Algoritmos de ordenamiento sobre secuencias

Algoritmos y Estructuras de Datos I

Ordenamiento de vectores

 Modificamos el vector solamente a través de intercambios de elementos.

```
proc swap(\text{inout } s : seq\langle \mathbb{Z} \rangle, \text{in } i,j : \mathbb{Z})\{

Pre \{0 \le i,j < |s| \land s = S_0\}

Post \{s[i] = S_0[j] \land s[j] = S_0[i] \land

(\forall k : \mathbb{Z})(0 \le k < |s| \land i \ne k \land j \ne k \rightarrow_L s[k] = S_0[k])\}

}
```

▶ Propiedad:

```
\{s = S_0\}

swap(s,i,j)

\{mismos(s, S_0)\}
```

▶ De esta forma, nos aseguramos que $mismos(s, S_0)$ a lo largo de la ejecución del algoritmo.

Ordenamiento de vectores

```
    proc ordenar(inout s : seq\langle \mathbb{Z} \rangle) {
        Pre \{s = S_0\}
        Post \{mismos(s, S_0) \land ordenado(s)\}
    }

    pred mismos(s, t : seq\langle \mathbb{Z} \rangle) {
        |s| = |t| \land
        (\forall i : \mathbb{Z})(0 \le i < |s|
        \rightarrow_L \#apariciones(s, s[i]) = \#apariciones(t, s[i]))
    }

    fun \#apariciones(s : seq\langle T \rangle, e : T) : \mathbb{Z} =
        \sum_{i=0}^{|s|-1} (\text{if } s[i] = e \text{ then } 1 \text{ else } 0 \text{ fi})

    pred ordenado(s : seq\langle \mathbb{Z} \rangle) {
        (\forall i : \mathbb{Z})(0 \le i < |s| - 1 \rightarrow_L s[i] \le s[i + 1])
    }
}
```

Ordenamiento por selección (Selection Sort)

- Notación: s[i,j] = subseq(s,i,j+1). s[i,j) = subseq(s,i,j).
- ▶ Observación: subseq nunca se indefine (devuelve la secuencia vacía ⟨⟩)
- ▶ **Idea:** Seleccionar el mínimo elemento e intercambiarlo con la primera posición del vector. Repetir con el segundo, etc.

```
void seleccion(vector<int> &s) {
    for(int i=0; i<s.size(); i++) {
        int pos = // ubicacion del minimo de s entre i y s.size()
        swap(s, i, pos);
    }
}</pre>
```

Ordenamiento por selección (Selection Sort)

Podemos refinar un poco el código:

```
void seleccion(vector<int> &s) {
    for(int i=0; i<s.size()-1; i++) {
        int pos = minimo(s, i, s.size());
        swap(s, i, pos);
    }
}</pre>
```

➤ Surge la necesidad de especificar el problema auxiliar de buscar el mínimo entre *i* y *s.size*():

```
proc minimo(inout \ s : seq\langle \mathbb{Z} \rangle, in \ d, h : Int, out \ result : \mathbb{Z})\{

Pre \{0 \le d < h \le |s|\}

Post \{d \le result < h

\land_L \ (\forall i : \mathbb{Z})(d \le i < h \rightarrow_L s[result] \le s[i])\}

}
```

Recap: Teorema de corrección de un ciclo

▶ **Teorema.** Sean un predicado I y una función $fv : \mathbb{V} \to \mathbb{Z}$ (donde \mathbb{V} es el producto cartesiano de los dominios de las variables del programa), y supongamos que $I \Rightarrow \text{def}(B)$. Si

```
1. P_C \Rightarrow I,

2. \{I \land B\} S \{I\},

3. I \land \neg B \Rightarrow Q_C,

4. \{I \land B \land v_0 = fv\} S \{fv < v_0\},

5. I \land fv < 0 \Rightarrow \neg B,
```

... entonces la siguiente tripla de Hoare es válida:

```
\{P_C\} while B do S endwhile \{Q_C\}
```

