

Conjuntos y Tablas Hash

Programación de Estructuras de Datos y Algoritmos Fundamentales
(TC1031)

M.C. Xavier Sánchez Díaz
sax@tec.mx



Outline

- 1 Revisión de Conjuntos
- 2 Operaciones con conjuntos
 - Conjuntos dinámicos y funciones
- 3 Hash Tables

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Un **conjunto** es un concepto abstracto, construido para referirse a una **colección de elementos**.

Usualmente representamos los **conjuntos** con letras mayúsculas (usualmente usando letras próximas a la A), y delimitamos sus contenidos con llaves (*curly brackets*):

Ejemplo

A es el conjunto de los primeros cinco **números naturales**, es decir aquellos que *nos sirven para contar*:

$$A = \{1, 2, 3, 4, 5\}$$

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Un **conjunto** es un concepto abstracto, construido para referirse a una **colección de elementos**.

Usualmente representamos los **conjuntos** con letras mayúsculas (usualmente usando letras próximas a la A), y delimitamos sus contenidos con llaves (*curly brackets*):

Ejemplo

A es el conjunto de los primeros cinco **números naturales**, es decir aquellos que *nos sirven para contar*:

$$A = \{1, 2, 3, 4, 5\}$$

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Un **conjunto** puede **enumerarse** o **describirse**:

Enumeración o Extensión

$$A = \{1, 2, 3, 4, 5\}$$

Descripción o Comprensión

- A = el conjunto de los primeros cinco números naturales
- B = el conjunto de personas en este *ZOOM Room*
- C = el conjunto de estudiantes del Campus Monterrey

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Un **conjunto** es una colección en la que **no existe orden alguno**:

Ejemplo

Si $A = \{1, 2, 3, 4, 5\}$ y $B = \{2, 3, 1, 5, 4\} \dots$

- ¿Cuál de los dos es el conjunto de los cinco primeros números naturales?
- ¿Cuáles son los elementos del primer conjunto y cuáles son los del segundo?

Podemos usar el símbolo \in para denotar *pertenencia*, e.g. $2 \in A$ significa que el 2 es un elemento *que pertenece* a A o *que está* en A .

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Podemos **contar los elementos** que hay dentro de un conjunto. A la **cantidad de elementos** dentro de un conjunto le llamamos **cardinalidad**.

Ejemplo

Si $A = \{1, 2, 3, 4, 5\} \dots$

- Q: ¿Cuál es la cardinalidad de A ?
- A: 5

Nota

Aunque es poco común, a veces pueden observarse conjuntos con elementos repetidos. Si este fuera el caso, asume que sólo existe una copia de cada elemento.

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Podemos **contar los elementos** que hay dentro de un conjunto. A la **cantidad de elementos** dentro de un conjunto le llamamos **cardinalidad**.

Ejemplo

Si $A = \{1, 2, 3, 4, 5\} \dots$

- Q: ¿Cuál es la cardinalidad de A ?
- A: 5

Nota

Aunque es poco común, a veces pueden observarse conjuntos con elementos repetidos. Si este fuera el caso, asume que sólo existe una copia de cada elemento.

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Podemos **contar los elementos** que hay dentro de un conjunto. A la **cantidad de elementos** dentro de un conjunto le llamamos **cardinalidad**.

Ejemplo

Si $A = \{1, 2, 3, 4, 5\} \dots$

- Q: ¿Cuál es la cardinalidad de A ?
- A: 5

Nota

Aunque es poco común, a veces pueden observarse conjuntos con elementos repetidos. Si este fuera el caso, asume que sólo existe una copia de cada elemento.

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Para denotar la **cardinalidad** de un conjunto *contable*, usualmente usamos el símbolo $\#(A)$, mientras que usamos dos barras verticales para denotar la cardinalidad de un conjunto no contable.

Ejemplo

Si $A = \{1, 2, 3, 4, 5\}$ entonces $\#(A) = 5$ o bien $|A| = 5$

Algunos autores usan una notación; otros, otra. No importa cuál usemos, intentemos ser consistentes.

¿Qué es un conjunto?

Definición y propiedades de los conjuntos

Para denotar la **cardinalidad** de un conjunto *contable*, usualmente usamos el símbolo $\#(A)$, mientras que usamos dos barras verticales para denotar la cardinalidad de un conjunto no contable.

