

Instituto Tecnológico de Costa Rica

Campus San Carlos

Escuela de ingeniería en Computación

Análisis de Algoritmos

Segundo Proyecto Programado

Profesor:

Ana Lorena Valerio Solís

Alumnos:

Johan Efrén Rodríguez Salas

Lamberth Aarón González Araya

I Semestre 2023

Tabla de contenidos

[Introducción 3](#_Toc136738508)

[Análisis del problema 3](#_Toc136738509)

[Solución del problema 4](#_Toc136738510)

[Datos utilizados 4](#_Toc136738511)

[Estructuras utilizadas 5](#_Toc136738512)

[ParsedData 5](#_Toc136738513)

[Professor 5](#_Toc136738514)

[Course 6](#_Toc136738515)

[Allele 6](#_Toc136738516)

[Chromosome 7](#_Toc136738517)

[Matriz 7](#_Toc136738518)

[Estrategia Genética 8](#_Toc136738519)

[Estrategia Backtracking 10](#_Toc136738520)

[Estrategia Programación Dinámica 10](#_Toc136738521)

[Análisis de resultados 10](#_Toc136738522)

[Conclusiones 10](#_Toc136738523)

[Recomendaciones 10](#_Toc136738524)

[Bitácoras y minutas 10](#_Toc136738525)

[Referencias 10](#_Toc136738526)

# Introducción

El presente proyecto tiene como objetivo desarrollar un programa en el lenguaje Java que utilice las ventajas de la programación orientada a objetos (POO) para abordar el problema de asignación de cursos a profesores en una universidad. El propósito principal es implementar tres estrategias de diseño diferentes (algoritmos genéticos, programación dinámica y backtracking) para resolver dicho problema y determinar la eficiencia de cada estrategia mediante mediciones empíricas y analíticas.

A lo largo de este proyecto se define el tipo de datos de prueba, como se implementa cada estrategia, que tipos de estructuras son utilizadas, cuáles son las medidas empíricas resultantes por cada estrategia de programación y los respectivos resultados correspondientes a cada estrategia, dando como conclusión los pros y contras de cada estrategia a la hora de solucionar la tarea designada en el proyecto.

# Análisis del problema

El problema planteado consiste en la asignación óptima de cursos a profesores en una universidad, considerando los datos históricos de calificación de cada profesor en los cursos disponibles. El objetivo es encontrar la mejor asignación que maximice la calidad de la educación y los servicios ofrecidos a los estudiantes. Para lograr esto, se deben tener en cuenta las siguientes consideraciones y restricciones.

En primer lugar, se debe asegurar que la carga de trabajo de cada profesor no exceda un tiempo semanal determinado, el cual equivale a la capacidad de impartir un máximo de 4 cursos. Esto implica que la asignación de cursos debe distribuirse de manera equitativa entre los profesores para evitar la sobrecarga de trabajo.

En segundo lugar, es necesario garantizar que todos los cursos disponibles sean asignados una única vez. Esto significa que no puede haber cursos sin asignar, ya que eso afectaría negativamente la calidad de la educación ofrecida por la universidad.

Por último, es importante tener en cuenta la posibilidad de que algunos profesores queden sin asignación. Esto puede ocurrir si la cantidad de profesores es mayor que la cantidad de cursos disponibles. En tal caso, se debe encontrar la asignación óptima que maximice la calidad de la educación dentro de las limitaciones existentes.

En resumen, el problema de asignación de cursos a profesores presenta restricciones cruciales que deben tenerse en cuenta durante el diseño de las estrategias. Estas restricciones incluyen la limitación de carga de trabajo semanal de los profesores, la asignación única de todos los cursos y la posibilidad de que algunos profesores queden sin asignación. Las estrategias de diseño deben abordar estas restricciones para lograr una solución óptima y eficiente que maximice la calidad educativa ofrecida por la universidad.

# Solución del problema

A continuación, se definen todas las etapas realizadas durante el proyecto para la solución del problema previamente analizado, empleando las distintas estrategias de programación y el manejo de estas.

## Datos utilizados

Para abordar el problema de asignación de cursos a profesores, se utiliza una representación de los datos generados aleatoriamente en formato JSON. Estos datos incluyen el nombre del profesor, la calificación del profesor en cada curso, la cantidad total de cursos y la población deseada para esos datos, según las especificaciones del proyecto. Cabe destacar que las calificaciones son valores enteros desde 0 hasta 10, y para determinar que un profesor no es elegible para un curso determinado se le otorga una calificación de -1 en el mismo.

