

Universidad Icesi

Facultad de ingeniería

Algoritmos y Estructuras de Datos

Profesor: Juan Manuel Reyes

### **Tarea integradora III**

Integrantes

Alejandra Diaz Parra

Juan Fernando Martinez

Santiago Rodas Rodriguez

## Método de la ingeniería

### Fase 1: Identificación del problema

El planeta Tierra ha llegado a tal punto crítico de inconvenientes y dificultades que una de las posibles soluciones para solucionar estos problemas es embarcarse en una misión para explorar los confines de la Vía Láctea. Es por esta razón que distintos países se han unido para construir una nave, en este caso llamada “USS DISCOVERY”. Ahora bien, al ser una nave interplanetaria, se necesita de un sistema completo de navegación en la nave, basado en una base de datos de sistemas planetarios existentes, claramente con rutas de navegación entre ellos. Con esto, no sólo se desea calcular el recorrido con menor tiempo de un planeta a otro, sino también de conocer la mayor cantidad posible de civilizaciones. Tomando en cuenta esa información introductoria, se tiene los siguientes requerimientos funcionales:

1. Indicar los sistemas más cercanos a la posición de la nave.
2. Calcular el orden en que se deben recorrer todos los sistemas desde la ubicación actual en el menor tiempo posible.
3. Calcular el recorrido que toma menos tiempo para viajar del sistema actual al sistema deseado.
4. Agregar un sistema planetario nuevo al programa y a su base de datos.
5. Eliminar un sistema planetario del software.
6. Implementar una pestaña donde se pueda buscar un sistema planetario para ver su información.
7. Calcular una ruta para recorrer la mayor cantidad de sistemas habitados en el menor tiempo posible partiendo de la ubicación actual de la nave.
8. (Opcional) Que cada campo de búsqueda despliegue una lista de autocompletar.
9. (Opcional) Que se calcule una ruta para recorrer los sistemas en el menor tiempo posible de acuerdo a los siguientes propósitos:
  - a. Explorar los sistemas inhabitados
  - b. Explorar las civilizaciones tipo 0 1 y 2 para la extracción y negociación de recursos.

- c. Explorar las civilizaciones tipo III para entablar relaciones diplomáticas e intercambiar conocimiento y recursos.

## **Fase 2: Recopilación de la información necesaria**

Para solucionar este problema tan complejo e importante, se necesita de una estructura eficiente que pueda almacenar grandes cantidades de datos, y al mismo tiempo tener una o varias relaciones entre ellos. Es por esa razón que, al analizar el problema y entenderlo de manera significativa, se logró encontrar una de las posibles soluciones al problema: los grafos. Ahora bien, al existir diferentes tipos de grafos se encontraron diferentes maneras de diseñarlos, por eso se detalla cada uno a continuación:

**Grafo dirigido:** Un grafo dirigido consta de un conjunto de vértices y aristas donde cada una se asocia de forma unidireccional a través de una flecha con otro. Las aristas, dependiendo de su salida o ingreso, reciben la calificación de entrante o saliente, la condición común es que siempre tienen un destino hacia un nodo.

**Grafo no dirigido:** Los grafos no dirigidos son aquellos que constan un conjunto de vértices que están conectados a un conjunto de aristas de forma no direccional. Esto significa que una arista puede indistintamente recorrer desde cualquiera de sus puntos y en cualquier dirección.

Además de los tipos, también se encuentran algunas propiedades interesantes: Una de las principales es que posee adyacencia. Ésta se trata de la relación entre dos aristas que comparten la conexión o relación con un vértice común. La incidencia se refiere simplemente al caso en el que un vértice está unido a otro. Por último, se deben entender las propiedades de ponderación de los grafos que corresponden a una función en la que cada arista es clasificada y cuantificada en diversos términos para aumentar la expresividad del modelo. Esta característica en especial es muy útil en estudios de optimización.

Tomando en cuenta la información introductoria y/o básica sobre la estructura de datos grafo, se pueden enumerar las posibles representaciones que se tienen de éste:

1. **Matrices de adyacencia:** Una de las maneras más fáciles de implementar un grafo es usar una matriz bidimensional. Aquí, cada una de las filas y columnas representa un vértice en el grafo. El valor que se almacena en la celda en la intersección de la fila V y la columna W indica si hay una arista desde el vértice v y el vértice w. Cuando dos vértices están conectados por una arista, decimos que son adyacentes.
2. **Matrices de peso:** Son similares a las matrices de adyacencia, dónde en vez de un 1 para representar una arista entre dos nodos se agrega un número que representa el peso de cada arista. Este número es 0 u infinito si no hay ninguna relación directa entre dos nodos del grafo, y cualquier otro número cuando sí existe una relación directa.