Buscar el Mínimo Elemento

```
▶ proc minimo(inout s : seq\langle \mathbb{Z} \rangle, in d, h : \mathbb{Z}, out result : \mathbb{Z}){

Pre \{0 \le d < h \le |s|\}

Post \{d \le result < h

\land_L(\forall i : \mathbb{Z})(d \le i < h \rightarrow_L s[result] \le s[i])\}
}
```

▶ Necesitamos implementar la especificación de **minimo**:

```
int minimo(vector<int> &s, int d, int h) {
    int result = d;
    for(int i=d+1; i<h; i++) {
        if (s[i]<s[result]) {
            result = i;
        }
    }
    return result;
}</pre>
```

Buscar el Mínimo Elemento

```
ightharpoonup P_C \equiv 0 \le d < h \le |s| \land result = d \land i = d+1
```

- $ightharpoonup B \equiv i < h$
- I ≡ ?
 - En cada iteración del ciclo, la variable **result** contiene el índice del menor elemento encontrado hasta ahora. Es decir, s[result] el menor elemento de s[d, i).
 - ▶ ¡Esto nos proporciona el invariante del ciclo!

```
d \leq result < i \leq h \land_L (\forall j : \mathbb{Z})(d \leq j < i \rightarrow_L s[result] \leq s[j])
```

```
ightharpoonup fv = h - i
```

Buscar el Mínimo Elemento

```
▶ P_C \equiv 0 \le d < h \le |s| \land result = d \land i = d + 1

▶ Q_C \equiv d \le result < h

\land L(\forall i : \mathbb{Z})(d \le i < h \rightarrow_L s[result] \le s[i])

▶ B \equiv i < h

▶ I \equiv d \le result < i \le h

\land L(\forall j : \mathbb{Z})(d \le j < i \rightarrow_L s[result] \le s[j])

▶ fv = h - i

int minimo(vector<int> &s, int d, int h) {

int result = s[d];

for(int i=d+1; i<h; i++) {

if (s[result] < s[i]) {

result = i;

}

return result;
```

Correctitud: Buscar el Mínimo Elemento

```
► I \equiv d \le result < i \le h

\land_L (\forall j : \mathbb{Z})(d \le j < i \rightarrow_L s[result] \le s[j])

► f_V = h - i
```

```
int minimo(vector<int> &s, int d, int h) {
    int result = s[d];
    for(int i=d+1; i<h; i++) {
        if (s[result] < s[i]) {
            result = i;
        }
    }
    return result;
}</pre>
```

- ¿I se preserva en cada iteración (punto 2.)? √
- ¿La función variante es estrictamente decreciente (punto 4.)?√

Correctitud: Buscar el Mínimo Elemento

- $P_C \equiv 0 \le d < h \le |s| \land \textit{result} = d \land i = d+1$
- \triangleright $B \equiv i < h$
- ► $I \equiv d \le result < i \le h$ $\land_L (\forall j : \mathbb{Z})(d \le j < i \rightarrow_L s[result] \le s[j])$
- ightharpoonup fv = h i
- ▶ ¿I es se cumple al principio del ciclo (punto 1.)? ✓
- ¿Se cumple la postcondición del ciclo a la salida del ciclo (punto 3.)? √
- ¿Si la función variante alcanza la cota inferior la guarda se deja de cumplir (punto 5.)? √

Ordenamiento por selección (Selection Sort)

▶ Volvamos ahora al programa de ordenamiento por selección:

```
void seleccion(vector<int> &s) {
for(int i=0; i<s.size(); i++) {
    int pos = minimo(s, i, s.size());
    swap(s, i, pos);
}
</pre>
```

- ho $P_C \equiv i = 0 \land s = S_0$
- $ightharpoonup Q_C \equiv mismos(s, S_0) \land ordenado(s)$
- $ightharpoonup B \equiv i < |s|$
- I ≡ ?
 - Luego de la i-ésima iteración, s[0, i) contiene los i primeros elementos ordenados! ¿Tenemos entonces el invariante del ciclo?
 - $I \equiv mismos(s, S_0) \land ordenado(s[0, i)) \land (0 < i < |s|)$
- ightharpoonup fv = |s| i

Ordenamiento por selección (Selection Sort)

