COMPUTACIÓN BIOLÓGICA

Pedro Isasi[1]

[1]Departamento de Informática
Universidad Carlos III de Madrid
Avda. de la Universidad, 30. 28911 Leganés (Madrid). Spain
email: isasi@ia.uc3m.es

Presentación

Temario

Temario

Temario

Temario

Temario

Temario

¿Qué son las restricciones?

- En algunos problemas las soluciones deben cumplir una serie de criterios para ser válidas
- En estos casos no todos los individuos codificados son aceptables
- Se dice que el problema tiene restricciones. Por ejemplo hay que asignar tareas a empleados, minimizando el tiempo:
  - Hay ciertos empleados que no pueden hacer ciertas tareas
  - Hay ciertas tareas que deben realizarse después de otras
  - Los empleados tienen horarios en los que asisten a reuniones o cursos
- No todas las asignaciones de tareas-empleados son válidas, sólo aquellas que cumplan las restricciones

PROBLEMÁTICA

- Los AGs canónicos no pueden tratar restricciones
- Se puede modificar codificando un superconjunto X del conjunto de soluciones factibles $\mathcal{X}$ ($X \supseteq \mathcal{X}$)
- El superconjunto debe:
  - Ser fácil de codificar
  - No ser mucho más grande que el conjunto de soluciones factibles, para que el AG no pierda excesivo tiempo en generar y procesar soluciones no factibles

Penalización

- Las restricciones se suelen expresar como un conjunto de inecuaciones:

$$g_i(x) \geq 0$$

- Para tratarlas se puede incluir un término de penalización
- Se puede penalizar más o menos en función del grado de incumplimiento de la restricción
- la penalización se introduce en la función de evaluación

- Matemáticamentge podría definirse un grado de incumplimiento por parte del individuo x de la restricción i-ésima como:

$$lr_i(x) = \begin{cases} -g_i(x) & Sii \quad g_i(x) < 0 \\ 0 & Sii \quad g_i(x) \geq 0 \end{cases}$$

- Este incumplimiento se introduce como penalización en la función de evaluación de cada individuo:

$$F(x) = F(x) - k \cdot \mathcal{P}(lr_1(x), ..., lr_p(x))$$

Donde: p es el número de restricciones K coeficiente de penalización $\mathcal{P}$ es una función de penalización creciente positiva dependiente del dominio

Se pueden proponer funciones de penalización variables con el tiempo:

$$\mathcal{P} = \left(\frac{k \cdot t}{m}\right)^2 \overline{F(P(t))} \sum_{i=1}^{p} lr_i(x(t))$$

Al avanzar la búsqueda la penalización se va haciendo cada vez más rígida.

Penalización: problemática

- Las funciones de penalización muy severas no son recomendables en problemas fuertemente restringidos
  - El AG gasta gasta mucho tiempo en procesar individuos no factibles
  - Cuando se encuentra un individuo factible éste puede destacar demasiado sobre los demás (convergencia prematura)
- Las funciones de penalización muy flexibles tienen el riesgo de dejar prosperar a individuos que, aún siendo no factibles, tengan una aptitud neta mayor que otros factibles
- A veces se habla de técnicas de gratificación: premian a los individuos factibles

Reparación o Corrección

Consisten en habilitar un procedimiento con el que corregir cualquier solución no factible que se genere