3. **Matrices de incidencia:** Son matrices binarias que se utilizan como una forma de representar relaciones binarias. Aquí se toman en cuenta varios aspectos: Las columnas de la matriz representan las *aristas* del grafo. Las filas representan a los distintos nodos. Por cada nodo unido por una arista, se indica un uno (1) en el lugar correspondiente, y se llena el resto de las ubicaciones con ceros (0).
4. **Listas de adyacencia:** Una lista de adyacencia es una representación de todas las aristas o arcos de un grafo mediante una lista. Si el grafo es no dirigido, cada entrada es un conjunto o multiconjunto de dos vértices conteniendo los dos extremos de la arista correspondiente. Si el grafo es dirigido, cada entrada es una tupla de dos nodos, uno denotando el nodo fuente y el otro denotando el nodo destino del arco correspondiente.
5. **Representación por medio de relaciones:** La posibilidad de representar de forma gráfica un conjunto de vértices y aristas es una solución importante en términos de análisis. La visualización de grafos no debe ser confundida con la representación del grafo en sí, sino del conjunto complejo que resulta de la posición de los vértices y la orientación de las aristas. Con esto se busca darle una imagen mental al conjunto de datos y su incorrecta representación puede afectar la comprensión, usabilidad y costo de la información.
6. **Listas de aristas o listas de incidencia:** Es una manera relativamente sencilla de almacenar un grafo en donde se tiene un arreglo con un tamaño equivalente al número de aristas. Éste almacena pares ordenados o grupos de dos nodos representando una arista. En un grafo dirigido la primera posición de cada par representa el nodo de salida y en la segunda el nodo de llegada. En un grafo no dirigido no hay distinción de orden.

Además de la información básica y las representaciones que se tuvieron en cuenta, es importante destacar las opciones de autocompletado que en transcurso del tiempo se han investigado e implementado:

1. **ControlsFx:** Librería que da ciertas herramientas útiles para complementar JavaFX. Entre sus herramientas se encuentra la clase `TextFields` que permite enlazar un campo de texto `TextField` con una lista de autocompletar. La librería de `ControlsFx` se encarga de mostrar el listado de predicciones cada vez que se escriba que se actualice el campo de texto.
2. **Trie:** Para obtener las palabras que se muestran en el autocompletar esta estructura es de bastante utilidad. Se encarga de almacenar cada palabra que se agregue en nodos que representan cada carácter de la palabra. Es muy eficiente también para la búsqueda de palabras debido a que la complejidad temporal depende únicamente del largo de la palabra a buscar.

Por último, se tiene como informacion adicional algunos algoritmos que se desean o se deben implementar en la estructura del grafo:

1. **BFS:** Es un algoritmo de búsqueda utilizado para recorrer o buscar elementos en un grafo (usado frecuentemente sobre árboles). Intuitivamente, se comienza en la raíz (eligiendo algún nodo como elemento raíz en el caso de un grafo) y se exploran todos los vecinos de este nodo. A continuación para cada uno de los vecinos se exploran sus respectivos vecinos adyacentes, y así hasta que se recorra todo el grafo.
2. **DFS:** Es un algoritmo de búsqueda utilizado para recorrer todos los nodos de un grafo de manera ordenada, pero no uniforme. Su funcionamiento consiste en ir expandiendo todos y cada uno de los nodos que va localizando, de forma recurrente, en un camino concreto. Cuando ya no quedan más nodos que visitar en dicho camino, regresa, de modo que repite el mismo proceso con cada uno de los hermanos del nodo ya procesado.
3. **Dijkstra:** La idea subyacente en este algoritmo consiste en ir explorando todos los caminos más cortos que parten del vértice origen y que llevan a todos los demás vértices; cuando se obtiene el camino más corto desde el vértice origen hasta el resto de los vértices que componen el grafo, el algoritmo se detiene. Se trata de una especialización de la búsqueda de costo uniforme y, como tal, no funciona en grafos con aristas de coste negativo (al elegir siempre el nodo con distancia menor, pueden quedar excluidos de la búsqueda nodos que en próximas iteraciones bajaría el costo general del camino al pasar por una arista con costo negativo).
4. **Floyd-Warshall:** El algoritmo de Floyd-Warshall compara todos los posibles caminos a través del grafo entre cada par de vértices. El algoritmo es capaz de hacer esto con sólo  $V^3$  comparaciones (esto es notable considerando que puede haber hasta  $V^2$  aristas en el grafo, y que cada combinación de aristas se prueba). Lo hace mejorando paulatinamente una estimación del camino más corto entre dos vértices, hasta que se sabe que la estimación es óptima.
5. **Prim:** es un algoritmo perteneciente a la teoría de grafos para encontrar un un árbol recubridor mínimo en un grafo conexo, no dirigido y cuyas aristas están etiquetadas. En otras palabras, el algoritmo encuentra un subconjunto de aristas que forman un árbol con todos los vértices, donde el peso total de todas las aristas en el árbol es el mínimo posible. Si el grafo no es conexo, entonces el algoritmo encontrará el árbol recubridor mínimo para uno de los componentes conexos que forman dicho grafo no conexo.
6. **Kruskal:** es un algoritmo de la teoría de grafos para encontrar un árbol recubridor mínimo en un grafo conexo y ponderado. Es decir, busca un subconjunto de artistas que, formando un árbol, incluyen todos los vértices y donde el valor de la suma de

todas las aristas del árbol es el mínimo. Si el grafo no es conexo, entonces busca un bosque expandido mínimo.

Complementando toda la información, también se tiene los requerimientos no funcionales:

1. Modelar e implementar el sistema de navegación en dos grafos diferentes.
2. Implementar la interfaz gráfica con JavaFx.
3. Contar con un menú superior en la ventana, el cual permite cambiar todos los elementos cuando se esté trabajando en opciones diferentes.
4. Implementar los algoritmos de recorridos sobre grafos BFS y DFS.
5. Codificar los algoritmos de camino de peso mínimo Dijkstra y Floyd-Warshall.
6. Desarrollar los algoritmos de árbol de recubrimiento mínimo Prim y Kruskal.
7. Cargar los nombres de los sistemas planetarios desde un archivo plano.
8. Generar sistemas a partir de un listado de nombres posibles.

### Fase 3: Búsqueda de soluciones creativas

Para llevar a cabo esta etapa, se formularon tres procesos:

- Analizar los requerimientos indicados para el proyecto para poder diseñar un problema cuya solución pueda plantearse en términos de lo solicitado en el enunciado.
- Realizar una lluvia de ideas entre los integrantes del equipo para determinar el tema del proyecto cuya solución será modelada como un software.
- Definir los atributos y exigencias propias del problema elegido para ser desarrollado.

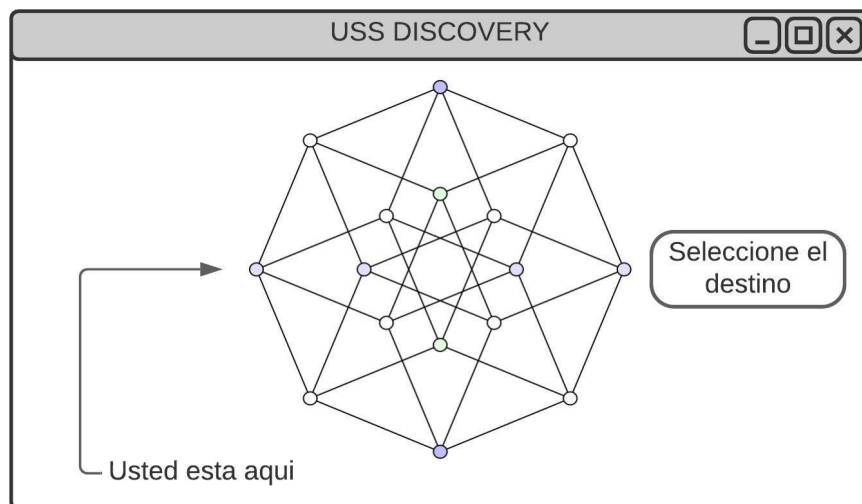
A diferencia de los proyectos trabajados anteriormente, en este proyecto no se realizó ninguna lluvia de ideas para elegir las estructuras con las que se podría resolver mejor el problema, puesto que uno de los requerimientos del proyecto es que sea modelado en grafos. Sin embargo, sí se realizó un análisis para la elección del tipo de grafo preferible para la cuestión a resolver, ya que entre las opciones estaban grafos dirigidos y no dirigidos, y cada uno de estos podían o no presentar aristas que tuvieran peso. Ahora bien, para las dos representaciones de grafos en las que se implementará el problema se tuvieron las ya mencionadas:

- Matrices de peso
- Matrices de incidencia
- Matrices de adyacencia
- Listas de adyacencia
- Representación por medio de relaciones (similar a un árbol)
- Listas de aristas o listas de incidencia

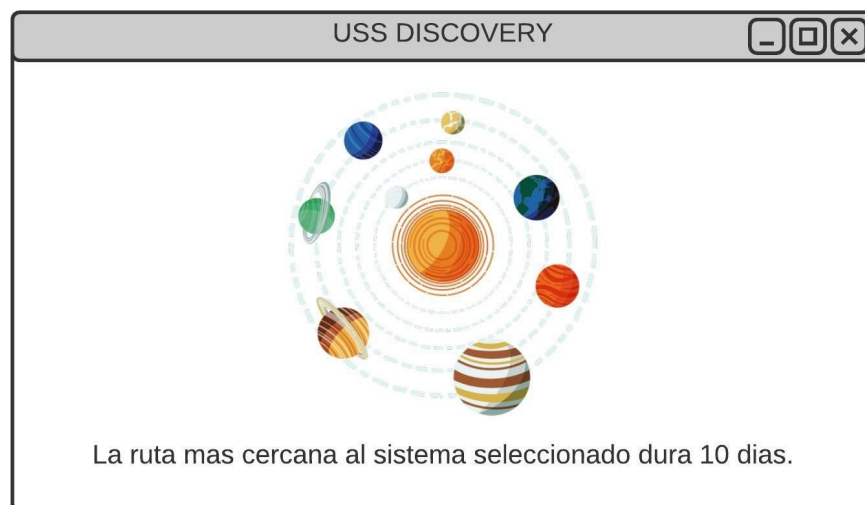
También se plantean como ideas el uso de todos los algoritmos planteados en la fase dos:

1. BFS
2. DFS
3. Dijkstra
4. Floyd-Warshall
5. Prim
6. Kruskal

Teniendo en cuenta todo lo mencionado anteriormente, se planeó la interfaz de usuario usando la herramienta online Lucidchart. Como resultado, se consiguieron los siguientes diseños básicos:







#### Fase 4: Transición de la formulación de ideas a los diseños preliminares

La primera decisión que se toma para la implementación de la solución de software es descartar el uso de grafos dirigidos, pues para la nave es posible navegar de un sistema A a un sistema B de la misma manera en que navegaría del sistema B al sistema A.

La segunda opción que se descarta es la representación a través de grafos no ponderados, ya que para el problema es clave que se tengan en cuenta los pesos de las aristas entre los nodos del grafo, pues estos representan el tiempo de viaje entre sistemas.

Con lo anterior se tiene entonces que los grafos con los que es posible modelar el problema son aquellos que permiten que sus aristas no sean dirigidas, y que además tengan pesos asociados. Por esto último, también se descarta el uso de matrices de adyacencia, resultando esto en las siguientes opciones viables:

- Usar una matriz de peso, ya que sus filas y columnas podrían representar los sistemas de la Vía Láctea, y las celdas de la matriz simularán el tiempo de viaje entre cada uno de estos sistemas.
- La lista de aristas también puede ser de utilidad ya que a cada par de nodos se le puede agregar un tercer campo que represente el tiempo de viaje entre dos sistemas. De igual forma puede crearse un objeto nodo que guarde dicha información.
- La matriz de incidencia puede llegar a resultar igualmente útil, ya que dentro de ésta sucede una situación importante: cada nodo que incide sobre una arista se puede tener un número que indique el tiempo de viaje entre los sistemas que inciden sobre una determinada arista, en vez de un 1 como indicador. También puede tenerse una lista de aristas para guardar la información del tiempo, donde la arista en la posición 1 representa la columna 1 de la matriz.