- $I \equiv 0 \le i \le |s| \land mismos(s, S_0) \land ordenado(s[0, i))$
- ightharpoonup fv = |s| i

```
void seleccion(vector < int > &s) {
    for(int i=0; i < s.size(); i++) {
        int pos = minimo(s, i, s.size());
        swap(s, i, pos);
    }
}</pre>
```

- ▶ ¿I se preserva en cada iteración (punto 2.)? X
- ► Contraejemplo:
 - ▶ Si arrancamos la iteración con i = 1 y $s = \langle 100, 2, 1 \rangle$
 - ► Terminamos con i = 2 y $s = \langle 100, 1, 2 \rangle$ que no satisface I

Debemos reforzar el invariante para probar la corrección:

```
I \equiv mismos(s, S_0) \land ordenado(s[0, i)) \land(0 \le i \le |s| \land_L (\forall j, k : \mathbb{Z}) ((0 \le j < i \land i \le k < |s|) \rightarrow_L s[j] \le s[k]))
```

Correctitud: Ordenamiento por selección (Selection Sort)

- $P_C \equiv i = 0 \land s = S_0$
- $ightharpoonup Q_C \equiv mismos(s, S_0) \land ordenado(s)$
- $ightharpoonup B \equiv i < |s|$
- ► $I \equiv mismos(s, S_0) \land ordenado(s[0, i)) \land (0 \le i \le |s| \land_L (\forall j, k : \mathbb{Z})((0 \le j < i \land i \le k < |s|) \rightarrow_L s[j] \le s[k]))$
- ightharpoonup fv = |s| i
- il es se cumple al principio del ciclo (punto 1.)? √
- ¿Se cumple la postcondición del ciclo a la salida del ciclo (punto 3.)?√
- ¿Si la función variante alcanza la cota inferior la guarda se deja de cumplir (punto 5.)?√

Correctitud: Ordenamiento por selección (Selection Sort)

```
 \begin{array}{ll} I & \equiv & mismos(s,S_0) \ \land \ ordenado(s[0,i)) \\ & \land \ (0 \leq i \leq |s| \land_L (\forall j,k:\mathbb{Z})((0 \leq j < i \land i \leq k < |s|) \rightarrow_L s[j] \leq s[k])) \end{array}
```

Gráficamente:

Correctitud: Ordenamiento por selección (Selection Sort)

```
▶ I \equiv mismos(s, S_0) \land ordenado(s[0, i)) \land (0 \le i \le |s| \land_L (\forall j, k : \mathbb{Z})((0 \le j < i \land i \le k < |s|) \rightarrow_L s[j] \le s[k]))
```

fv = |s| - i

```
void seleccion(vector<int> &s) {
    for(int i=0; i<s.size(); i++) {
        int pos = minimo(s, i, s.size());
        swap(s, i, pos);
    }
}</pre>
```

- ▶ ¿I se preserva en cada iteración (punto 2.)? ✓
- ¿La función variante es estrictamente decreciente (punto 4.)?√

Ordenamiento por selección (Selection Sort)

```
int minimo(vector<int> &s, int d, int h) {
    int result = s[d];
    for(int i=d+1; i<h; i++) {
        if (s[result] < s[i]) {
            result = i;
        }
      }
    return result;
    }
    void seleccion(vector<int> &s) {
      for(int i=0; i<s.size(); i++) {
        int pos = minimo(s,i,s.size());
        swap(s, i, pos);
    }
}</pre>
```

- ¿Cómo se comporta este algoritmo?
- ▶ Veámoslo en https://visualgo.net/es/sorting.

Ordenamiento por selección (Selection Sort)

- ► Variantes del algoritmo básico:
 - 1. Cocktail sort: consiste en buscar en cada iteración el máximo y el mínimo del vector por ordenar, intercambiando el mínimo con i y el máximo con |s| i 1.
 - 2. Bingo sort: consiste en ubicar todas las apariciones del valor mínimo en el vector por ordenar, y mover todos los valores mínimos al mismo tiempo (efectivo si hay muchos valores repetidos).
- Ambas variantes también son algoritmos cuadráticos (iteran una cantidad cuadrática de veces).

