Tesis de Doctorado Recientes

Métodos de Acceso y Procesamiento de Consultas Espacio-Temporales

Alumno: Gilberto Gutiérrez.

Profesores Guía: Gonzalo Navarro (Universidad de Chile), Andrea Rodríguez (Universidad de Concepción).

Fecha de Examen: Abril de 2007.

Existe una necesidad creciente por contar con aplicaciones espacio-temporales que necesitan modelar la naturaleza dinámica de los objetos espaciales. Las bases de datos espacio-temporales intentan proporcionar facilidades que permitan apoyar la implementación de este tipo de aplicaciones. Una de estas facilidades corresponde a los métodos de acceso, que tienen por objetivo construir índices para permitir el procesamiento eficiente de las consultas espacio-temporales.

Gilberto Gutiérrez en la actualidad se desempeña como director y profesor del Departamento de Ciencias de la Computación y Tecnologías de la Información, Facultad de Ciencias Empresariales, Universidad del Bío-Bío, Chile

En esta tesis se describen nuevos métodos de acceso basados en un enfoque que combina dos visiones para modelar información espacio-temporal: snapshots y eventos. Los snapshots se implementan por medio de un índice espacial y los eventos que ocurren entre snapshots consecutivos, se registran en una bitácora. Se estudió el comportamiento de nuestro enfoque considerando diferentes granularidades del espacio. Nuestro primer método de acceso espacio-temporal (SEST-Index) se obtuvo teniendo en cuenta el espacio completo y el segundo (SESTL) considerando las divisiones más finas del espacio producidas por el índice espacial. En esta tesis se realizaron varios estudios comparativos entre nuestros métodos de acceso y otros métodos propuestos en la literatura (HR-tree y MVR-tree) para evaluar las consultas espacio-temporales tradicionales (time-slice y time-interval). Los estudios muestran la superioridad de nuestras estructuras de datos en términos de almacenamiento y eficiencia para procesar tales consultas en un amplio rango de situaciones. Para nuestros dos métodos de acceso se definieron modelos de costos que permiten estimar tanto el almacenamiento como el tiempo de las consultas. Estos modelos se validaron experimentalmente presentando una buena capacidad de estimación.

Basándonos en nuestros métodos propusimos algoritmos para procesar otros tipos de consultas espacio-temporales, más allá de time-slice y time-interval. Específicamente diseñamos algoritmos para evaluar la operación de reunión espacio-temporal, consultas sobre eventos y sobre patrones

espacio-temporales. Se realizaron varios experimentos con el propósito de comparar el desempeño de nuestros métodos frente a otros propuestos en la literatura (3D R-tree, MVR-tree, HR-tree y CellList) para procesar estos tipos de consultas. Los resultados muestran un rendimiento, en general, favorable a nuestros métodos. En resumen, nuestros métodos son los primeros que resuelven de manera eficiente no sólo las consultas de tipo time-slice y time-interval, sino también varias otras de interés en aplicaciones espacio-temporales.

Indexación Efectiva de Espacios Métricos usando Permutaciones

Alumno: Karina Figueroa.

Profesores Guía: Gonzalo Navarro (Universidad de Chile), Edgar Chávez (Universidad Michoacana de San Nicolás de Hidalgo Mexico).

Fecha de Examen: Junio de 2007.

En muchas aplicaciones multimedia y de reconocimiento de patrones es necesario hacer consultas por proximidad a grandes bases de datos modelándolas como un espacio métrico, donde los elementos son los objetos de la base de datos y la proximidad se mide usando una distancia, generalmente costosa de calcular. El objetivo de un índice es preprocesar la base de datos para responder consultas haciendo el menor número de evaluaciones de distancia.

Los índices métricos existentes hacen uso de la desigualdad triangular para responder consultas de proximidad, ya sea partiendo el espacio en regiones compactas o utilizando distancias precalculadas a un conjunto distinguido de elementos. En esta tesis presentamos una nueva manera de resolver el problema, representando los elementos como permutaciones. La permutación se obtiene eligiendo un conjunto de objetos, llamados permutantes, y considerando el orden relativo en el que se ven los permutantes desde cada elemento a indexar.

Nuestra contribución principal es el haber descubierto que la proximidad entre elementos se puede predecir con mucha precisión midiendo la distancia entre las permutaciones que representan esos elementos.

Karina Figueroa en la actualidad se desempeña como profesora e investigadora, y administradora de Sistemas del Centro de Cómputo, Facultad de Ciencias Físico Matemáticas, Universidad Michoacana (UMSNH), México.

Una aplicación directa de nuestra técnica deriva en un método probabilístico simple y eficiente: Se ordena la base de datos por proximidad de las permutaciones de los elementos a la permutación de la consulta, y se recorre en ese orden. De la comparación experimental de esta técnica contra el estado del arte, en diversos espacios reales y sintéticos, se concluye que las permutaciones son mucho mejores predictores de proximidad que las técnicas hasta ahora usadas, sobre todo en dimensiones altas. Generalmente basta revisar una pequeña fracción de la base de datos para tener un alto porcentaje de la respuesta correcta.

Otra aplicación menos directa de nuestra técnica consiste en modificar el algoritmo exacto AESA, que por 20 años ha sido el índice más eficiente, en términos de cálculos de distancia, para buscar en espacios métricos. Nuestra variante, iAESA, utiliza las permutaciones para determinar el siguiente candidato a compararse contra la consulta. Los resultados experimentales muestran que es posible mejorar el desempeño de AESA hasta en 35 %. Esta técnica es adaptable a otros algoritmos existentes.

