

Memoria dinámica

Algoritmos y Estructuras de Datos II - 2023C1

Memoria dinámica: motivación

Queremos implementar una versión simplificada de `std::vector`¹:

```
template<class T>
class Vec<T> {
public:
    Vec();
    int size() const;
    T get(int i) const;
    void set(int i, T x);
    void push_back(T x);
private:
    ...
};
```

¹...sin usar `std::vector`.

Repaso: arreglos estáticos

Arreglos estáticos

C++ soporta nativamente arreglos estáticos, cuyo tamaño está fijo en tiempo de compilación:

```
int main() {  
    int arreglo_estatico[10];  
    for (int i = 0; i < 10; i++) {  
        arreglo_estatico[i] = i * i;  
    }  
    for (int i = 0; i < 10; i++) {  
        cout << arreglo_estatico[i] << endl;  
    }  
}
```

Memoria dinámica: motivación

¿Qué representación elegimos?

- ▶ No alcanza con un arreglo estático.

Modelo de memoria

En C++ la memoria es un arreglo de *bytes*.

Un byte es un entero de 8 bits (0..255).

Cada byte de la memoria tiene una única *dirección*.

Representación de variables locales

```
int main() {  
    int foo = 123;  
    int bar = 1000000;  
    char baz = 'A';  
    ...  
}
```

	Dirección	Byte

foo:0	9000	123
foo:1	9001	0
foo:2	9002	0
foo:3	9003	0
bar:0	9004	64
bar:1	9005	66
bar:2	9006	15
bar:3	9007	0
baz	9008	65

Modelo de memoria

Representación de estructuras

```
struct Par {
```

```
    int x;
```

```
    char y;
```

```
};
```

```
int main() {
```

```
    Par pares[2];
```

```
    pares[0].x = 10;
```

```
    pares[0].y = 'A';
```

```
    pares[1].x = 20;
```

```
    pares[1].y = 'B';
```

```
    ...
```

```
}
```

	Dirección	Byte

pares[0].x:0	9000	10
pares[0].x:1	9001	0
pares[0].x:2	9002	0
pares[0].x:3	9003	0
pares[0].y	9004	65
pares[1].x:0	9005	20
pares[1].x:1	9006	0
pares[1].x:2	9007	0
pares[1].x:3	9008	0
pares[1].y	9009	66

Nota: los detalles de representación pueden variar dependiendo de la arquitectura y del compilador.

Punteros

El tipo T^* es el tipo de los **punteros a T**.

Un puntero a T representa una dirección de memoria en la que (presumiblemente) hay almacenado un valor de tipo T.

- ▶ `int*`: puntero a int
- ▶ `char*`: puntero a char
- ▶ `vector<int>*`: puntero a vector de int
- ▶ `vector<int*>`: vector de punteros a int
- ▶ `int**`: puntero a puntero, por ejemplo, para tener una lista dinámica de listas
- ▶ ...

Punteros

Operaciones con punteros

- ▶ Dirección de memoria de una variable. (&variable)
 si `variable` es de tipo T
 &variable es de tipo T*
- ▶ Valor almacenado en una dirección de memoria (operador de **desreferencia**). (*puntero)
 si `puntero` es de tipo T*
 *`puntero` es de tipo T

Punteros

Punteros a variables locales

```
int main() {  
    int x = 10;  
    int* px = &x;  
    cout << px << endl;  
    cout << *px << endl;  
    *px = *px + 1;  
    cout << x << endl;  
  
    int* q = &7;  
}
```

Punteros

Punteros a variables locales

```
int main() {  
    int x = 10;  
    int* px = &x;  
    cout << px << endl;  
    cout << *px << endl;  
    *px = *px + 1;  
    cout << x << endl;
```

```
    int* q = &7;  
}
```

co.cpp:10:13: error: lvalue required as unary '&' operand

```
    int* q = &7;
```

Punteros

Punteros a estructuras

```
struct Par {  
    int x;  
    char y;  
};  
  
int main() {  
    Par pares[2];  
    Par* p = &pares[1];  
    (*p).x = 10;  
    p->y = 'b';  
    cout << p->x << endl;  
    char* q = &p->y;  
    *q = 'c';  
    cout << pares[1].y << endl;  
}
```

Puntero `nullptr`

La palabra clave `NULL`

Históricamente, la dirección de memoria 0 está reservada para representar un puntero que no referencia ninguna posición de memoria. En C++ se puede escribir `NULL` para denotar dicho puntero.

