

# Algoritmo Genético

## Calendario Generativo

# Reporte

## Inteligencia Artificial

**Autor:** Adolfo González Juárez.  
**Docente:** Dr. Carlos Alberto Díaz Hernández.

15 de marzo de 2024

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Análisis del Problema</b>	<b>3</b>
2.1. Problemática . . . . .	3
<b>3. Resolución de problemas</b>	<b>3</b>
3.1. Captura de materias . . . . .	4
3.2. Filtrado de materias . . . . .	4
3.2.1. Seriación . . . . .	5
3.2.2. Delimitación de materias . . . . .	6
3.3. Cuantificación de los datos . . . . .	8
3.3.1. Identificador . . . . .	8
3.3.2. Umbral de estrés . . . . .	9
3.3.3. Umbral de relevancia . . . . .	9
3.4. Procesamiento de la información . . . . .	10

## Índice de figuras

1. Diagrama del funcionamiento general del algoritmo. . . . .	3
2. Captura de pantalla de la sección de selección de materias cursadas, obtenida del maquetado en Figma. . . . .	4
3. Mapa curricular de la carrera de Ingeniería en Desarrollo de Software del año 2017. . . . .	5
4. Mapa curricular simplificado de ejemplo: Caso 1. . . . .	6
5. Mapa curricular simplificado de ejemplo: Caso 2. . . . .	7
6. Ejemplo ilustrativo de cálculo de relevancia. . . . .	9

## Lista de ecuaciones

1. Arreglo de materias sin procesar. . . . .	4
2. Atributos de las asignaturas sin procesar. . . . .	5
3. Condición explícita de materias disponibles. . . . .	7
4. Definición de materia apta para procesamiento. . . . .	8
5. Definiciones de datos de entrada. . . . .	8

## **1. Introducción**

El presente documento tiene como objetivo presentar un algoritmo genético diseñado para abordar una problemática común en el ámbito estudiantil: la selección óptima de materias para un estudiante en un estado de irregularidad académica. Esta herramienta de software propone una solución que determina la distribución más adecuada de materias teniendo en cuenta las preferencias del estudiante, así como la secuencia recomendada dentro del plan de estudios y el progreso académico del mismo. La exposición detallada del análisis, la metodología empleada y el desarrollo del software facilita la comprensión de su funcionamiento y su aplicación práctica en el contexto educativo.

## 2. Análisis del Problema

Para abordar la resolución de la problemática planteada, se llevó a cabo una exhaustiva identificación de los elementos pertinentes que influyen en el funcionamiento del algoritmo, así como su efectiva traducción para su implementación en un algoritmo genético.

### 2.1. Problemática

El núcleo del problema radica en la necesidad de planificar el horario académico del próximo cuatrimestre para un estudiante, basándose en su historial académico, el cual comprende un conjunto de materias que ha cursado o reprobado. Este desafío implica la evaluación de cada materia según su importancia relativa, el respeto a las preferencias del estudiante y la evitación de solapamientos de horarios entre las asignaturas. El funcionamiento general del algoritmo puede esquematizarse de la siguiente manera:

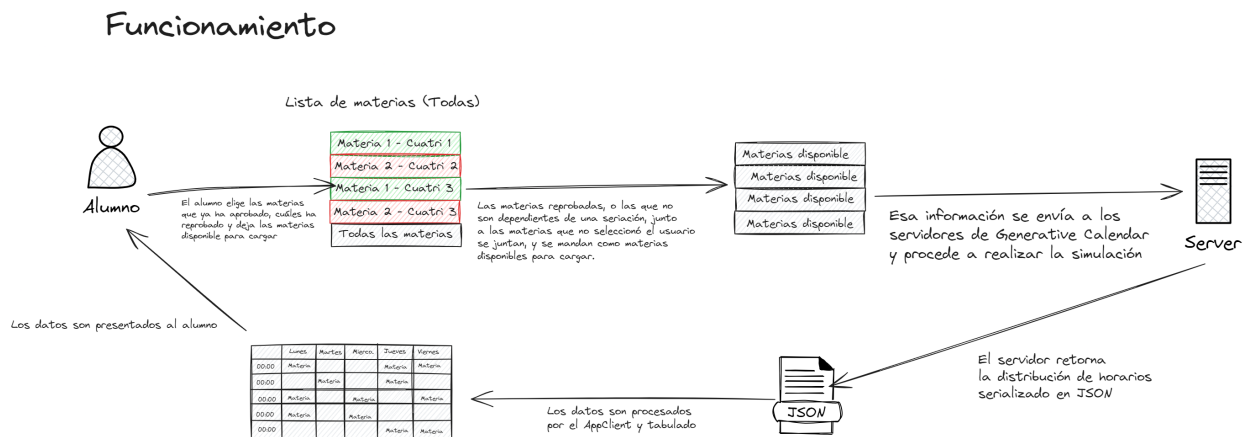


Figura 1: Diagrama del funcionamiento general del algoritmo.

## 3. Resolución de problemas

En esta sección se describe el enfoque adoptado para abordar el problema y alcanzar los objetivos del programa. La resolución del problema se divide en los siguientes puntos, que se detallarán a través de las diferentes subsecciones del documento.

1. Captura de materias. 3.1
2. Filtrado de materias. 3.2
3. Cuantificación de los datos. 3.3
4. Procesamiento de la información. 3.4

### 3.1. Captura de materias

La primera etapa del proceso del algoritmo se centra en la captura de las materias cursadas por el alumno, lo que refleja su estado académico actual. En esta fase, el alumno seleccionará manualmente las materias que ha aprobado y las que ha reprobado, dejando libres aquellas que aún no ha cursado. Esta acción proporciona al sistema un contexto claro sobre el estado y el progreso del usuario. Además, la implementación de la seriación de materias, tema que se abordará en la siguiente sección, facilitará el completado y filtrado de las materias.

Selecciona tu estado actual

CUATRIMESTRE 1	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7	CUATRIMESTRE 2	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7	CUATRIMESTRE 3	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7	CUATRIMESTRE 4	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7
CUATRIMESTRE 5	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7	CUATRIMESTRE 6	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7	CUATRIMESTRE 7	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7	CUATRIMESTRE 8	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7
CUATRIMESTRE 9	MATERIA 1 MATERIA 2 MATERIA 3 MATERIA 4 MATERIA 5 MATERIA 6 MATERIA 7						

Figura 2: Captura de pantalla de la sección de selección de materias cursadas, obtenida del maquetado en Figma.

### 3.2. Filtrado de materias

El proceso de filtrado de materias tiene como objetivo eliminar aquellas materias que, debido a restricciones de seriación, no deben considerarse al generar el nuevo horario académico. Estas materias, denominadas "posibles materias de carga", son aquellas que el algoritmo debe tomar en cuenta al generar el horario. Es importante destacar que el Algoritmo Genético (AG) procesa únicamente las posibles combinaciones de materias, sin validar directamente la existencia de dependencias entre ellas.

Una vez que el usuario ha seleccionado las materias aprobadas, estas se guardan en un arreglo de materias  $M$ , donde se almacenan todas las materias iniciales  $N$  etiquetadas según la selección del alumno. De esta forma,  $M$  se define como:

$$M : \{N_1, N_2, \dots, N_i\} \quad (1)$$

Donde  $N_i$  es la materia alojada en la posición  $i$  del arreglo  $M$ , y se podría definir a  $N_i$  de la siguiente manera:

$$N_i : (\text{id}_i, \text{status}_i, [\text{dependencias}]_i, [\text{dependientes}]_i, \text{cuatrimestre}_i) \quad (2)$$

Donde cada atributo de los objetos materia obedece las siguientes definiciones:

- id: es el identificador único de la materia conformado por caracteres alfabéticos de longitud 4.
- status: es el estado que ha elegido el alumno, que puede ser: aprobado, reprobado, o sin cursar.
- [dependencias]: es un arreglo de longitud  $n$  que contiene todos los identificadores de las materias que se deben cursar antes de poder cursar la materia actual.
- [dependientes]: es un arreglo de todas las materias que dependen de la materia actual, es decir, aquellas materias que se podrán cursar una vez que la materia actual haya sido aprobada.
- cuatrimestre: corresponde a un identificador numérico del 1 al 9, que indica el cuatrimestre regular en el que debería ser cursada dicha materia.

