

Recuperación de Información Multimedia

Jacqueline Fernández, Verónica Ludueña, Nora Reyes, Patricia Roggero

Departamento de Informática
Universidad Nacional de San Luis
{jmfer, vlud, nreyes, proggero}@unsl.edu.ar

Edgar Chávez
Escuela de Ciencias Físico–Matemáticas
Universidad Michoacana de San Nicolás de Hidalgo
elchavez@umich.mx

Resumen

Las principales dificultades al momento de realizar una búsqueda multimedia son, por un lado, la dificultad de los usuarios de especificar sus intereses en forma clara y como una consulta bien definida, y por otro lado, el problema de la extracción de las características relevantes de los objetos multimedia. Además las respuestas se ven afectadas, entre otras cosas, por la forma de representación y almacenamiento de los datos y por el costo de la transferencia de los mismos ya sea entre distintos dispositivos de almacenamiento en la jerarquía de memorias, o sobre una red. Dada una consulta al sistema, el objetivo clave de un sistema de recuperación de información es recuperar la información que podría ser útil o relevante para el usuario.

Por lo tanto, nuestra propuesta se enfoca en tratar de mejorar las herramientas de recuperación de información multimedia desarrollando nuevas técnicas capaces de soportar la interacción con el usuario, diseñando nuevas estructuras de datos (índices) capaces de manipular eficientemente datos multimedia y buscando representaciones que reflejen más adecuadamente las características de interés de los objetos multimedia.

Palabras Claves: *Recuperación de Información, Bases de Datos Multimedia, Búsqueda por Similitud*

Contexto

Esta línea de investigación se encuentra enmarcada dentro de un nuevo proyecto presentado en la Universidad Nacional de San Luis y en el Programa de Incentivos: “Nuevas Tecnologías para el Tratamiento Integral de Datos Multimedia”. El mismo será financiado por la Universidad Nacional de San Luis.

Además nuestras investigaciones se encuadran en el marco de un proyecto dentro del Programa de Promoción de la Universidad Argentina para el Fortalecimiento de Redes Interuniversitarias III en los que participa nuestra universidad junto con la Universidad de Zaragoza (España) y la Universidad Michoacana de San Nicolás de Hidalgo (México).

En este proyecto se propone avanzar en la integración de las investigaciones sobre adquisición, procesamiento y análisis de datos no estructurados y su aplicación en dominios no convencionales, concentrándose principalmente en los siguientes tipos de datos no estructurados: texto, sonido, imágenes y video. Se espera que el principal aporte de esta propuesta sea la incorporación de información no estructurada en los procesos de toma de decisiones y resolución de problemas que, normalmente, queda fuera de consideración en los enfoques clásicos.

Dentro de este contexto nuestra línea se dedica principalmente al diseño y desarrollo de índices que sirvan de apoyo a sistemas de recuperación dedicados a datos no estructurados tales como: datos multimedia, texto, etc. Se espera así contribuir a estos sistemas diseñando estructuras de datos más eficientes para memorias jerárquicas, gracias por ejemplo a la compactidad (el trabajar con estructuras de datos compactas) y/o con I/O eficiente. Se pretende principalmente analizar las estructuras de datos existentes, y eventualmente proponer nuevas, para manipular y recuperar algunos de los siguientes tipos de datos: secuencias, textos, video, imágenes, entre otros, estudiando los problemas desde ambos puntos de vista teórico y empírico.

1. Introducción y Motivación

Es común en nuestros días que los sistemas de computación hagan uso intensivo de información estructurada, es decir datos elementales o estructuras, generadas con un formato específico por un programa determinado. Una característica principal en estos casos, es que la estructura o formato de esta información puede ser fácilmente interpretada y directamente utilizada por un programa de computadora, una vez que la información es accedida desde estructuras clásicas de almacenamiento como matrices, registros de bases de datos, etc.

Si bien la información estructurada ha sido la principal materia prima utilizada en los sistemas computacionales hasta la fecha el hecho de restringirse al uso de este tipo de información conduce, muchas veces, a representar una visión parcial del problema y dejar fuera de consideración información que podría ser de gran importancia para la resolución efectiva del mismo. En este contexto, comienza a reconocerse hoy en día, que gran parte de la información que se requiere para la toma de decisiones y la resolución de problemas de índole general proviene de información no estructurada, principalmente aquella almacenada en forma de texto, audio, imagen y video.

Una manera de responder eficientemente consultas sobre una base de datos multimedia es usando métodos de acceso (índices) [11, 1, 9]. En general, en el ámbito de la recuperación de información, se usan índices debido al volumen de datos con el que se trabaja.