La palabra clave `nullptr`

Desde nuestro punto de vista `NULL` y `nullptr` son sinónimos. Hay algunas diferencias técnicas y `nullptr` se puede considerar más prolijo.

Puntero nullptr

```
int* p = nullptr;

if (p != nullptr) {
    cout << *p << endl;
} {
    cout << "p no esta disponible en este momento";
}
```

Regiones de memoria

La memoria en C++ se divide en tres tipos/regiones:

Global (estática) \Rightarrow en el ejecutable

La memoria estática se encuentra incrustada en el ejecutable.

Local (automática) \Rightarrow en la pila (*stack*)

La memoria en la pila se administra automáticamente.

Dinámica (manual) \Rightarrow en el *heap*

La memoria en el *heap* se administra manualmente.

Global (estática)

Las variables estáticas existen durante todo el programa

```
int x = 42;
```

```
int main() {  
    int* p = &x;  
    cout << *p << endl;  
    return 0;  
}
```

La pila

La memoria en la pila se administra **automáticamente**.

En C++ las variables locales y los parámetros se almacenan en la pila. El tiempo de vida de una variable está dado por su *scope*.

- ▶ Al declarar una variable local, se apila su valor.
- ▶ Cuando el *scope* de la variable finaliza, se desapila automáticamente su valor.

La pila

Las variables en la pila sólo son referenciables desde su scope

```
void f() {  
    int x = 42;  
}
```

```
int main() {  
    f();  
    int* p = &x;  
    cout << *p << endl;  
}
```

¿Qué sucede al intentar compilar?

La pila

Las variables en la pila sólo son referenciables desde su scope

```
void f() {  
    int x = 42;  
}
```

```
int main() {  
    f();  
    int* p = &x;  
    cout << *p << endl;  
}
```

¿Qué sucede al intentar compilar?

test.cpp: En la función "int main()":

test.cpp:8:13: error: 'x' no se declaró en este ámbito

```
    int* p = &x;
```

La pila

Tiempo de vida de una variable en la pila

```
void g(int* p) {  
    cout << *p << endl;  
}
```

```
int* f() {  
    int x = 42;  
    g(&x);  
    return &x;  
}
```

```
int main() {  
    int* p = f();  
    cout << *p << endl;  
}
```

El *heap*

La memoria en el *heap* se administra **manualmente**.

C++ provee dos operaciones para administrar la memoria dinámica:

- ▶ **new** T — reserva espacio en el *heap* para almacenar un valor de tipo T. Devuelve un puntero de tipo T* a la dirección de memoria donde comienza ese espacio.
- ▶ **delete** p — libera la memoria asociada al puntero p.

El *heap*

Tiempo de vida de una variable en el *heap*

```
int* f() {  
    int* p = new int(0);  
    *p = 42;  
    return p;  
}
```

```
int main() {  
    int* q = f();  
    cout << *q << endl;  
    delete q;  
}
```

Luego de hacer `delete`, se puede reutilizar ese espacio.

El *heap*

También se pueden reservar arreglos de tamaño *dinámico*, cuyo tamaño se elige en tiempo de ejecución:

- ▶ `new T[n]` — reserva espacio en el *heap* para almacenar contiguamente *n valores* de tipo T. Devuelve un puntero de tipo T^* a la dirección de memoria donde comienza ese espacio.
- ▶ `delete[] p` — libera la memoria asociada al arreglo que empieza en la dirección p.

