

La Matemática del Conocimiento.

Joos Heintz

La época del internet, de las grandes bases de datos y de la comunicación universal nos imponen una reflexión sobre el significado del conocimiento y su papel en la sociedad.

Hoy en día la sociedad mundial dispone sobre un históricamente nunca alcanzado caudal de información (“conocimiento”), que se encuentra almacenada en bases de datos (bibliotecas, archivos, ficheros de investigación, historias clínicas etc.) dispersados sobre todo el globo terráqueo. Es técnicamente posible acceder a esta información desde cualquier ordenador.

La filosofía nos conduce al intento de una definición del conocimiento por su contenido, una tarea difícil con soluciones que suelen obviar aspectos importantes (por ejemplo, la cuestión de la representación del conocimiento). Por otra parte, la búsqueda de la perfecta definición del conocimiento no es una necesidad fundamental para la sociedad, pero sí las cuestiones de flujo y del acceso a la información. Justamente este aspecto del conocimiento se brinda a la modelación matemática, lo que a su vez nos permite estructurar y organizar mejor el conocimiento, aumentar su accesibilidad y, también,.....darnos cuenta de los límites de tales mejoras.

Las bases de datos usuales son de tipo relacional y almacenan información discreta. Típicos ejemplos son bibliotecas, archivos y también imágenes digitalizados. Existe una teoría informática desarrollada de bases de datos relacionales, en gran medida basada en la lógica de las estructuras combinatorias (Finite Model Theory). Otro tipo de bases de datos de gran difusión en el mundo real está representado por las bases de datos continuas (Constraint Databases - CDB). Como dice su nombre, estas bases de datos almacenan información continua (representada por vectores fijos o dinámicos constituidos por valores reales “arbitrarios”). Típicos ejemplos están representados por análisis clínicos, tablas meteorológicas o geológicas o mapas con información catastral (por ejemplo, en el contexto de Geographic Information Systems – GIS).

Una base de datos (relacional o continua) almacena información implícita que se transforma en explícita mediante consultas en un lenguaje previamente fijado. Una consulta puede ser de tipo booleano (respondiendo con un si o un no o con un nombre) o puede reducir una base de datos que contiene información redundante en una más sucinta que permite una evaluación más eficiente de preguntas.

A pesar de que las bases de datos continuas tienen amplia difusión en la vida práctica (por ejemplo, en medicina), su teoría informática se encuentra insuficientemente desarrollada. Este hecho se debe a la estrecha relación que existe entre la complejidad de la evaluación de una consulta con las tareas tradicionales de la eliminación geométrica (resolución efectiva y eficiente de sistemas de ecuaciones y desigualdades polinomiales en varias variables, proyección y reconstrucción de figuras geométricas dadas de forma implícita). La dificultad intrínseca (en términos de complejidad computacional) de la eliminación geométrica impidió hasta ahora el desarrollo de software eficiente para consultas en bases de datos continuas. En este contexto hay que

observar que la simple optimización de los algoritmos con respecto a las estructuras de datos tradicionales (representación de los enteros por dígitos, de polinomios por sus coeficientes y de sistemas de ecuaciones y desigualdades por formulas) no da resultados satisfactorios en la práctica y que la tarea de la evaluación eficiente de consultas en bases de datos continuas requiere la optimización simultanea de los tipos y estructuras de datos que representan los objetos matemáticos involucrados.

Un candidato promisorio para una estructura de datos eficiente consiste en la representación de enteros, polinomios y formulas por circuitos aritméticos. Como consecuencia de este cambio de la estructura de datos se obtiene una mejora exponencial de la complejidad de los algoritmos que evalúan consultas en bases de datos continuas. Utilizando parámetros *sintácticos* y *semánticos* (geométricos) adecuados para medir el tamaño de la entrada, es posible afirmar que los nuevos algoritmos evalúan las consultas en tiempo polinomial. Sin embargo, si el tamaño de la entrada se mide *exclusivamente* en términos *sintácticos*, la evaluación de consultas preserva el carácter intrínsecamente exponencial de los algoritmos de eliminación geométrica, bajo mínimos requisitos de uniformidad (branching parsimonious procedures).

Finalmente observamos que el lenguaje de la (todavía rudimentaria) teoría de las bases de datos continuas permite la especificación descriptiva de las tareas que resuelven los conocidos algoritmos de eliminación geométrica.