Los datos generados se almacenan en archivos JSON para facilitar su manejo y acceso desde el programa Java. Se crearon un total de seis archivos de datos, correspondientes a diferentes tamaños y poblaciones, siguiendo las siguientes especificaciones del proyecto:

* Tamaño 3p, 9c: Se generó un archivo llamado "data0.json" que contiene información sobre tres profesores y nueve cursos.
* Tamaño 5p, 15c: Se generó un archivo llamado "data1.json" que contiene información sobre cinco profesores y quince cursos.
* Tamaño 6p, 18c: Se generó un archivo llamado "data2.json" que contiene información sobre seis profesores y dieciocho cursos.
* Tamaño 9p, 27c: Se generó un archivo llamado "data3.json" que contiene información sobre nueve profesores y veintisiete cursos.
* Tamaño 10p, 20c: Se generó un archivo llamado "data4.json" que contiene información sobre diez profesores y veinte cursos.
* Tamaño 10p, 30c: Se generó un archivo llamado "data5.json" que contiene información sobre diez profesores y treinta cursos.

Estos archivos se encuentran ubicados en la carpeta "data" en la raíz del proyecto. Se decidió utilizar un script de Python para generar los datos, debido a su facilidad de uso y agilidad a la hora de la creación de datos de manera aleatoria. Al guardar los datos en formato JSON, se logra una estructura de datos legible y fácil de manipular desde el programa Java.

Debido a que los datos se generaron de manera aleatoria, se hicieron modificaciones posteriores para definir datos de control, los cuales no cumplen con las restricciones del proyecto, de forma que se pueda manejar errores en etapas posteriores del programa. La mayoría de estas modificaciones fueron en “data0.json” debido a que es el archivo menos extenso y más sencillo de manejar a la hora de hacer comprobaciones. Los datos de control van desde definir un conjunto de datos invalido para el cálculo, hasta abordar los puntos débiles de cada estrategia de diseño, los cuales posteriormente se menciona como se resolvieron de manera detallada.

Por último, en cuanto al manejo de estos archivos en el programa Java respecta. El archivo **LoadJson.java** se encarga de cargar archivos JSON y convertirlos en objetos de tipo Alelo para el manejo de los datos. Utiliza la biblioteca **json.simple** para analizar y extraer la información necesaria del archivo. Los objetos de tipo Alelo se almacenan en una lista llamada **alleles**, mientras que los valores del total de cursos y la población se obtienen y almacenan en variables correspondientes en un objeto de tipo **ParsedData**. En resumen, **LoadJson.java** permite cargar y convertir los datos de los archivos JSON para su posterior uso en la solución del problema de asignación de cursos a profesores.

## Estructuras utilizadas

A continuación, se definen las estructuras utilizadas para el desarrollo del proyecto, se detalla para que se usa cada una de ellas, que proporciona como utilidad al proyecto y su respectivo diagrama.

### ParsedData

La clase ParsedData es una estructura de datos utilizada para almacenar los resultados del análisis de los archivos JSON. Tiene los siguientes atributos:

* alleles (ArrayList<Allele>): Una lista de objetos Allele que representan la información de las calificaciones de los profesores en los cursos.
* courses (int): El número total de cursos disponibles.
* population (int): La cantidad de población, es decir, la cantidad de individuos en el algoritmo genético que se utilizará para resolver el problema.

La clase ParsedData también tiene un método toString() sobrescrito para proporcionar una representación en forma de cadena de la instancia de ParsedData, que muestra los valores de los atributos alleles, courses y population.

El diagrama de la estructura ParsedData se muestra a continuación:

### Professor

La clase Professor se utiliza para representar a un profesor en el algoritmo.

La clase tiene los siguientes atributos privados:

* name (String): El nombre del profesor.
* index (int): El índice del profesor en el array.

La clase Professor tiene un constructor que recibe el nombre del profesor y lo utiliza para inicializar el atributo name. Además, calcula el índice del profesor a partir del nombre.

La clase proporciona métodos para obtener y establecer el nombre del profesor (getName() y setName()), obtener el índice del profesor (getIndex()), y un método toString() que devuelve el nombre del profesor como una cadena.

El diagrama de la estructura Professor se muestra a continuación:

### Course

La clase Course se utiliza para representar un curso en el algoritmo.

La clase tiene los siguientes atributos privados:

* name (String): El nombre del curso.
* index (int): El índice del curso.