Ejemplo

Si $A = \{1, 2, 3, 4, 5\}$ entonces $\#(A) = 5$ o bien $|A| = 5$

Algunos autores usan una notación; otros, otra. No importa cuál usemos, intentemos ser consistentes.

Inclusión

Operaciones con conjuntos

Podemos **comparar** dos conjuntos en cuanto a tamaño, pero también podemos saber si uno está **incluido** dentro de otro.

Ejemplo

Si A = el conjunto de habitantes de Nuevo León y B = es el conjunto de habitantes de Monterrey, entonces sabemos que B es un **subconjunto** de A .

Usamos la notación $B \subseteq A$ para decir que B es un **subconjunto** de A ; *cada elemento de B está en A ...*

PERO ESPERA

Inclusión

Operaciones con conjuntos

Podemos **comparar** dos conjuntos en cuanto a tamaño, pero también podemos saber si uno está **incluido** dentro de otro.

Ejemplo

Si A = el conjunto de habitantes de Nuevo León y B = es el conjunto de habitantes de Monterrey, entonces sabemos que B es un **subconjunto** de A .

Usamos la notación $B \subseteq A$ para decir que B es un **subconjunto** de A ; *cada elemento de B está en A ...*

PERO ESPERA

Inclusión

Operaciones con conjuntos

Podemos **comparar** dos conjuntos en cuanto a tamaño, pero también podemos saber si uno está **incluido** dentro de otro.

Ejemplo

Si A = el conjunto de habitantes de Nuevo León y B = es el conjunto de habitantes de Monterrey, entonces sabemos que B es un **subconjunto** de A .

Usamos la notación $B \subseteq A$ para decir que B es un **subconjunto** de A ; *cada elemento de B está en A ...*

PERO ESPERA

Inclusión

Operaciones con conjuntos

Podemos **comparar** dos conjuntos en cuanto a tamaño, pero también podemos saber si uno está **incluido** dentro de otro.

Ejemplo

Si A = el conjunto de habitantes de Nuevo León y B = es el conjunto de habitantes de Monterrey, entonces sabemos que B es un **subconjunto** de A .

Usamos la notación $B \subseteq A$ para decir que B es un **subconjunto** de A ; *cada elemento de B está en A ...*

PERO ESPERA

Inclusión

Operaciones con conjuntos

Si A = el conjunto de habitantes de Nuevo León y B = es el conjunto de habitantes de Monterrey, entonces sabemos que B es un **subconjunto propio** de A :

Inclusión propia

- Si **todos** los elementos de B están en A , sabemos que $B \subseteq A$.
- Si **todos** los elementos de B están en A , pero no todos los elementos de A están en B , entonces $B \subset A$

A esto último se le llama inclusión propia (que es el caso de los de Monterrey y los de Nuevo León), y da *más información* que la simple inclusión.

Inclusión

Operaciones con conjuntos

Si A = el conjunto de habitantes de Nuevo León y B = es el conjunto de habitantes de Monterrey, entonces sabemos que B es un **subconjunto propio** de A :

Inclusión propia

- Si **todos** los elementos de B están en A , sabemos que $B \subseteq A$.
- Si **todos** los elementos de B están en A , pero no todos los elementos de A están en B , entonces $B \subset A$

A esto último se le llama inclusión propia (que es el caso de los de Monterrey y los de Nuevo León), y da *más información* que la simple inclusión.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

Identidad

Operaciones con conjuntos

¿Qué información tenemos en cada caso? Reflexiona un momento. . .

- $A \subseteq B$
- $B \subseteq A$
- $A \subseteq B$ y $B \subseteq A$

Cuando dos conjuntos tienen lo mismo, decimos que son **idénticos** (duh).

Observa la diferencia entre $A \subseteq B$ y $A \subset B$.

¿Cómo implementar un conjunto?

Conjuntos dinámicos

Necesitamos un **contenedor** que guarde elementos que no están ordenados y en donde cada elemento es único.

Si usamos un arreglo, lista o vector...

- ¿Cuántas operaciones toma buscar un elemento?
- ¿Cuántas operaciones toma acceder a un elemento?

¿Cómo cambia si lo guardamos en un árbol? ¿Y un grafo?

¿Cómo implementar un conjunto?