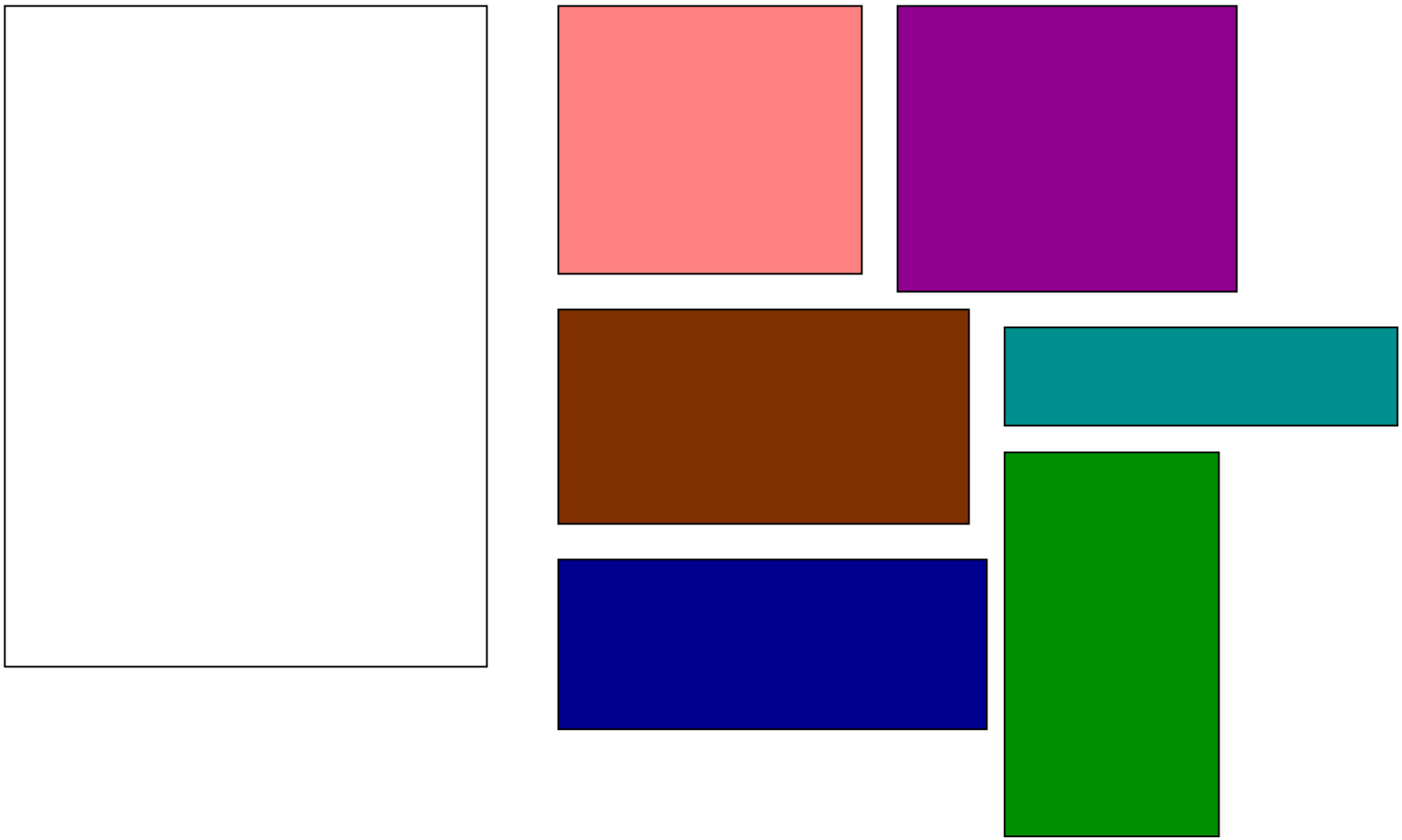
- Inconveniente: suele ser difícil encontrar un procedimiento de corrección lo suficientemente general y, si se encuentra, normalmente consume muchos recursos de computación
- Este procedimiento es necesariamente específico para cada problema