Una forma de implementar la búsqueda por similitud en bases de datos multimedia es usando información de anotaciones que describan el contenido de los objetos multimedia. Sin embargo, este modo de hacerlo no es práctico en grandes repositorios multimedia porque la mayoría de las descripciones textuales deben generarse a mano dado que son difíciles de obtener automáticamente. Además, en general, son subjetivas y en la mayoría de los casos no pueden caracterizar toda la información disponible del objeto multimedia. Más aún, otra restricción de este enfoque es que el procesamiento del índice debe “adivinar”, dejando fijado de antemano, cuáles clases de consultas se podrán hacer sobre los datos.

Un enfoque más prometedor para implementar sistemas de recuperación usando búsqueda por similitud es una búsqueda basada en contenidos, la cual usa el dato multimedia mismo. Para calcular la simi-

litud entre dos objetos multimedia, se debe definir una función de distancia. Dicha función mide la similitud, o más bien la disimilitud, entre dos objetos.

El concepto de búsqueda por similitud se puede definir a partir del concepto de espacios métricos, lo cual da un marco formal que es independiente del dominio de aplicación. Un espacio métrico está compuesto por un *universo* U de objetos y una función de distancia $d : U \times U \rightarrow R^+$, que satisface las propiedades que hacen de ella una métrica.

Las consultas por similitud, sobre una base de datos $S \subseteq U$, son usualmente de dos tipos:

Búsqueda por rango: recuperar todos los elementos de S a distancia r de un elemento q dado.

Búsqueda de los k vecinos más cercanos: dado q , recuperar los k elementos más cercanos a q .

Si la base de datos S posee n objetos, las consultas pueden ser trivialmente respondidas llevando a cabo n evaluaciones de distancia. Sin embargo, la mayoría de las aplicaciones requieren distancias costosas de computar, por ejemplo, la comparación de huellas digitales, búsquedas por contenido en bases de datos multimedia, etc. Por lo tanto, la búsqueda secuencial no escala para problemas de tamaño medio o grande, que son los tamaños más habituales de las bases de datos multimedia. Así el objetivo es preprocesar la base de datos, construyendo un índice, de manera tal que las consultas puedan ser respondidas con la menor cantidad de cálculos de distancia.

Un caso particular de dato multimedia son las imágenes. Una imagen es un arreglo de píxeles que resultan de la convolución de una función señal y de una función de “rendering”. Dos imágenes son consideradas copia una de otra si tienen la misma señal, aunque tengan diferente función de “rendering”. La recuperación de imágenes basadas en su contenido es un aspecto fundamental de muchos problemas en visión por computadora, incluyendo reconocimiento de objetos o de escenas. Para ello se consideran características visuales de la imagen tal como el color, forma, textura y capa espacial para representar e indexar la imagen [3, 4, 5]. Además una imagen puede sufrir distorsiones geométricas como RST (rotación, escalado y traslación), “cropping”, perspectiva; distorsiones de iluminación (“light distortion”, “sunburnt distortion”), así como también distorsiones de calidad (“blurring” o “halftone scan-line”). En sistemas típicos los contenidos visuales de las imágenes son extraídos y descritos mediante vectores característicos multi-dimensionales.

Por lo tanto, nuestra propuesta se enfoca en tratar de mejorar las herramientas de recuperación desarrollando nuevas técnicas que soporten la interacción con el usuario, diseñando nuevas estructuras de datos (índices) capaces de manipular eficientemente datos multimedia y buscando representaciones que reflejen más adecuadamente las características de interés de los objetos multimedia.

2. Líneas de Investigación y Desarrollo

Como se ha mencionado previamente se planea investigar respecto de dos aspectos importantes para los sistemas de recuperación de información multimedia: diseñar nuevos índices capaces de manipular eficientemente datos multimedia y buscar representaciones que reflejen las características de interés de los objetos multimedia.

Diseño de Índices

Un catálogo importante de índices para espacios métricos aparece en [9, 1, 11]. La mayoría de los índices usan la desigualdad triangular para evitar el análisis secuencial de la base de datos. La distancia entre la consulta y los objetos de la base de datos puede ser estimada calculando algunas distancias de antemano hacia unos objetos distinguidos llamados *pivotes* y sin calcular las distancias reales desde el objeto de consulta a los objetos de la base de datos durante una búsqueda. Otra técnica común es indexar una partición del espacio en regiones denominadas *particiones compactas*.

Entre todas las técnicas para indexación en espacios métricos son de especial interés las estructuras de datos dinámicas, donde la base de datos no se conoce de antemano y además tanto los objetos como las consultas arriban al azar. En cambio, las estructuras estáticas se benefician desde el conocimiento de la base de datos, y de esta manera seleccionan los mejores puntos de referencia para una estructura de datos determinada.