Implementación de Vec<T>

Podemos completar la implementación de Vec<T>:

```
template<class T>
class Vec<T> {
public:
    Vec();
    int size() const;
    T get(int i) const;
    void set(int i, T x);
    void push_back(T x);
private:
    ???
};
```

Implementación de Vec<T>

Podemos completar la implementación de Vec<T>:

```
template<class T>
class Vec<T> {
public:
    Vec();
    int size() const;
    T get(int i) const;
    void set(int i, T x);
    void push_back(T x);
private:
    int _capacidad;
    int _tam;
    T* _valores;
};
```


Implementación de Vec<T>

```
template<class T> Vec<T>::Vec() : _capacidad(1),  
                                _tam(0),  
                                _valores(new T[1]) { }  
  
template<class T> int Vec<T>::size() const {  
    return _tam;  
}  
  
template<class T> T Vec<T>::get(int i) const {  
    return _valores[i];  
}  
  
template<class T> void Vec<T>::set(int i, T x) {  
    _valores[i] = x;  
}
```

Implementación de Vec<T>

Cada vez que la capacidad del arreglo se llena:

1. Solicitar un arreglo del doble de la capacidad.
2. Copiar los datos de la vieja estructura a la nueva.
3. Liberar la memoria de la vieja estructura.
4. Actualizar las variables.

Implementación de Vec<T>

```
template<class T>
void Vec<T>::push_back(T x) {
    if (_tam == _capacidad) {
        T* nuevo = new T[2 * _capacidad];
        for (int i = 0; i < _capacidad; i++) {
            nuevo[i] = _valores[i];
        }
        delete[] _valores;
        _capacidad = 2 * _capacidad;
        _valores = nuevo;
    }
    _valores[_tam] = x;
    _tam++;
}
```

Problemas con punteros

Problema con punteros: *leaks*

- ▶ Cada vez que se hace un `new T`, se debe hacer un `delete` de esa dirección de memoria posteriormente.
- ▶ De lo contrario el programa *pierde memoria* (tiene un *leak*).

```
int main() {  
    int* p = new int;  
}
```

Nuestra implementación de `Vec<T>` tiene un *leak*.

¿Dónde?

(En breve lo arreglaremos).

Problemas con punteros

Otro problema con punteros: acceso a memoria liberada *dangling pointers*

- Una vez que hicimos `delete` de una dirección de memoria, no deberíamos acceder a su contenido.

```
int main() {  
    int* p = new int;  
    *p = 42;  
    delete p;  
    cout << *p << endl;  
}
```

Destructores (motivación)

- ▶ Cuando termina el *scope* de una variable local *x* de tipo *T*, esa memoria se recupera automáticamente.
- ▶ ¿Qué pasa si *x* tiene internamente punteros a estructuras que están almacenadas en el *heap*?

Por ejemplo:

```
int main() {  
    Vec<int> v;  
    v.push_back(1);  
}
```

Destructores (motivación)

- ▶ Cuando termina el *scope* de una variable local *x* de tipo *T*, esa memoria se recupera automáticamente.
- ▶ ¿Qué pasa si *x* tiene internamente punteros a estructuras que están almacenadas en el *heap*?

Por ejemplo:

```
int main() {  
    Vec<int> v;  
    v.push_back(1);  
}
```

- ▶ **Problema:** Finaliza el scope de *v* pero nunca se hizo `delete[]` del arreglo privado `v._valores`.

Destruyores

- ▶ Cada vez que se libera la memoria de un objeto de tipo T, C++ invoca implícitamente al **destructor** del tipo T.
- ▶ El destructor de una clase T se llama T::~~T().
- ▶ El programador nunca debe llamar explícitamente al destructor.

```
template<class T>
class Vec {
    ...
public:
    ~Vec();
};

template<class T>
Vec<T>::~~Vec() {
    delete[] _valores;
}
```


Referencias

Otra forma de usar punteros: referencias

- ▶ Una variable local o parámetro se puede declarar como una referencia a un valor de tipo T, dándole tipo T&.
- ▶ Una referencia es un puntero pero que debe ser inicializado apuntando a algún lugar y no puede cambiar el lugar al que apunta durante su ciclo de vida.

