

## **Web Crawling**

Carlos Castillo (Universidad de Chile)

El recorrido de la Web debe ser realizado de una manera eficiente y escalable, debido a que el ancho de banda disponible no es ni infinito ni gratis, y el número de páginas Web es esencialmente no-acotado. Discutiremos métodos de ordenamiento para el corto- y largo-plazo al recorrer la Web, así como asuntos de arquitectura e implementación e incluso asuntos de índole práctica que surgen al realizar recorridos sobre muchas páginas. Esta charla presenta los problemas de un Web crawler en muchos niveles diferentes, desde eficiencia de red hasta los desafíos algorítmicos de priorizar páginas en un grafo que es sólo parcialmente conocido.