En particular es de interés mejorar el desempeño de índices dinámicos jerárquicos (árboles), que es el caso de algunos de los índices para espacios métricos. Estos índices dinámicos, en general, se construyen incrementalmente vía inserciones. De tal manera, la raíz del árbol es el primer objeto que llega, y esto se repite recursivamente en cada nivel del árbol.

En esta línea se ha propuesto una técnica donde el “buffering” logra un buen compromiso entre una estructura estática construida con toda la información necesaria y una dinámica con conocimiento local de los datos. Entonces, en lugar de elegir al primer elemento como la raíz, se demora la selección hasta que hayan arribado suficientes elementos para estar en condiciones de realizar dicha selección, y de esta manera se toma una decisión en base a más información. Dado que las consultas arriban a un ritmo desconocido, para mantener el dinamismo es necesario contar con un índice que responda a las consultas con mejor desempeño que la técnica de fuerza bruta. La idea ha sido, entonces, dar una estructura propia al “buffer” de manera que fuera capaz de responder consultas. Es por ello que el índice del “buffer” debería ser rápido y eficiente. En este sentido, se han analizado dos elecciones posibles: la primera fue usar un índice del mismo tipo como estructura del “buffer”, reconstruyendo una vez que el “buffer” estuviera completo. La segunda alternativa fue usar otra estructura de datos, como AESA [10], donde se asume el conocimiento completo de las distancias entre los elementos del “buffer”. Para este propósito se han analizado diferentes estrategias de selección de la raíz.

Esta técnica provee un marco adecuado para diseñar estructuras de datos dinámicas estables. Por lo tanto, tener un “buffer” en todos los niveles de una estructura jerárquica debería ser útil cuando se diseñan estrategias de ruteo para guiar las búsquedas, lo cual aparece como un área promisorio de investigación.

Representación de Objetos Multimedia

Un enfoque más prometedor para implementar sistemas de recuperación usando búsqueda por similitud es realizar una búsqueda basada en contenidos, la cual usa el dato multimedia mismo. En este caso para calcular la similitud entre dos objetos multimedia, es necesario definir una función de distancia.

En muchos casos para modelar la similitud de objetos multimedia se transforman los objetos en puntos de un espacio vectorial, el cual es un tipo particular de espacio métrico. Luego de definir ciertas características de interés para los objetos, se extraen valores numéricos para ellas desde los objetos multimedia y se construye un vector de características o descriptor, generalmente de alta dimensionalidad. Sobre espacios vectoriales se han definido numerosas funciones de distancia; por ejemplo: la distancia

Euclidiana.

Ambos enfoques de la búsqueda por similitud tienen sus ventajas y desventajas. En el caso de los espacios métricos, la función de similitud (la distancia) normalmente mide el mínimo esfuerzo (costo) necesario para transformar un objeto en otro (como en el caso de la distancia de edición entre cadenas de símbolos, que mide el número mínimo de inserciones, eliminaciones o reemplazo de caracteres necesario para transformar una cadena en otra). Por lo tanto, es necesario resolver un problema de optimización. Aunque ésta es una manera formal de definir la similitud entre objetos, dependiendo de los tipos de datos multimedia reales la función de similitud puede ser muy compleja y puede no siempre satisfacer las propiedades de una métrica (lo cual es importante con vistas a la posterior recuperación a través de la indexación).

En el caso de espacios vectoriales existen muchas funciones métricas disponibles que son fáciles de calcular, y los vectores de características poseen propiedades geométricas que se pueden usar para mejorar la indexación del espacio. Sin embargo, no siempre está claro cuáles características deben extraerse para obtener una buena representación de los datos multimedia originales. Una solución práctica, que proviene del área de procesamiento de señales, recurre a la utilización de transformaciones numéricas, tal como las transformadas de Fourier o de Wavelet. Si los datos se pueden representar por medio de una señal, la transformación da los coeficientes de una función base (es decir, funciones de base sinusoidal en el caso de la transformada de Fourier) cuya combinación lineal produce la señal original. Algunos de los coeficientes obtenidos se usan para formar los vectores de características.

Dependerá realmente de la aplicación final cuál de ambos modelos deberían usarse. El enfoque más práctico, como ya se mencionó, es modelar los datos multimedia como un espacio vectorial. Sin embargo, el enfoque de espacios métricos puede ser también útil si la similitud satisface las propiedades de una métrica, debido a que una función de distancia definida por expertos para un tipo de datos multimedia específico puede modelar la similitud mejor que el enfoque de vectores de características.

Cuando los datos multimedia se han modelado como un espacio métrico o como un espacio vectorial, la búsqueda por similitud se reduce a una búsqueda de objetos o puntos cercanos en el espacio.