Ordenamiento por selección (Selection Sort)

- ¿Cuántas iteraciones ejecuta este programa en peor caso?
 - ► Para ello contamos la cantidad de veces que se ejecuta el **if** de minimo

$$\operatorname{ejecuciones}_{if} \ = \ \sum_{i=0}^{|s|-1} |s|-i \ = |s| \times |s| - \frac{(|s|-1) \times |s|}{2} \le (|s|)^2.$$

- Decimos que el algoritmo de ordenamiento por selección es un algoritmo cuadrático (¡más información en algo2!).
- ▶ ¿Puede ejecutarse una cantidad menor de veces?
 - ► Siempre se ejecuta la misma cantidad de veces. El peor caso es igual al mejor caso.

Intervalo

Break!

Ordenamiento por inserción (Insertion Sort)

Veamos un segundo algoritmo, en el que usaremos como invariante únicamente que:

```
I \equiv 0 \le i \le |s| \land_L mismos(s, S_0) \land ordenado(s[0, i))
```

► Esto implica que (a diferencia del algoritmo de selection sort) en cada iteración los primeros *i* elementos están ordenados, sin ser necesariamente los *i* elementos más pequeños del vector.

Ordenamiento por inserción (Insertion Sort)

```
void insercion(vector<int> &s) {
    for(int i=0; i<s.size(); i++) {
        for(int j=i; j>0 && s[j] < s[j-1] ; j--) {
            swap(s, j, j-1);
        }
    }
}
```

```
I_{ext} \equiv mismos(s, S_0) \land ordenado(s[0, i)) \land (0 \le i \le |s|)
```

Li Cuál es el invariante del for interno?

```
\begin{array}{ll} I_{int} & \equiv & 0 \leq j \leq i \\ & \wedge & mismos(s[0,i+1),S_0[0,i+1)) \\ & \wedge & s[i+1,|s|) = S_0[i+1,|s|) \\ & \wedge & ordenado(s[0,j)) \ \wedge & ordenado(s[j,i+1)) \\ & \wedge & (\forall k:\mathbb{Z})(j < k \leq i \rightarrow_L s[j] < s[k]) \end{array}
```

Ordenamiento por inserción (Insertion Sort)

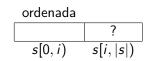
```
I \equiv 0 \leq i \leq |s| \land_L \ mismos(s, S_0) \land \ ordenado(s[0, i))
void insercion(vector<int> &s) {
for(int i=0; i<s.size(); i++) {
      // Tenemos que preservar el invariante ...
}
}
```

- ▶ ¿/ es se cumple al principio del ciclo (punto 1.)? ✓
- ¿Se cumple la postcondición del ciclo a la salida del ciclo (punto 3.)? √
- ▶ ¿I se preserva en cada iteración (punto 2.)?
 - Sabiendo que los primeros i elementos están ordenados, tenemos que hacer que los primeros i+1 elementos pasen a estar ordenados!
 - ¿Cómo lo podemos hacer?
 - Insertando s[i] en la posición temporaria que le corresponda!

Ordenamiento por inserción (Insertion Sort)

Gráficamente:

► Invariante del ciclo exterior



► Invariante del ciclo interior

ordenada		ordenada	
	X	> x	?
s[0,j)	s[j]	s[j+1,i+1)	s[i+1, s)

Ordenamiento por inserción (Insertion Sort)

```
void insercion(vector<int> &s) {
    for(int i=0; i<s.size(); i++) {
        for( int j=i; j>0 && s[j] < s[j-1] ; j--) {
            swap(s, j, j-1);
        }
    }
}</pre>
```

Li Cómo son las funciones variantes de cada ciclo?

$$fv_{ext} = |s| - i$$

 $fv_{int} = j$

Ordenamiento por inserción (Insertion Sort)

- Luántas veces se ejecuta el swap del ciclo interior?
 - ► ¡Depende de los datos!
- Analizamos el peor caso (es decir, que el while interior realice i + 1 iteraciones).