La clase Course tiene un constructor que recibe el nombre del curso y lo utiliza para inicializar el atributo name. Además, calcula el índice del curso a partir del nombre.

La clase proporciona métodos para obtener y establecer el nombre del curso (getName() y setName()), obtener el índice del curso (getIndex()), y un método toString() que devuelve el nombre del curso como una cadena.

El diagrama de la estructura Course se muestra a continuación:

### Allele

La clase Allele representa un alelo en el algoritmo genético. Un alelo es una combinación de un profesor, un curso y una calificación.

La clase tiene los siguientes atributos privados:

* professor (Professor): El objeto Professor que representa al profesor asociado al alelo.
* course (Course): El objeto Course que representa al curso asociado al alelo.
* grade (int): La calificación del profesor para el curso.

La clase Allele tiene un constructor que recibe un profesor, un curso y una calificación para inicializar los atributos.

Además, la clase Allele proporciona métodos para acceder y modificar los atributos, como getProfessor(), setProfessor(), getCourse(), setCourse(), getGrade(), setGrade(), así como métodos getAllele() y setAllele() que permiten obtener y establecer el alelo completo.

La clase también tiene un método toString() que devuelve una representación en forma de cadena del alelo, mostrando el nombre del profesor, el nombre del curso y la calificación.

El diagrama de la estructura Allele se muestra a continuación:

### Chromosome

La clase Chromosome representa un cromosoma en el algoritmo genético.

La clase tiene los siguientes atributos privados:

* alleles (ArrayList<Allele>): Una lista de objetos Allele que representan los alelos del cromosoma.

La clase Chromosome tiene varios constructores, uno que recibe un ArrayList<Allele> y otro que recibe una List<Allele>, ambos utilizados para inicializar los alelos del cromosoma.

La clase proporciona métodos para generar un cromosoma aleatorio (generateRandom()), verificar si un alelo es óptimo para la mutación (isOptimalAllele()), verificar si un alelo es válido para un curso (isValidCourse()), verificar si un alelo es válido para un profesor (isValidProfessor()), calcular la función de aptitud (fitness()), obtener y establecer los alelos del cromosoma (getAlleles() y setAlleles()), y un método toString() que devuelve una representación en forma de cadena del cromosoma.

El diagrama de la estructura Chromosome se muestra a continuación:

### Matriz

La matriz se utiliza como entrada para la estrategia de programación dinámica y backtracking en el algoritmo. La matriz representa la asignación de notas a los profesores para cada curso. Las filas de la matriz representan a los profesores, mientras que las columnas representan a los cursos. La intersección entre una fila y una columna contiene la nota asignada al profesor en el curso correspondiente.

El diagrama para la matriz se muestra a continuación:

A picture containing black, darkness

Description automatically generated

Figura: Diagrama de la matriz.

Donde:

* Nota\_{i,j} representa la nota asignada al profesor i en el curso j.
* m es el número de profesores (filas de la matriz).
* n es el número de cursos (columnas de la matriz).

## Estrategia Genética

El objetivo de la estrategia genética es encontrar la mejor solución posible a través de la evolución de una población de cromosomas. A continuación, se detalla los pasos realizados por el algoritmo para alcanzar este objetivo, además de un diagrama de flujo.

En primer lugar, se define una clase llamada Genetic que contiene los métodos y variables necesarios para ejecutar el algoritmo genético.

La función principal es runGenetic, que recibe un objeto Genetic y un nombre de archivo como parámetros. Esta función carga los datos del archivo y configura los parámetros iniciales del algoritmo genético, como el tamaño de la población, el número total de profesores y cursos, y el número total de generaciones. Luego, ejecuta el algoritmo genético iterativamente durante el número especificado de generaciones.

Cabe destacar que la cantidad total de generaciones que realiza el programa es igual al valor de población extraído de cada archivo JSON.

Dentro del bucle principal del algoritmo genético, se realiza la selección, el cruce, la mutación y la evaluación de la aptitud para cada par de cromosomas seleccionados. La selección se realiza mediante el método selection, que elige un cromosoma con respecto a un índice designado previamente en la población general.

El cruce se realiza mediante el método crossover, que implementa el cruce de dos cromosomas utilizando el operador de cruce de puntos de cruce mixtos (PMX). Los puntos de cruce se definen como pmxPoint1 y pmxPoint2 y se seleccionan según la proporción del total de cursos, en caso de que el cruce retorne un resultado invalido, se continua con el siguiente par de cromosomas.