Decodificación

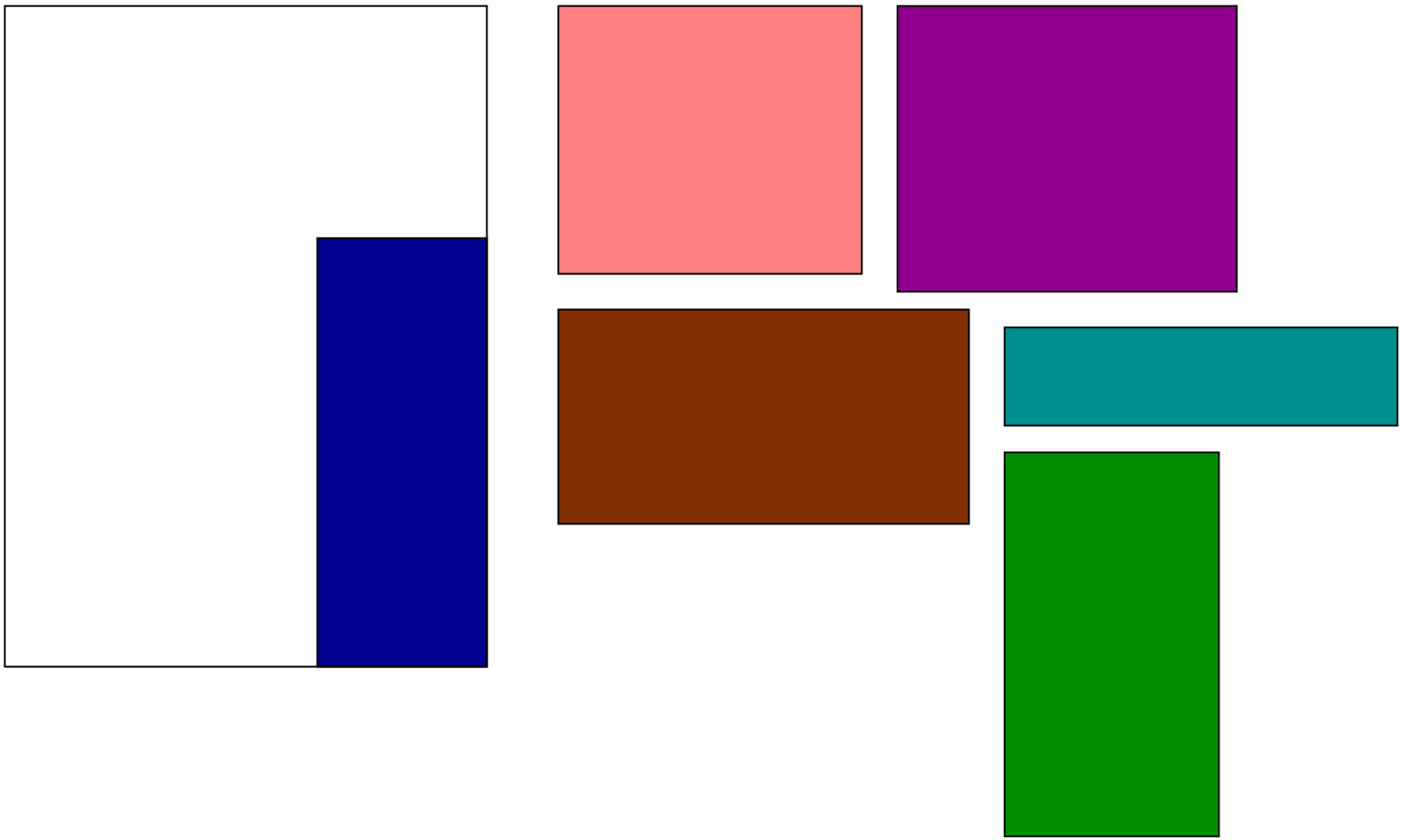
Consisten en realizar la codificación de tal modo que sea imposible generar soluciones no factibles

- Se modifica X a través de la codificación para que se aproxime lo máximo posible a $\mathcal{X}$
- Suelen requerir modificar los operadores genéticos para que conserven la factibilidad de los individuos
- Inconvenientes similares a las técnicas de reparación
- De modo indirecto todas estas técnicas están añadiendo conocimiento específico.