### 3.2.1. Seriación

Es importante destacar que las materias contienen arreglos internos que hacen referencia a las materias dependientes y las materias dependencias. Esta característica se debe a la seriación, la cual dicta qué materias se podrán cursar en función del progreso del alumno.

La seriación, comúnmente conocida, se ajusta al mapa curricular de la carrera para la cual se desarrolló el algoritmo. En el caso actual, dicho mapa curricular corresponde a la carrera de **Ingeniería en Desarrollo de Software (IDS)**.

El algoritmo se basó en el mapa curricular de IDS del año 2017. Como se puede observar en la figura 3, existen materias que solo se podrán cursar una vez que todas las materias requeridas sean aprobadas.

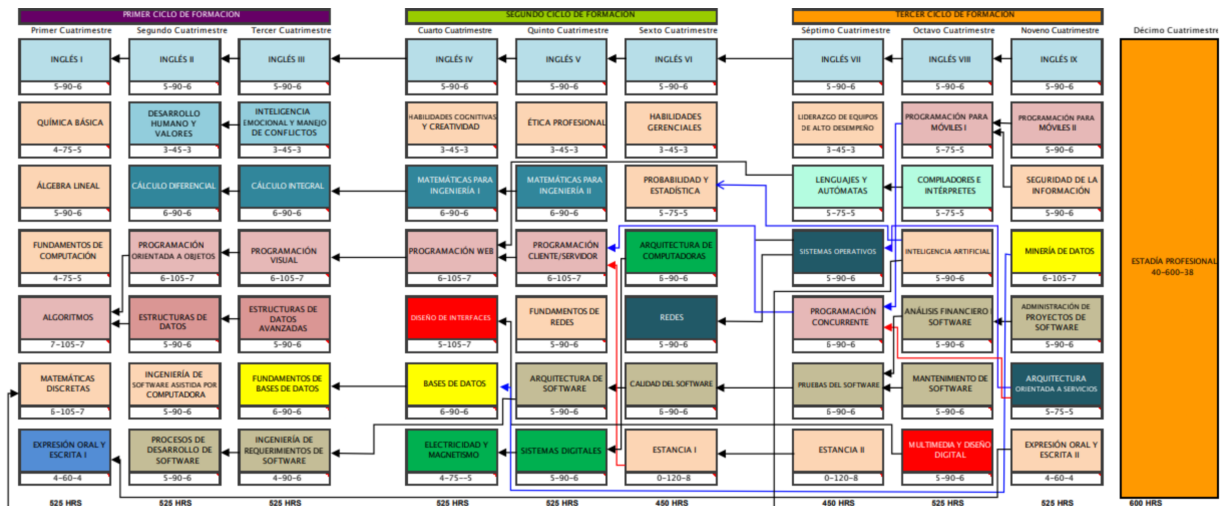


Figura 3: Mapa curricular de la carrera de Ingeniería en Desarrollo de Software del año 2017.

### 3.2.2. Delimitación de materias

Es fundamental tener en cuenta que existen materias que no pueden ser cursadas hasta que se aprueben una o más asignaturas de su línea de seriación. Por lo tanto, el programa debe eliminar de la lista de materias todas aquellas asignaturas que, según el caso de cada alumno, no sean aptas para ser cargadas debido a motivos de reprobación o porque su esquema de seriación está incompleto.

Para ilustrar este concepto, consideremos un ejemplo y simplifiquemos el mapa curricular de la siguiente manera:

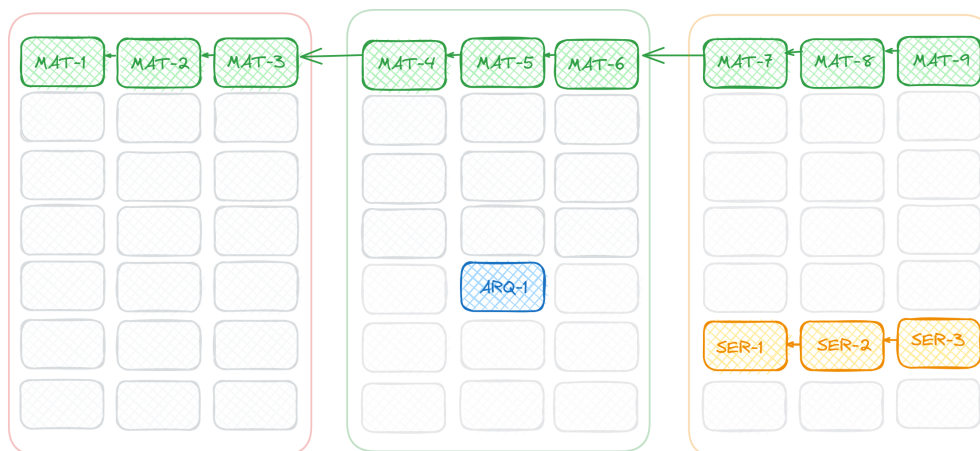


Figura 4: Mapa curricular simplificado de ejemplo: Caso 1.

Teniendo un mapa curricular como el de la figura 4, cada materia tiene una lista de asignaturas que dependen de ella. Por ejemplo:

- MAT-7 tiene una lista de materias dependientes donde se hallan MAT-8 y MAT-9.
- MAT-6 tiene una lista de materias dependientes donde se hallan MAT-7, MAT-8 y MAT-9.

Cuando una materia, por ejemplo ARQ-1, no tiene ninguna asignatura en su conjunto de materias dependientes, se le considera una materia débilmente seriada. Esto significa que ARQ-1 puede ser cursada sin tener ninguna otra materia como requisito previo. En otras palabras, no existen restricciones de seriación que limiten la capacidad de cursar ARQ-1 en función de otras asignaturas.

Ahora, si una materia es marcada como 'reprobada', todas las materias pertenecientes a la lista de materias dependientes de esa materia reprobada se marcarán como "deshabilitadas". Este proceso se aplica recursivamente a todas las materias dependientes de las materias deshabilitadas. Es decir, si una materia directamente dependiente de la materia reprobada se deshabilita, entonces todas las materias que dependen de esa materia también se deshabilitarán. La materia originalmente marcada como reprobada, sin embargo, será apta para volver a cargarse en futuros intentos académicos.

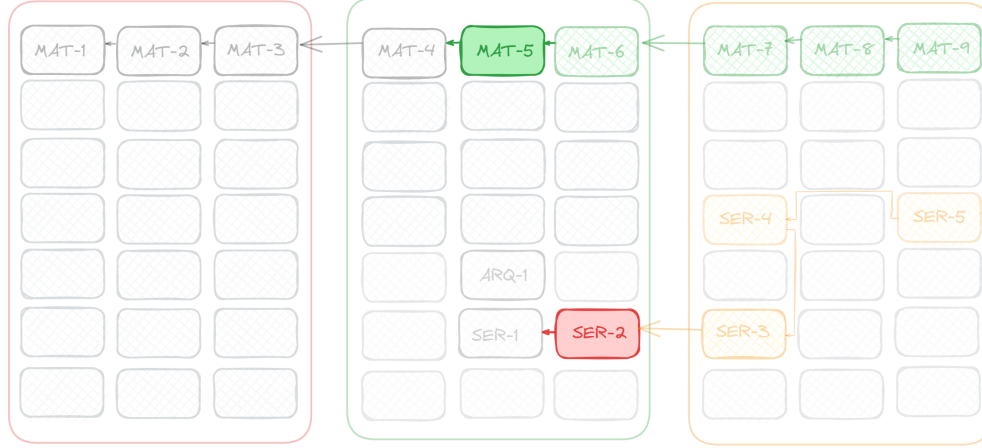


Figura 5: Mapa curricular simplificado de ejemplo: Caso 2.

De manera similar, si una materia es seleccionada por el usuario y se encuentra dentro de al menos un conjunto de materias marcadas como reprobadas o no cursadas, esta materia también será marcada como "deshabilitada". Este proceso garantiza que no se programen materias que tengan dependencias no cumplidas o que hayan sido reprobadas previamente, garantizando así la coherencia del horario académico del estudiante.

Esta delimitación se realiza mediante una serie de condiciones algebraicas que describen el proceso de manera precisa.