- La representación por medio de relaciones puede mantener en cada relación de los nodos a otros nodos el tiempo de viaje como un atributo adicional o la relación a un objeto arista que se encarga de almacenar el tiempo y conectar los nodos.
- La lista de adyacencia con pesos se adapta la representación original para incluir un par vértice y peso de la arista.

Ahora bien, también es importante analizar cuáles algoritmos son factibles a la hora de solucionar el problema. Esto teniendo en cuenta la restricción de que tiene que haber al menos  $\frac{2}{3}$  de grupos utilizados (búsqueda, caminos mínimos o árboles de recubrimiento mínimo) aunque todos los algoritmos deban estar implementados.

- Dijkstra (sistemas más cercanos y cálculo de la mejor ruta)
- BFS (búsqueda de un sistema específico)
- DFS (Búsqueda normal)

#### **Fase 5: Selección y evaluación de la mejor solución**

Al tener esta problemática tan importante, se deben de utilizar estructuras que sean totalmente eficaces, y que al mismo tiempo no sean tan complejas de implementar. Es por esa razón, que la resolución más emblemática que se encontró fueron los grafos.

Para seleccionar, evaluar y desarrollar las mejores soluciones, en una tabla que se muestra a continuación se muestra cada una de las representaciones posibles de grafos, junto a las complejidades, desventajas y ventajas respectivas:

| Representación       | Complejidad de implementación (1-5) Muy fácil a muy difícil | Desventajas   | Ventajas  |
|----------------------|---|---|---|
| Matriz de pesos      | 3   | Ocupa el máximo espacio posible sin importar la cantidad de aristas, se requiere una lista de vértices aparte.  | Uso amplio; no requiere lista de aristas; fácil adaptación a algoritmos.                                      |
| Listas de adyacencia | 3   | Cada vértice puede aparecer más de una vez; cuando se tienen muchas aristas ocupa más espacio que la matriz de pesos; la complejidad temporal aumenta | Se tiene una sola lista con vértices y pesos de las aristas; adaptación inmediata a la mayoría de algoritmos. |

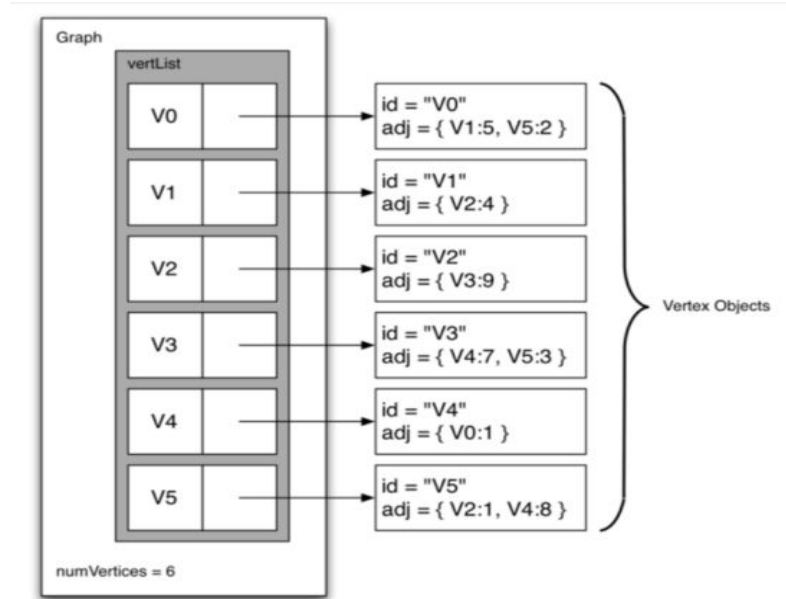
|                      |   |   |  |
|----------------------|---|---|--|
|                      |   | drásticamente en grafos densos.   |  |
| Listas de incidencia | 4 | Se tarda en encontrar el vértice correspondiente a una arista. La lista crece mucho en tamaño (grafos densos).  | Una sola lista con aristas; implementación más sencilla de un grafo. |
| Matriz de incidencia | 4 | El peso de la arista es duplicado; se requiere una lista de vértices y posiblemente una de aristas.   | -  |
| Por relaciones       | 5 | Puede ser complejo de adaptar a los algoritmos; no es tan sencillo y eficiente el añadir vértices y aristas; se debe tener un arreglo de vértices ya que no hay una raíz. | Representación más intuitiva.  |

