Curso de Datos Masivos I

Licenciatura en Ciencia de Datos, IIMAS, UNAM

Tablas y funciones *hash*

Fecha límite: 16 de marzo.

Formato: Libreta Jupyter.

Forma de entrega: Subir en Google Classroom.

Ejercicio 1

Programa una tabla *hash* que realice sondeo cuadrático. Elige un número primo grande para m y $c_1 = c_2 = 1$ y prueba la tabla almacenando un conjunto de 200 números.

$$h(x, i) = (h'(x) + c_1 \cdot i + c_2 \cdot i^2) \quad \text{mód } m$$

Ejercicio 2

Desarrolla un programa que cuente el número de ocurrencias de todas las subcadenas de longitud entre 5 a 10 (sin considerar espacios en blanco ni signos de admiración/interrogación ni caracteres especiales) usando funciones y tablas *hash*. Prueba este programa en el conjunto de documentos de 20 newsgroups.