Ejemplo: ambas funciones hacen lo mismo

```
int main() {  
    int a = 41;  
    int& b = a;  
    b = b + 1;  
    cout << a << endl;  
}
```

```
int main() {  
    int a = 41;  
    int* b = &a;  
    *b = *b + 1;  
    cout << a << endl;  
}
```

Referencias

Pasaje de parámetros por referencia

```
void f(int& x, int y) {  
    x++;  
    y++;  
}
```

```
int main() {  
    int a = 1;  
    int b = 1;  
    f(a, b);  
    cout << a << endl;  
    cout << b << endl;  
}
```

Referencias

Devolución de resultados por referencia

```
template<class T>
class Vec { ...
public:
    T& operator[](int i) ;
};

template<class T>
T& Vec<T>::operator[](int i) {
    return _valores[i];
}

int main() {
    Vec v;
    v.push_back(1);
    v[0] = 10;
    cout << v[0] << endl;
}
```

Referencias `const` (motivación)

Consideremos la función que recibe un vector y suma sus primeros dos elementos:

```
int sumaPrimeros(vector<int> v) {  
    return v[0] + v[1];  
}
```

Problema: el parámetro se pasa por copia. Esto es extremadamente ineficiente.

Referencias `const` (motivación)

Podemos arreglar el problema de eficiencia si recibimos el vector por referencia:

```
int sumaPrimeros(vector<int>& v) {  
    return v[0] + v[1];  
}
```

Nuevo problema: no hay ninguna garantía de que la función no modifique su parámetro.

Referencias `const`

El tipo `const T&` representa una referencia **immutable** a un valor de tipo `T`:

```
int sumaPrimeros(const vector<int>& v) {  
    return v[0] + v[1];  
}
```

Referencias const

Tenemos un conjunto implementado sobre un arreglo sin repetidos:

```
template<class T>
class Conj {
public:
    void agregar(const T& x);
    bool pertenece(const T& x) const;
private:
    vector<T> _elementos;
};
```

Referencias `const`

¿Cómo agregamos un método para obtener un vector con todos los elementos del conjunto? Comparar las siguientes cuatro opciones:

1. `vector<T>& Conj<T>::elementos()`
2. `vector<T> Conj<T>::elementos() const`
3. `vector<T>& Conj<T>::elementos() const`
4. `const vector<T>& Conj<T>::elementos() const`

1. Doy una referencia modificable de elementos.
2. Doy una copia modificable.
3. Intenta dar una referencia modificable, pero no compila porque el metodo es `const`.
4. Doy una referencia no modificable.

Preguntas

- ▶ ¿Qué pasa si termina el *scope* del conjunto y queremos usar sus elementos?
- ▶ ¿Qué pasa si el usuario modifica el vector de elementos?

Punteros `const`

También se puede usar `const` con punteros.

```
int main()
{
    int valor = 5 ;

    // puntero que apunta a un int
    int* puntero0 = &valor ;

    // puntero que apunta a un const int
    const int* puntero1 = &valor;

    // puntero const que apunta a un int
    int* const puntero2 = &valor;

    // puntero const que apunta a un const int
    const int* const puntero3 = &valor;

    return 0;
```

Testing

¿Cómo comprobamos que la implementación no tiene problemas de memoria?

- ▶ *Leaks.*
- ▶ Acceso a memoria liberada *Dangling pointers.*
- ▶ Doble `delete`.
- ▶ Desreferencia de `NULL` (`*NULL`).

Es un problema difícil en general.