Se define  $M$  como una materia,  $R$  como una materia marcada como reprobada o no cursada,  $P$  como el conjunto de materias con posibilidad de carga, y  $N$  como el conjunto de materias sin posibilidad de carga.

Entonces, estas condiciones pueden expresarse de la siguiente manera:

- Si  $M$  no pertenece al conjunto de dependientes de  $R$  y no está en el conjunto  $N$ , entonces  $M$  pertenece al conjunto  $P$ .
- Si  $M$  pertenece al conjunto de dependientes de  $R$ , entonces  $M$  se coloca en el conjunto  $N$  y se elimina de  $P$ .

Esto se puede expresar en términos algebraicos de la siguiente manera:

$$\begin{aligned} (M \notin R[\text{dependientes}] \wedge (M \notin N) &\implies M \in P \\ M \in R[\text{dependientes}] &\implies (M \in N) \wedge (M \notin P) \end{aligned} \quad (3)$$

Esta expresión algebraica define de manera clara y precisa las condiciones bajo las cuales una materia puede ser considerada para su carga en función del estado de otras materias y sus dependencias.

Para ilustrar este proceso, supongamos que el usuario ha aprobado las asignaturas MAT-4 y ARQ-1, pero ha reprobado SER-2. En este caso, las únicas materias disponibles para ser cargadas serán MAT-5, ya que es la continuación de MAT-4 y SER-2, que es la materia que podrá volver a cursar en un nuevo intento académico, como se muestra en el mapa curricular simplificado de ejemplo en la Figura 5.



La materia SER-5, aunque no sea directamente dependiente de SER-2, está dentro de la línea de seriación gracias de SER-4 que depende de SER-3 que a su vez depende de SER-2, lo que significa que todas esas materias listadas deberán ser deshabilitadas.

### 3.3. Cuantificación de los datos

Otra parte importante del proceso previo al procesamiento de las materias por el algoritmo genético, es la cuantificación de los datos, es decir, añadir valores numéricos a las materias con el fin de realizar operaciones algebraicas y definir de una manera más concreta si una materia es apta para el alumno.

Tras la depuración de las materias que se especificó en la sección 3.2.2, nos queda una lista de materias que ahora sí son aptas para el procesado del algoritmo genético, sin embargo debemos añadir información adicional y remover información innecesarias para que los datos que se le provea al algoritmo genético sean los óptimos para generar un horario ideal.

A la información de las materias disponibles de carga, se le removerá las propiedades "dependientes" "dependencias", ya que sólo fueron usados en la sección 3.2.2 para eliminar las materias con seriaciones incompletas.

Por otro lado, se añade tres propiedades nuevas a la definición de la materia, que son Umbral de Estrés ( $\varphi$ ), Umbral de Relevancia ( $\mu$ ), y un conjunto de horarios disponibles ( $H$ ).

En este sentido podemos definir a una materia apta para procesamiento  $M$  de la siguiente manera:

$$M : (id, \varphi, \mu, [H]) \quad (4)$$

Estas materias pertenecen a un conjunto de materias aptas para procesamiento  $D$  que se adjunta a otro conjunto de entrada  $E$  que se enviará al servidor del algoritmo genético para su procesamiento. Este conjunto  $E$  posee propiedades adicionales que hacen referencia a las preferencias del usuario como nivel de estrés del alumno  $\varepsilon$  y cuatrimestre actual del alumno  $C$ .

Descrito explícitamente podríamos decir

$$\begin{aligned} D : \{M_1, M_2, \dots, M_i\} \\ E : (D, \varepsilon, C) \end{aligned} \quad (5)$$

Este conjunto de datos son los que serán procesados por el algoritmo genético para generar el horario más óptimo para el alumno, abordemos cada uno de los campos a continuación.

#### 3.3.1. Identificador

Es una forma única de identificar una materia dentro del plan curricular de la carrera, de esta forma podemos utilizar las materias de una forma mas precisas, este identificador es compoene únicamente de caracteres alfabéticos de longitud 4 y son irrepetibles dentro de todas las materias de la carrera.