ejecuciones
$$_{if}$$
 = $\sum_{i=0}^{|s|} i + 1 = \frac{|s| \times (|s|+1)}{2} + |s| \leq |s|^2$

- ► El algoritmo de ordenamiento por inserción también es un algoritmo cuadrático (itera una cantidad cuadrática de veces)
- ▶ **Observación:** Selection sort e insertion sort se pueden generalizar a secuencias de tipo T con un predicado de orden \leq (no solamente funcionan con secuencias de \mathbb{Z})

Ordenamiento por inserción (Insertion Sort)

```
void insercion(vector<int> &s) {
    for(int i=0; i<s.size(); i++) {
    for( int j = i; j>0 && s[j] < s[j-1] ; j--) {
        swap(s, j, j-1);
    }
}</pre>
```

- > ¿Cómo se comporta este algoritmo de ordenamiento?
- ▶ Veámoslo en https://visualgo.net/es/sorting.

Dutch National Flag Problem

Dado una secuencia que contiene colores (rojo, blanco y azul) ordenarlos de modo que respeten el orden de la bandera holandesa (primero rojo, luego blanco y luego azul)



Por ejemplo, si la secuencia es:

(White, Red, Blue, Blue, Red)

El programa debe modificar la secuencia para que quede:

⟨Red, Red, White, Blue, Blue⟩

Dutch National Flag Problem

- ➤ Si Red=0,White=1 y Blue=2, ¿Cuál sería la especificación del problema?
- ▶ proc $dutchNationalFlag(inout s : seq\langle \mathbb{Z} \rangle)$ {
 Pre $\{s = S_0 \land (\forall e : \mathbb{Z})(e \in s \leftrightarrow (e = 0 \lor e = 1 \lor e = 2))\}$ Post $\{mismos(s, S_0) \land ordenado(s)\}$ }
- L'Cómo podemos implementar una solución a este problema?
 - → ¿Podemos usar algún algoritmo de ordenamiento que conozcamos? Rta: podemos usar insertionSort o selectionSort.
 - Luál es el peor caso? Rta: $(|s|)^2$
 - ▶ ¿Podemos hacer que tenga un peor caso mas eficiente?

Dutch National Flag Problem

```
#define RED 0
#define WHITE 1
#define BLUE 2

void dutchNationalFlag(vector<int> &s) {
// contamos la cantidad de apariciones de cada color
vector<int> colorCount = fillColorCount(s);

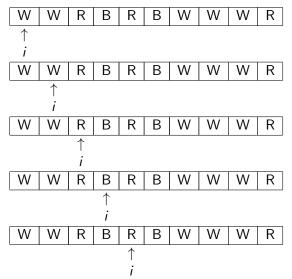
// usamos la cantidad de apariciones para repoblar la secuencia
populate(s,colorCount);
}
```

Dutch National Flag Problem

- 1. En una pasada contar la cantidad de apariciones de cada color: RED (0), WHITE (1) y BLUE (2), almacenándolas en una secuencia de tres elementos.
- 2. En una segunda pasada repoblar la secuencia usando las cantidades leídas.
- 3. ¿Qué cantidad de iteraciones se ejecutaría en peor caso?
 - **Rta:** $2 * |s| \le (|s|)^2$

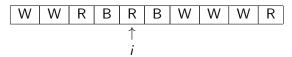
Dutch National Flag Problem

¿Qué invariante de ciclo podemos usar para contar la cantidad de apariciones de cada color?



Dutch National Flag Problem

¿Qué invariante de ciclo podemos usar para contar la cantidad de apariciones de cada color?



```
I \equiv (0 \le i \le |s|) \land_L

(colorCount[RED] = \#apariciones(subseq(s, 0, i), RED) \land

colorCount[WHITE] = \#apariciones(subseq(s, 0, i), WHITE) \land

colorCount[BLUE] = \#apariciones(subseq(s, 0, i), BLUE))
```

¿Qué función variante podemos usar en la primer pasada?

$$fv \equiv |s| - i$$

Dutch National Flag Problem

```
vector<int> fillColorCount(vector<int> &s) {
     vector<int> colorCount = \{0,0,0\}:
     for(int i=0; i<s.size(); i++) {
       if (s[i]==RED) {
4
         colorCount[RED]++;
5
       } else if (s[i]==WHITE) {
         colorCount[WHITE]++;
7
        } else {
8
         colorCount[BLUE]++;
10
11
     return colorCount;
12
13
```

¿Qué cantidad de iteraciones ejecuta fillColorCount en **peor caso? Rta:** Realiza 3+|s|-iteraciones en peor caso.