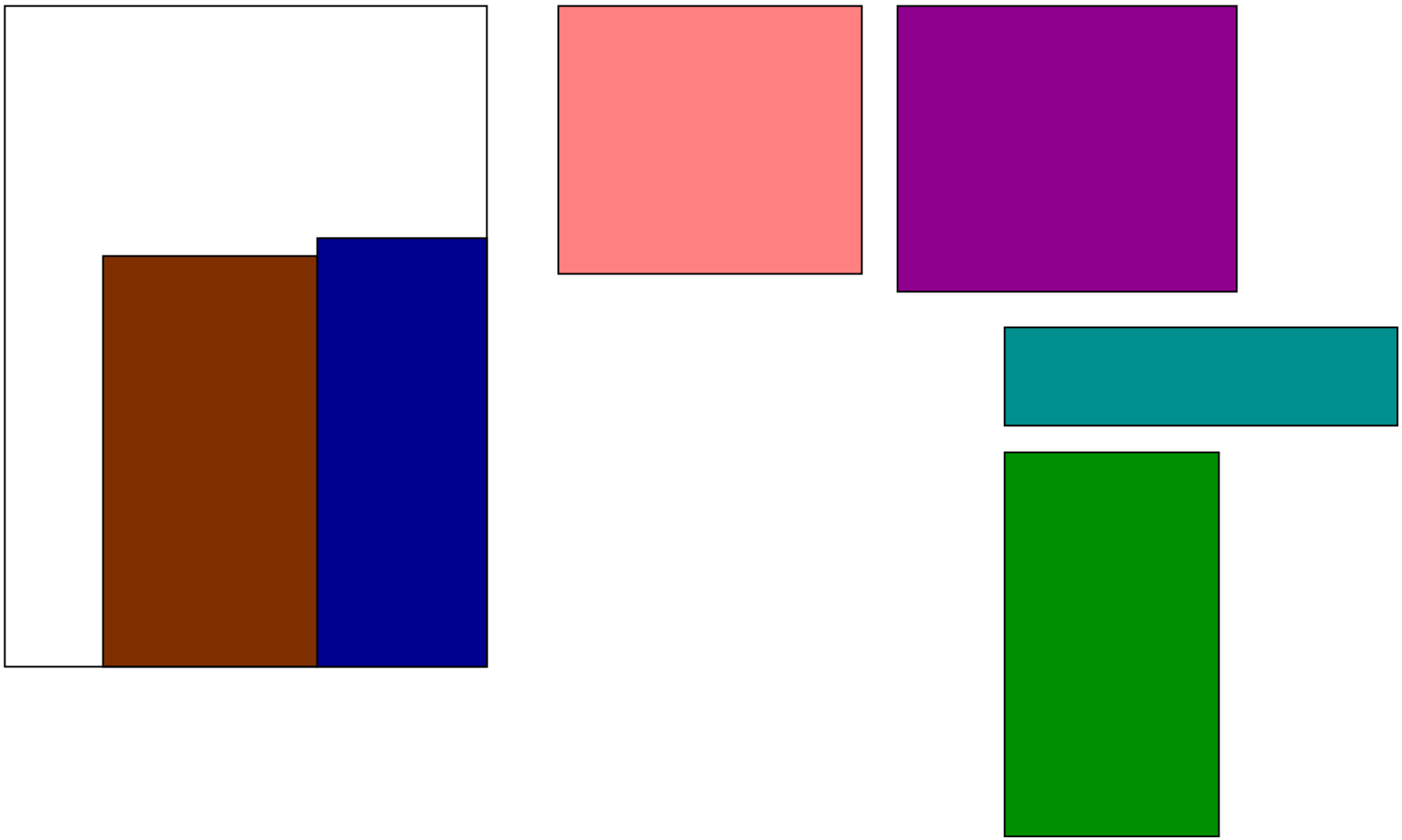
El problema de la mochila



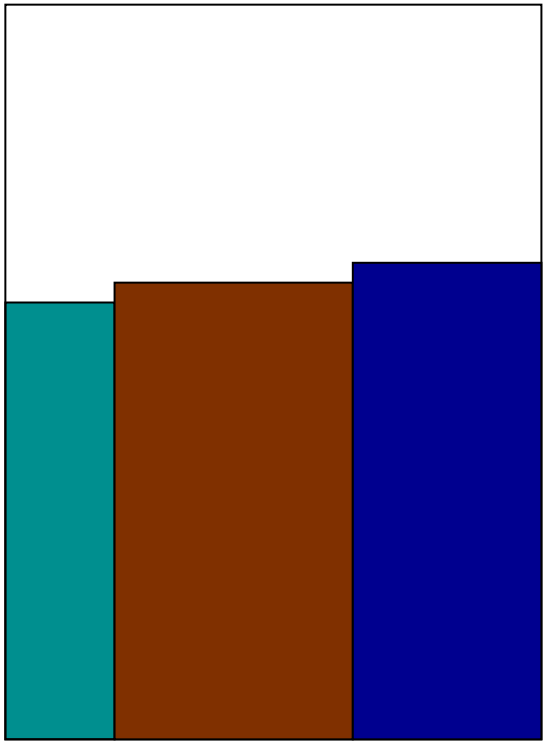
El problema de la mochila

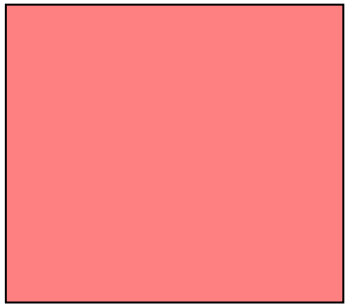