Conjuntos dinámicos

Necesitamos un **contenedor** que guarde elementos que no están ordenados y en donde cada elemento es único.

Si usamos un arreglo, lista o vector...

- ¿Cuántas operaciones toma buscar un elemento?
- ¿Cuántas operaciones toma acceder a un elemento?

¿Cómo cambia si lo guardamos en un árbol? ¿Y un grafo?

¿Cómo implementar un conjunto?

Conjuntos dinámicos

Necesitamos un **contenedor** que guarde elementos que no están ordenados y en donde cada elemento es único.

Si usamos un arreglo, lista o vector...

- ¿Cuántas operaciones toma buscar un elemento?
- ¿Cuántas operaciones toma acceder a un elemento?

¿Cómo cambia si lo guardamos en un árbol? ¿Y un grafo?

¿Cómo implementar un conjunto?

Conjuntos dinámicos

Necesitamos un **contenedor** que guarde elementos que no están ordenados y en donde cada elemento es único.

Si usamos un arreglo, lista o vector...

- ¿Cuántas operaciones toma buscar un elemento?
- ¿Cuántas operaciones toma acceder a un elemento?

¿Cómo cambia si lo guardamos en un árbol? ¿Y un grafo?

¿Cómo implementar un conjunto?

Conjuntos dinámicos

Necesitamos un **contenedor** que guarde elementos que no están ordenados y en donde cada elemento es único.

Si usamos un arreglo, lista o vector...

- ¿Cuántas operaciones toma buscar un elemento?
- ¿Cuántas operaciones toma acceder a un elemento?

¿Cómo cambia si lo guardamos en un árbol? ¿Y un grafo?

¿Cómo implementar un conjunto?

Conjuntos dinámicos

Necesitamos un **contenedor** que guarde elementos que no están ordenados y en donde cada elemento es único.

Si usamos un arreglo, lista o vector...

- ¿Cuántas operaciones toma buscar un elemento?
- ¿Cuántas operaciones toma acceder a un elemento?

¿Cómo cambia si lo guardamos en un árbol? ¿Y un grafo?

Funciones

Conjuntos dinámicos

Definición de función o *mapeo*

Una **función** f de A en B o bien $f: A \rightarrow B$ es una relación de la forma aRb o bien (a, b) , que existe entre los conjuntos A y B , tal que $a \in A$ y $b \in B$, y considerando que:

- Para **cada** $a \in A$ existe **siempre** un $b \in B$
- Para **cada** $a \in A$ existe un **único** $b \in B$
- Para **cada** $a \in A$ existe siempre el mismo $b \in B$

Es decir que es una caja mágica que por cada ingrediente a siempre devuelve resultados: un único resultado b que siempre será el mismo por cada ingrediente a con el que se trabaje.

Funciones

Conjuntos dinámicos

Definición de función o *mapeo*

Una **función** f de A en B o bien $f: A \rightarrow B$ es una relación de la forma aRb o bien (a, b) , que existe entre los conjuntos A y B , tal que $a \in A$ y $b \in B$, y considerando que:

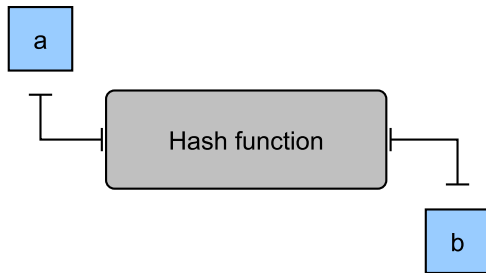
- Para **cada** $a \in A$ existe **siempre** un $b \in B$
- Para **cada** $a \in A$ existe un **único** $b \in B$
- Para **cada** $a \in A$ existe siempre el mismo $b \in B$

Es decir que es una caja mágica que por cada ingrediente a siempre devuelve resultados: un único resultado b que siempre será el mismo por cada ingrediente a con el que se trabaje.

Hash *function*

Hash Tables

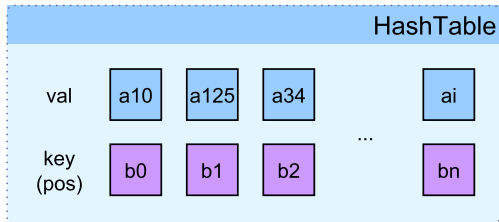
Ésa es la idea con una **hash table**: una *tabla* que guarda cada (a, b) de una **función** donde b es la **posición** (o *bucket*) y a es la **llave** (o valor) del elemento a guardar:



Al elemento a que queremos guardar se le aplica la función *hash* que hayamos elegido, y generará una posición que indica **dónde** debe guardarse.