Se aplicó nuestra técnica al problema de identificación de rostros en imágenes, y se lograron resultados hasta ahora no alcanzados por los típicos algoritmos vectoriales usados en estas aplicaciones. Asimismo, dado que nuestra técnica no aplica explícitamente la desigualdad triangular, la probamos en algunos espacios de similaridad no métrica, obteniendo un índice que permite la búsqueda por proximidad con resultados semejantes al caso de los espacios métricos.

Minería de Datos en Motores de Búsqueda

Alumno: Marcelo Mendoza.
Profesores Guía: Ricardo Baeza (Universidad de Chile).
Fecha de Examen: Junio de 2007.

La Web es un gran espacio de información donde muchos recursos como documentos, imágenes u otros contenidos multimediales pueden ser accesados. En este contexto, varias tecnologías de la información han sido desarrolladas para ayudar a los usuarios a satisfacer sus necesidades de búsqueda en la Web, y las más usadas de estas son los motores de búsqueda. Los motores de búsqueda permiten a los usuarios encontrar recursos formulando consultas y revisando una lista de respuestas.

Marcelo Mendoza en la actualidad se desempeña como director y profesor de la Carrera de Ingenieria Civil Informática, Universidad de Valparaíso, Chile.

Uno de los principales desafíos para la comunidad de la Web es diseñar motores de búsqueda que permitan a los usuarios encontrar recursos semánticamente conectados con sus consultas. El gran tamaño de la Web y la vaguedad de los términos más comúnmente usados en la formulación de consultas son los principales obstáculos para lograr este objetivo.

En esta tesis se exploraron las selecciones de los usuarios registradas en los logs de los motores de búsqueda para aprender cómo los usuarios buscan y también para diseñar algoritmos que permitieran mejorar la precisión de las respuestas recomendadas a los usuarios. Se comenzó explorando las propiedades de estos datos. Esta exploración nos permitió determinar la naturaleza dispersa de estos datos. Además se presentaron modelos que permitieron entender cómo los usuarios buscan en los motores de búsqueda.

Luego, se exploraron las selecciones de los usuarios para encontrar asociaciones útiles entre consultas registradas en los logs. Los esfuerzos se concentraron en el diseño de técnicas que permitieran a los usuarios encontrar mejores consultas que la consulta original. Como una aplicación, se diseñaron métodos de reformulación de consultas que ayudaran a los usuarios a encontrar términos más útiles mejorando la representación de sus necesidades.

Usando términos de documentos se construyeron representaciones vectoriales para consultas. Aplicando técnicas de clustering se pudieron determinar grupos de consultas similares. Usando estos grupos de consultas, se introdujeron métodos para recomendación de consultas y documentos que permitieron mejorar la precisión de las recomendaciones.

Finalmente, se diseñaron técnicas de clasificación de consultas que nos permitieron encontrar conceptos semánticamente relacionados con la consulta original. Para lograr esto, se clasificaron las consultas de los usuarios en directorios Web. Como una aplicación, se introdujeron métodos para la mantención automática de los directorios.

Improving Learning-Object Metadata Usage During Lesson Authoring

Alumno: Olivier Motelet.
Profesor Guía: Nelson Baloian; José A. Pino (Universidad de Chile).
Fecha de Examen: Octubre 2007.

Para lograr coherencia y flexibilidad en unidades de aprendizaje basadas en documentos multimedia, varios autores han recomendado estructurar los componentes de los cursos en grafos. En un grafo de curso, los recursos educacionales son encapsulados como objetos de aprendizaje (LO - Learning Objects) con sus respectivos metadatos (LOM - Learning-Object Metadata) y son interconectados con relaciones de varios tipos retóricos y/o semánticos. Los grafos de recursos son almacenados en repositorios en los cuales los metadatos sirven para facilitar su recuperación y reutilización. Sin embargo, tales sistemas se enfrentan con problemas serios en cuanto al uso de los LOMs: los metadatos son difíciles de instanciar y los autores de cursos generalmente no tienen estímulos para cumplir con esta tediosa tarea ya que ellos mismos no se benefician de los metadatos que generan.

Olivier Motelet se desempeña como responsable de Indentificación de Comercialización Proyecto Innovador para Desarrolladores, Empresa IDTM, Francia.

La generación automática de metadatos resuelve este problema. Sin embargo, este método se limita a ciertos metadatos excluyendo la mayor parte de los metadatos subjetivos tales como los metadatos educacionales. Esta limitación motivó el enfoque de esta tesis sobre una técnica complementaria: un método híbrido basado en la sinergia entre procesos automáticos e intervención

humana. La generación híbrida de LOMs puede ser aplicada sobre los atributos que no pueden ser automáticamente generados. Sin embargo, este enfoque está basado en la contribución de usuarios no siempre cooperativos, quienes necesitarían ver beneficios para motivar su participación.

Proponemos estudiar los usos de LOM durante la creación de cursos, no sólo desde la perspectiva de la generación híbrida sino también desde la perspectiva de los beneficios que pueden brindar los LOMs. Esta estrategia tiene como objetivo soportar una retroacción positiva en la cual los beneficios puedan motivar la generación de LOMs de buena calidad, y la buena calidad de los LOMs pueda mejorar los beneficios.

En particular, esta tesis investiga métodos para (1) integrar sin transición la generación híbrida de LOMs dentro de una herramienta de creación de cursos, (2) procesar un conjunto de LOMs aunque ciertos metadatos quedaran incompletos, incorrectos, o faltantes, (3) mejorar los resultados de los métodos clásicos de recuperación de LOs usando los metadatos de los LOs que componen un curso.

Desarrollamos una herramienta de código abierto para validar las propuestas de esta tesis. Experimentos preliminares mostraron que los LOMs pueden mejorar significativamente la recuperación de LOs adicionales durante el proceso de creación de cursos.