- ▶ En algunos lenguajes modernos (ej. rust) el compilador puede garantizar, a través del sistema de tipos, que el programa usa la memoria de manera segura.
- ▶ En C++ tenemos que hacer *testing*. Usaremos la herramienta `valgrind`:

```
valgrind --leak-check=full ./programa
```

Leak

lib.cpp

```
1  int* crear() {  
2      int* p = new int;  
3      *p = 42;  
4      return p;  
5  }
```

tests.cpp

```
6  TEST(punteros, leak) {  
7      int* ps[3];  
8      ps[0] = crear();  
9      ps[1] = crear();  
10     ps[2] = crear();  
11  
12     delete ps[0];  
13     delete ps[1];  
14 }
```

Leak

```
==4849== HEAP SUMMARY:
==4849==      in use at exit: 4 bytes in 1 blocks
==4849==    total heap usage: 159 allocs, 158 frees, 108,612 bytes allocated
==4849==
==4849== 4 bytes in 1 blocks are definitely lost in loss record 1 of 1
==4849==    at 0x483BE63: operator new(unsigned long) (in /usr/lib/x86_64-linux-gnu/valgrind/vgpreload_men
==4849==    by 0x110A8F: crear() (lib.cpp:2)
==4849==    by 0x110B33: punteros_leak_Test::TestBody() (tests.cpp:10)
==4849==    by 0x1437D8: void testing::internal::HandleSehExceptionsInMethodIfSupported<testing::Test, vo
==4849==    by 0x13C8A6: void testing::internal::HandleExceptionsInMethodIfSupported<testing::Test, void>
==4849==    by 0x11924B: testing::Test::Run() (gtest-all.cc:3973)
==4849==    by 0x119C08: testing::TestInfo::Run() (gtest-all.cc:4149)
==4849==    by 0x11A2CF: testing::TestCase::Run() (gtest-all.cc:4267)
==4849==    by 0x1256FC: testing::internal::UnitTestImpl::RunAllTests() (gtest-all.cc:6633)
==4849==    by 0x144C77: bool testing::internal::HandleSehExceptionsInMethodIfSupported<testing::internal
==4849==    by 0x13D968: bool testing::internal::HandleExceptionsInMethodIfSupported<testing::internal::U
==4849==    by 0x1240D7: testing::UnitTest::Run() (gtest-all.cc:6242)
==4849==
==4849== LEAK SUMMARY:
==4849==    definitely lost: 4 bytes in 1 blocks
==4849==    indirectly lost: 0 bytes in 0 blocks
==4849==    possibly lost: 0 bytes in 0 blocks
==4849==    still reachable: 0 bytes in 0 blocks
==4849==    suppressed: 0 bytes in 0 blocks
==4849==
==4849== For lists of detected and suppressed errors, rerun with: -s
==4849== ERROR SUMMARY: 1 errors from 1 contexts (suppressed: 0 from 0)
```

Invalid read

lib.cpp

```
7  int incrementar(int* p) {  
8      return *p + 10;  
9  }
```

tests.cpp

```
20  TEST(punteros, invalid_read) {  
21      int* x = crear();  
22      delete x;  
23      int y = incrementar(x);  
24  }
```

Invalid read

```
==4876== Invalid read of size 4
==4876==    at 0x110AB4: incrementar(int*) (lib.cpp:8)
==4876==    by 0x110B38: punteros_invalid_read_Test::TestBody() (tests.cpp:23)
==4876==    by 0x14379C: void
↳ testing::internal::HandleSehExceptionsInMethodIfSupported<testing::Test,
↳ void>(testing::Test*, void (testing::Test::*)(), char const*) (gtest-all.cc:3899)
==4876==    by 0x13C86A: void
↳ testing::internal::HandleExceptionsInMethodIfSupported<testing::Test, void>(testing::Test*,
↳ void (testing::Test::*)(), char const*) (gtest-all.cc:3935)
==4876==    by 0x11920F: testing::Test::Run() (gtest-all.cc:3973)
==4876==    by 0x119BCC: testing::TestInfo::Run() (gtest-all.cc:4149)

...

==4876== HEAP SUMMARY:
==4876==    in use at exit: 0 bytes in 0 blocks
==4876==    total heap usage: 159 allocs, 159 frees, 108,666 bytes allocated
==4876==
==4876== All heap blocks were freed -- no leaks are possible
==4876==
==4876== For lists of detected and suppressed errors, rerun with: -s
==4876== ERROR SUMMARY: 1 errors from 1 contexts (suppressed: 0 from 0)
```