Representación visual

Hash Tables



- Al elemento a_{10} se le aplicó una función que dio como resultado b_0 por lo que su lugar es el b_0 .
- Al elemento a_{125} se le aplicó la misma función y dio como resultado b_1 así que ése es el lugar que le corresponde. . .
- . . .y así para todos los elementos que queramos guardar.

Simple enough. . .

... Or is it?

Conflictos en Hash Tables

Considera la función hash $h(d_1d_2d_3d_4d_5d_6d_7d_8) = d_5d_6$ que toma los dígitos quinto y sexto del valor que queremos guardar para generar una posición.

¿Cuántos *buckets* distintos podemos tener usando esta función hash?

- ¿Qué posición le toca al valor 12345678?
- ¿Qué posición le corresponde al valor 24535655?
- ¿Cuál le corresponde al 16895674?

¡A los tres elementos les toca la misma posición!

... Or is it?

Conflictos en Hash Tables

Considera la función hash $h(d_1d_2d_3d_4d_5d_6d_7d_8) = d_5d_6$ que toma los dígitos quinto y sexto del valor que queremos guardar para generar una posición.

¿Cuántos *buckets* distintos podemos tener usando esta función hash?

- ¿Qué posición le toca al valor 12345678?
- ¿Qué posición le corresponde al valor 24535655?
- ¿Cuál le corresponde al 16895674?

¡A los tres elementos les toca la misma posición!

... Or is it?

Conflictos en Hash Tables

Considera la función hash $h(d_1d_2d_3d_4d_5d_6d_7d_8) = d_5d_6$ que toma los dígitos quinto y sexto del valor que queremos guardar para generar una posición.

¿Cuántos *buckets* distintos podemos tener usando esta función hash?

- ¿Qué posición le toca al valor 12345678?
- ¿Qué posición le corresponde al valor 24535655?
- ¿Cuál le corresponde al 16895674?

¡A los tres elementos les toca la misma posición!

... Or is it?

Conflictos en Hash Tables

Considera la función hash $h(d_1d_2d_3d_4d_5d_6d_7d_8) = d_5d_6$ que toma los dígitos quinto y sexto del valor que queremos guardar para generar una posición.

¿Cuántos *buckets* distintos podemos tener usando esta función hash?

- ¿Qué posición le toca al valor 12345678?
- ¿Qué posición le corresponde al valor 24535655?
- ¿Cuál le corresponde al 16895674?

¡A los tres elementos les toca la misma posición!

... Or is it?

Conflictos en Hash Tables

Considera la función hash $h(d_1d_2d_3d_4d_5d_6d_7d_8) = d_5d_6$ que toma los dígitos quinto y sexto del valor que queremos guardar para generar una posición.

¿Cuántos *buckets* distintos podemos tener usando esta función hash?

- ¿Qué posición le toca al valor 12345678?
- ¿Qué posición le corresponde al valor 24535655?
- ¿Cuál le corresponde al 16895674?

¡A los tres elementos les toca la misma posición!

Resolución de conflictos

Conflictos en Hash Tables

Hay dos opciones:

- **Open addressing** o *Rehashing* donde la estrategia es la **re-ubicación**:
 - ▶ *Linear probing*: $h(k) + i$
 - ▶ *Quadratic probing*: $h(k) + i^2$
 - ▶ *Double hashing*: $h_1(k) + i \cdot h_2(k)$
- **Chaining** donde la estrategia es el **encadenamiento**:
 - ▶ Externo: el *bucket* apunta a una lista que buscamos secuencialmente para encontrar lo que buscamos
 - ▶ Interno: el *bucket* apunta directamente a otra casilla en un espacio extra designado para eso dentro de la misma tabla

h es la función de hashing, k es la llave, e i es un número natural

¿Qué función debo usar?
¿Qué método de resolución de conflictos debo utilizar?

Depende¹

¿Y los ejemplos?

- Revisar las slides originales del curso
- Revisar el excelente material de Calvin Newton de Georgetown University

¹La tarea especifica qué usar. Para todo lo demás usen `std::map` por favor