### 3.3.2. Umbral de estrés

El umbral de estrés  $\varphi$ , es un número flotante de dos decimales de precisión que indica qué tan estresante resulta una materia según datos recopilados por usuarios. Dicho dato puede variar entre 0.00, siendo este el nivel mas bajo alcanzable, significando un materia nada estresante y 1.00, siendo el valor máximo, significando una materia altamente estresante.

### 3.3.3. Umbral de relevancia

El umbral de relevancia es un número entero mayor a 0 asociado a cada materia. Este número indica la importancia de la materia y es utilizado por la función que determina el horario final del alumno.

El cálculo del umbral se rige por las siguientes reglas:

- Cada materia tiene un valor predeterminado de 1 como umbral.
- Por cada materia dependiente, se añade 1 punto al umbral.
- Si la materia se encuentra en un cuatrimestre menor al cuatrimestre actual del alumno, se suma 1 punto al umbral por cada cuatrimestre de diferencia entre el último cuatrimestre actual y el cuatrimestre de la materia.

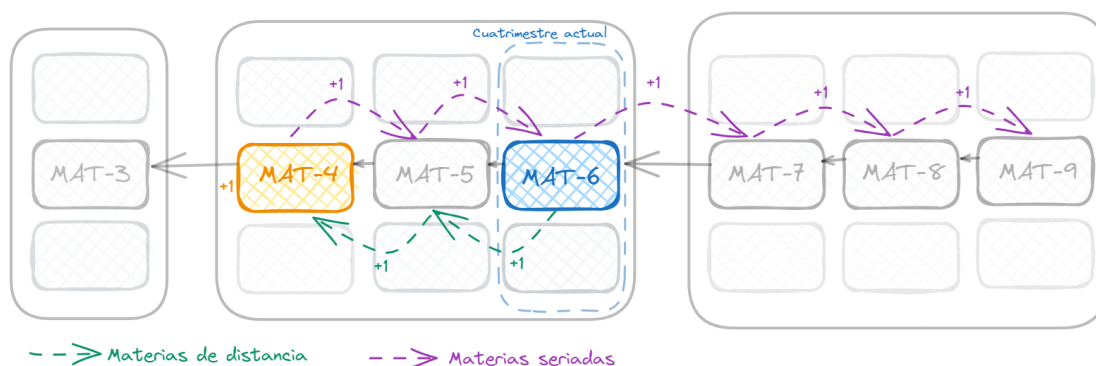


Figura 6: Ejemplo ilustrativo de cálculo de relevancia.

Por ejemplo, consideremos el caso del umbral de la materia MAT-4 de la figura 6 para un alumno en sexto semestre. En este caso, se sumará 1 punto al umbral base de 1 debido a que MAT-4 es una materia dependiente. Además, si el último cuatrimestre actual del alumno es mayor al cuatrimestre en el que se ofrece la materia MAT-4, se añadirá un punto adicional al umbral por cada cuatrimestre de diferencia entre el último cuatrimestre actual y el cuatrimestre de la materia.

Para ilustrar el cálculo del umbral de relevancia de la materia MAT-4 para un alumno en sexto semestre, consideremos los siguientes factores:

MAT-4 comienza con un umbral de 1, que es el valor predeterminado de todas las materias. Al ser dependiente de las materias MAT-5 hasta MAT-9, lo que suma un total de 5 materias dependientes, se añade 1 punto al umbral por cada una de estas materias, lo que suma un total de 5 puntos adicionales al umbral original de 1, resultando en un umbral de 6.

Además, el alumno se encuentra en el cuatrimestre 6, mientras que la carga donde se halla la materia MAT-4 está a 2 cuatrimestres de distancia. Por cada cuatrimestre de diferencia entre el último cuatrimestre

actual del alumno y el cuatrimestre de la materia, se suma un punto adicional al umbral de relevancia de la materia. Por lo tanto, se añadirán 2 puntos al umbral, lo que resulta en un umbral final de relevancia de 8 para la materia MAT-4 en este ejemplo.

Este proceso ilustra cómo se calcula el umbral de relevancia de una materia, teniendo en cuenta tanto sus dependencias como la distancia en cuatrimestres entre el último cuatrimestre actual del alumno y el cuatrimestre en el que se ofrece la materia.

### **3.4. Procesamiento de la información**