Invalid write

lib.cpp

```
11 void sobrecribir(int* p) {  
12     *p = 20;  
13 }
```

tests.cpp

```
30 TEST(punteros, invalid_write) {  
31     int* x = crear();  
32     delete x;  
33     sobrecribir(x);  
34 }
```

Invalid write

```
==4908== Invalid write of size 4
==4908==    at 0x110ACB: sobreescribir(int*) (lib.cpp:12)
==4908==    by 0x110B38: punteros_invalid_write_Test::TestBody() (tests.cpp:33)
==4908==    by 0x143798: void
↳ testing::internal::HandleSehExceptionsInMethodIfSupported<testing::Test,
↳ void>(testing::Test*, void (testing::Test::*)(), char const*) (gtest-all.cc:3899)
==4908==    by 0x13C866: void
↳ testing::internal::HandleExceptionsInMethodIfSupported<testing::Test, void>(testing::Test*,
↳ void (testing::Test::*)(), char const*) (gtest-all.cc:3935)
==4908==    by 0x11920B: testing::Test::Run() (gtest-all.cc:3973)
==4908==    by 0x119BC8: testing::TestInfo::Run() (gtest-all.cc:4149)

...

==4876== HEAP SUMMARY:
==4876==    in use at exit: 0 bytes in 0 blocks
==4876==    total heap usage: 159 allocs, 159 frees, 108,666 bytes allocated
==4876==
==4876== All heap blocks were freed -- no leaks are possible
==4876==
==4876== For lists of detected and suppressed errors, rerun with: -s
==4876== ERROR SUMMARY: 1 errors from 1 contexts (suppressed: 0 from 0)
```


Double free

lib.cpp

```
15 void limpiar(int* p) {  
16     delete p;  
17 }
```

tests.cpp

```
40 TEST(punteros, double_free) {  
41     int* x = crear();  
42     limpiar(x);  
43     limpiar(x);  
44 }
```

Double free

```
==4944== Invalid free() / delete / delete[] / realloc()
==4944==    at 0x483D1CF: operator delete(void*, unsigned long) (in
↳ /usr/lib/x86_64-linux-gnu/valgrind/vgpreload_memcheck-amd64-linux.so)
==4944==    by 0x110AF9: limpiar(int*) (lib.cpp:16)
==4944==    by 0x110B2E: punteros_double_free_Test::TestBody() (tests.cpp:43)
==4944==    by 0x14378E: void
↳ testing::internal::HandleSehExceptionsInMethodIfSupported<testing::Test,
↳ void>(testing::Test*, void (testing::Test::*)(), char const*) (gtest-all.cc:3899)
==4944==    by 0x13C85C: void
↳ testing::internal::HandleExceptionsInMethodIfSupported<testing::Test, void>(testing::Test*,
↳ void (testing::Test::*)(), char const*) (gtest-all.cc:3935)
==4944==    by 0x119201: testing::Test::Run() (gtest-all.cc:3973)

...

==4944== HEAP SUMMARY:
==4944==    in use at exit: 0 bytes in 0 blocks
==4944==    total heap usage: 159 allocs, 160 frees, 108,666 bytes allocated
==4944==
==4944== All heap blocks were freed -- no leaks are possible
==4944==
==4944== For lists of detected and suppressed errors, rerun with: -s
==4944== ERROR SUMMARY: 1 errors from 1 contexts (suppressed: 0 from 0)
```