3. Resultados Obtenidos y Esperados

Se ha podido comprobar experimentalmente que las estrategias de “buffering” mejoran la performance de una estructura de datos dinámica [2]. Se seleccionó como caso de estudio el Árbol de Aproximación Espacial Dinámico [7] y se obtuvo una mejora sistemática en los costos de las consultas usando un “buffer” en el primer nivel del árbol.

En particular, se ha podido verificar que esta estructura es mejor que su versión estática [6], porque deja como “vecinos” de un nodo algunos objetos alejados. Es decir, que en algunos casos parece permitir avanzar en la exploración espacial con “pasos más grandes” al pasar a los vecinos de un nodo. Es por esto que ahora se pretende analizar el efecto de elegir como vecinos de un nodo una muestra de objetos cercanos y lejanos. Si se clasificaran los objetos por distancia a la raíz, usando la información del histograma de distancias a ella, se podría elegir con esa misma densidad a los vecinos: muchos donde el histograma sea denso y pocos donde el histograma sea raro. Se espera que mediante esta estrategia se mejore, aún más, el desempeño de este índice y que esto pueda extenderse a otros índices jerárquicos.

En este mismo sentido, se espera conseguir versiones de estos índices que permitan trabajar con grandes volúmenes de datos y por lo tanto sean diseñadas específicamente para trabajar en memoria secundaria.

Respecto a la representación de datos multimedia, en particular sobre imágenes, un problema abierto corresponde a la construcción de una distancia que permita identificar imágenes completas [8]. Es decir, el estado del arte ([3, 4, 5]) permite identificar puntos de interés que son semejantes; pero esto es sólo un paso para poder identificar imágenes completas que contienen muchos puntos de interés. Por lo tanto, para una imagen se desea llegar a determinar los puntos de interés de la misma, que sean invariantes a distorsiones. Entonces, el conjunto de estos puntos de interés para una imagen formarían su representación, donde los diferentes puntos de interés serían a su vez vectores. Luego, la recuperación basada en contenido considerará, como imágenes similares, a aquellas imágenes donde haya semejanza entre los grafos que se forman con los respectivos conjuntos de puntos de interés. Para ello uno de los objetivos consistirá en poder determinar una función de distancia entre los respectivos grafos de puntos de

interés.

4. Formación de Recursos Humanos

Considerando la importancia de formar profesionales que puedan contribuir al desarrollo de sistemas de recuperación de información multimedia, dentro de esta línea de investigación se están formando los siguientes docentes:

- Trabajo Final de Licenciatura en Ciencias de la Computación: un alumno está desarrollando su trabajo final sobre el diseño e implantación de un índice dinámico para búsquedas por similitud en espacios métricos especialmente diseñado para memoria secundaria. Este trabajo se está realizando gracias a una Beca Estímulo de la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis.
- Tesis de Maestría en Ciencias de la Computación: uno de los integrantes está desarrollando su tesis de Maestría sobre el tema de recuperación de imágenes.
- Tesis de Doctorado en Ciencias de la Computación: uno de los integrantes se encuentra definiendo su plan de doctorado sobre temas de diseño y optimización de índices para realizar búsquedas por similitud en espacios métricos.

5. Referencias

- [1] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, sep 2001.
- [2] E. Chávez, N. Reyes, and P. Roggero. Data buffering strategies to improve dynamic metric indexing. In *Proc. XXVIII International Conference of the Chilean Computer Science Society (SCCC'09)*. IEEE CS Press, 2009. To appear.
- [3] Haibin Ling and David W. Jacobs. Deformation invariant image matching. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1466–1473, Washington, DC, USA, 2005. IEEE Computer Society.
- [4] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [5] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *Int. J. Comput. Vision*, 60(1):63–86, 2004.
- [6] G. Navarro. Searching in metric spaces by spatial approximation. *The Very Large Databases Journal (VLDBJ)*, 11(1):28–46, 2002.
- [7] G. Navarro and N. Reyes. Dynamic spatial approximation trees. *Journal of Experimental Algorithmics*, 12:1–68, 2008.
- [8] A. Ocsa, A. Rodríguez, H. Chuctaya, and G. Humpire. Reconocimiento de rostros mediante puntos característicos locales. In *Anales del II Simposio Peruano de Computación Gráfica y Procesamiento de Imágenes*, pages 1–6, 2008.
- [9] H. Samet. *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.
- [10] E. Vidal. An algorithm for finding nearest neighbors in (approximately) constant average time. *Pattern Recognition Letters*, 4:145–157, 1986.
- [11] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach (Advances in Database Systems)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. XVIII, 220 p., Hardcover ISBN: 0-387-29146-6.