Dutch National Flag Problem

```
vector<int> fillColorCount(vector<int> &s) {
            ...
            }
```

Dutch National Flag Problem

Ahora nos falta repoblar la secuencia con los valores. ¿Qué invariante de ciclo tendríamos que respetar?

```
\begin{split} I &\equiv \textit{mismos}(s, S_0) \land \\ & \big( (0 \leq i \leq |s|) \land_L \\ & (\textit{ordenado}(\textit{subseq}(s, 0, i)) \land \\ & \textit{colorCount}[\textit{RED}] = \#a...(S_0, \textit{RED}) - \#a...(\textit{subseq}(s, 0, i), \textit{RED})) \land \\ & \textit{colorCount}[\textit{WHITE}] = \#a...(S_0, \textit{WHITE}) - \#a...(\textit{subseq}(s, 0, i), \textit{WHITE})) \land \\ & \textit{colorCount}[\textit{BLUE}] = \#a...(S_0, \textit{BLUE}) - \#a...(\textit{subseq}(s, 0, i), \textit{BLUE})))) \end{split}
```

¿Qué función variante podemos usar en la primer pasada?

$$fv \equiv |s| - i$$

Dutch National Flag Problem

```
void populate(vector<FlagColor> &s, vector<int> &colorCount) {
     for(int i=0; i<s.size(); i++) {
       if (colorCount[RED]>0) {
         s[i] = RED:
5
         colorCount[RED]--;
       } else if (colorCount[WHITE]>0) {
         s[i] = WHITE;
         colorCount[WHITE]--;
9
       } else {
10
         s[i] = BLUE;
11
         colorCount[BLUE]--;
12
13
14
15
```

¿Qué cantidad de iteraciones ejecuta populate en **peor caso**? **Rta:** Realiza |s|-iteraciones en peor caso.

Eficiencia de los Algoritmos de ordenamiento

- Tanto selection sort como insertion sort son algoritmos cuadráticos (iteran una cantidad cuadrática de veces)
- ¿Hay algoritmos con comportamiento más eficiente?
 - Quicksort y BubbleSort: Peor caso cuadrático $(|s|)^2$
 - ▶ Mergesort y Heapsort: Peor caso: $|s| \times log_2(|s|)$
 - ► Counting sort (para secuencias de enteros). Peor caso: |s|
 - Radix sort (para secuencias de enteros). Peor caso: 2³²
- ▶ Bubble sort está en la práctica 8. El resto los van a ver en algo2.

Dutch National Flag Problem

```
void dutchNationalFlag(vector < int > &s) {

// realiza (3+|s|)-iteraciones

vector < int > colorCount = fillColorCount(s);

// realiza |s|-iteraciones

populate(s,colorCount);

}
```

¿Qué cantidad de iteraciones ejecuta dutchNationalFlag en **peor caso**?

- ▶ Rta: Realiza 3 + |s| + |s|-iteraciones en peor caso.
- Por lo tanto, a partir de |s| > 3, se cumple que $3 + 2 * |s| < |s|^2$
- ► En otras palabras, en peor caso la implementación es mas eficiente que selectionSort o insertionSort

Bibliografía

- ▶ Vickers et al. Reasoned Programming
 - ▶ 6.5 Insertion Sort
- ▶ NIST- Dictionary of Algorithms and Data Structures
 - ► Selection Sort https://xlinux.nist.gov/dads/HTML/selectionSort.html
 - ▶ Bingo Sort https://xlinux.nist.gov/dads/HTML/bingosort.html
 - Cocktail Sort https://xlinux.nist.gov/dads/HTML/bidirectionalBubbleSort.html