Los motivos para que el método crossover retorne un resultado invalido se debe a que a la hora de hacer el cruce no existe una manera de realizarlo sin que se incumplan las restricciones prestablecidas para el proyecto y las distintas estrategias de diseño.

La mutación se realiza mediante el método mutation, el cual aplica una mutación aleatoria a un cromosoma. En este caso, la mutación se lleva a cabo seleccionando un alelo aleatorio del conjunto inicial de alelos. Sin embargo, se aplican restricciones para garantizar que la mutación sea válida y mejore la aptitud del cromosoma.

Estas restricciones son las siguientes: el alelo obtenido no puede tener una nota igual a -1 (lo cual indica que el profesor no puede dar ese curso) y tampoco puede pertenecer a un profesor que ya tenga asignados 4 cursos. Estas restricciones aseguran que no se exceda el límite de cursos por profesor y que no se agreguen cursos inválidos.

Si el alelo seleccionado cumple con las restricciones, se introduce en el cromosoma reemplazando al curso con el mismo nombre. Luego, se calcula el fitness tanto para el cromosoma original como para el cromosoma posterior a la mutación.

Para decidir si se mantiene la mutación o no, se compara la aptitud de ambos cromosomas. Si el cromosoma mutado tiene una aptitud superior al cromosoma original, se conserva la mutación. De lo contrario, se descarta la mutación y se mantiene el cromosoma original.

La función de aptitud o fitness calcula un valor numérico que representa la calidad de un cromosoma. En este caso, se calcula sumando todas las notas de la combinación de cursos. Un valor mayor indica un cromosoma mejor. La función de aptitud también se utiliza en el algoritmo dinámico para encontrar la mejor solución, ya que no se tiene un objetivo específico definido.

Después de la mutación y la evaluación de la aptitud, se actualiza la población reemplazando los cromosomas menos aptos con los descendientes mutados si su aptitud es mayor. Además, se mantiene una lista de los 5 mejores cromosomas (diferentes entre sí) en la variable bestChromosomes, para mostrar los mejores cromosomas generados posterior a la ejecución del algoritmo.

Finalmente, se selecciona el mejor cromosoma de la población y se almacena en la variable result. Se realiza una comparación de la aptitud de cada cromosoma con el cromosoma anterior y se actualiza el result si se encuentra un cromosoma con una aptitud superior.

En resumen, la estrategia genética implementada en este proyecto utiliza el operador de cruce PMX, una mutación aleatoria y la evaluación de la aptitud para evolucionar una población de cromosomas y encontrar la mejor solución al problema de optimización.

## Estrategia Backtracking

## Estrategia Programación Dinámica

* Realice un diagrama de flujo o seudocódigo donde explique la lógica que desarrollo, para aplicar cada uno de los algoritmos.
* Describa la estrategia de la programación dinámica, si los datos se ordenan por curso o profesor de forma descendente o ascendente, que criterio se utiliza para la selección o asignación de cada curso en cada etapa de avance…
* Describa la estrategia del backtracking, como va construyendo la respuesta en cada llamada recursiva, como descarta las posibilidades que no cumplen con las restricciones del problema. Para que tamaños de problemas es posible dar respuesta.

# Análisis de resultados

* Resultados finales, indique que partes están completas, cuales defectuosos, y cuáles no se realizaron y el porqué.
* Adjunte las tablas de todas las mediciones realizadas a sus algoritmos.
* Cálculos realizados.
* Clasificación en notación O grande.
* Gráfico y su respectivo análisis donde compare los comportamientos de los algoritmos.
* Análisis de las tablas de mediciones (empíricas y analíticas).

# Conclusiones

* Según la medición realizada indique cuál estrategia es más eficiente para resolver este problema, justifique su aseveración.
* Responda la siguiente pregunta ¿Conforme crece la talla cuál algoritmo se va haciendo más eficiente?

# Recomendaciones

* Con respecto al alcance del proyecto.

# Bitácoras y minutas

# Referencias

GeeksforGeeks. (2023b). Backtracking Algorithms. *GeeksforGeeks*. <https://www.geeksforgeeks.org/backtracking-algorithms/>

GeeksforGeeks. (2023). Dynamic Programming. *GeeksforGeeks*. <https://www.geeksforgeeks.org/dynamic-programming/>

GeeksforGeeks. (2023a). Genetic Algorithms. *GeeksforGeeks*. <https://www.geeksforgeeks.org/genetic-algorithms/>

OpenAI. (2023, May). ChatGPT Conversation.