversidad de áreas de investigación en Multimedia en general. A continuación detallo algunos de los problemas abiertos en la actualidad:

- **Gap semántico.** Un problema importante del área es que aún es difícil extraer contenido de alto nivel (información semántica) directamente del contenido multimedia (datos de bajo nivel). Por ejemplo, es posible que dos fotos sean muy parecidas si uno analiza exclusivamente la distribución de colores en ellas, pero podrían ser fotos cuyo significado sea completamente distinto. Una de las formas en las que se está abordando este problema es combinar la información de bajo nivel con metadatos o información generada por el usuario (tags, información de bitácoras de búsqueda, etc.), de forma de enriquecer la información multimedia y por ende mejorar la calidad de los resultados obtenidos.
- **Big data.** Las técnicas actuales de indexamiento de datos multimedia pueden funcionar bien en colecciones de millones de objetos, pero hay pocas que tengan escalabilidad a tamaños mayores. Esto es lo que se requeriría para poder buscar en la Web o para buscar en datos que están siendo generados constantemente (datos de sensores, por ejemplo). Esto es tema activo de investigación actual.
- **Heterogeneidad.** ¿Cómo combinar la información digital proveniente de distintas fuentes y que son de distinto tipo?, ¿cómo identificar si pertenecen al mismo contexto?, ¿cómo enriquecer la información digital a partir de sus metadatos? etc.

- **Sistemas distribuidos.** Este tema es de particular importancia, ya que podría ser la forma de resolver algunos de los problemas planteados en Big Data. Es muy interesante por ejemplo, el estudio de algoritmos de búsquedas en redes P2P, donde las técnicas tradicionales de indexamiento no se pueden utilizar directamente.
- **Espacios no-métricos.** Para la búsqueda por similitud, es interesante el estudio de funciones de distancia no-métricas para comparar objetos, ya que en ciertas aplicaciones dichas distancias pueden obtener resultados más eficaces que funciones de distancias métricas más “tradicionales”. Sin embargo, si no se cumplen las propiedades métricas, en particular la desigualdad triangular, no se pueden utilizar las técnicas de indexación tradicionales. Queda entonces el problema de cómo indexar en forma eficiente conjuntos de datos multimedia modelados como espacios no-métricos.

## AGRADECIMIENTOS

Mis agradecimientos a Tomas Skopal, profesor de la Charles University in Prague, por permitirme usar algunas de sus figuras de ejemplo en este artículo.<sup>BITS</sup>

## REFERENCIAS

[ACM12] ACM Multimedia 2012 Full/Short Paper Area Description. <http://www.acmmm12.org/fullshort-paper-area-description/>

[Cisco12] Cisco Visual Networking Index: Forecast and Methodology, 2011–2016. Mayo 2012. White paper. Disponible en: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-481360.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf)

[Quora11] <http://www.quora.com/How-many-photos-are-uploaded-to-Facebook-each-day>

[Zezula05] Zezula, P., Amato, G., Dohnal, V., and Batko, M. 2005. Similarity Search: The Metric Space Approach (Advances in Database Systems). Springer, Berlin, Germany.