Pero, como se desea desarrollar un software de calidad, y sin ningún tipo de error o inconveniente secundario, con la implementación de grafos se piensa representarlos de dos maneras:

1. Matriz de pesos: En la siguiente imagen se muestra una imagen asociada al dato anteriormente mencionado.

$$W = \begin{array}{c|cccc|c} & A & B & C & D & \\ \hline A & \infty & 2 & 8 & \infty & A \\ B & \infty & \infty & \infty & 21 & B \\ C & \infty & \infty & \infty & 4 & C \\ D & \infty & \infty & \infty & \infty & D \end{array} \quad W = \begin{array}{c|cccc|c} & A & B & C & D & \\ \hline A & \infty & 3 & 26 & 11 & A \\ B & \infty & \infty & \infty & \infty & B \\ C & \infty & \infty & \infty & \infty & C \\ D & \infty & \infty & \infty & \infty & D \end{array}$$

2. Lista de adyacencia: En la siguiente imagen se muestra una imagen asociada al dato anteriormente mencionado.



### Fase 6: Preparación de informes y especificaciones

- Diseño de casos de pruebas

#### Collections

#### Graph

| Name     | Class     | Scenario   |
|----------|-----------|--|
| setup1() | TestGraph | Creates an empty weighted undirected graph.  |
| setup2() | TestGraph | <p>Creates a weighted and undirected graph with vertex from 1 to 20, with the following values in the following positions:</p> <ul style="list-style-type: none"> <li>• (0,1) = 4</li> <li>• (5, 7) = 69</li> <li>• (18,11) = 33</li> <li>• (6,12) = 20</li> <li>• (20,20) = 55</li> <li>• (17,19) = 82</li> <li>• (3,16) = 8</li> <li>• (2,5) = 76</li> <li>• (11,14) = 13</li> <li>• (4,8) = 45</li> </ul> |
| setup3() | TestGraph | One thousand vertices with values from 1 to 1000.  |

| Class     | Method              | Setup    | Input   | Output   |
|-----------|---------------------|----------|---|--|
| TestGraph | testAddVertex1()    | setup1() | x = 21  | True, the element has been added correctly.                                |
| TestGraph | testAddVertex2()    | setup2() | x = 64  | True, the element has been added correctly.                                |
| TestGraph | testAddVertex3()    | setup1() | One thousand vertices with values from 0 to 1000. | True, all the elements have been added successfully.                       |
| TestGraph | testAddEdge1()      | setup2() | v1 = 2, v2 = 15, w = 67                           | True, the element has been added to the graph. The weight in (2,15) is 67. |
| TestGraph | testAddEdge2()      | setup2() | v1 = 3, v2 = 16, w = 9                            | False because an edge between those 2 nodes already exists.                |
| TestGraph | testAddEdge3()      | setup3() | v1 = 13, v2 = 35                                  | True, the element has been added to the graph.                             |
| TestGraph | testIsEmpty1()      | setup1() | -   | True, because there aren't elements in the graph.                          |
| TestGraph | testIsEmpty2()      | setup2() | -   | False, because there are elements in the graph.                            |
| TestGraph | testIsEmpty3()      | setup3() | -   | False, because there are elements in the graph.                            |
| TestGraph | testRemoveVertex1() | setup1() | v = 70  | False, because the graph is empty.   |
| TestGraph | testRemoveVertex2() | setup2() | v = 20  | True. The element has been removed for the graph.                          |
| TestGraph | testRemoveVertex3() | setup3() | v = 1459  | False, because the element does not exist in the graph.                    |
| TestGraph | testRemoveEdge1()   | setup2() | v1 = 18, v2 = 11                                  | True. The edge has been removed for the graph.                             |

|           |                   |          |                   |  |
|-----------|-------------------|----------|-------------------|--|
| TestGraph | testRemoveEdge2() | setup2() | v1 = 7, v2 = 1    | False, because there is not an edge between those two nodes. |
| TestGraph | testRemoveEdge3() | setup3() | v1 = 50, v3 = 700 | False, because there is not an edge between those two nodes. |

- [Diagrama de clases](#)
- [Diagrama de pruebas](#)

### Fase 7: Implementación del diseño

[Ver el repositorio en GitHub.](#)