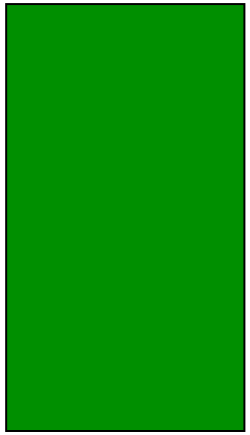
El problema de la mochila



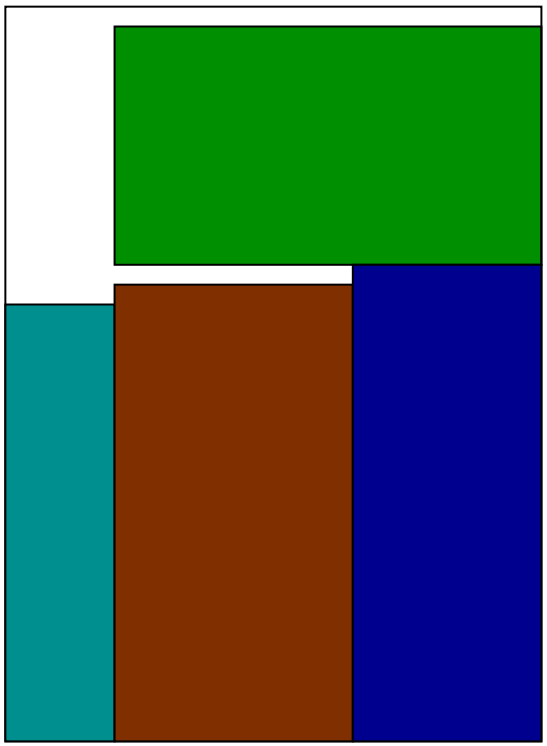
El problema de la mochila

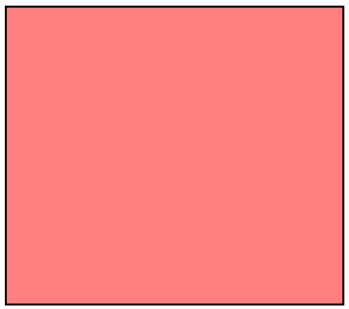




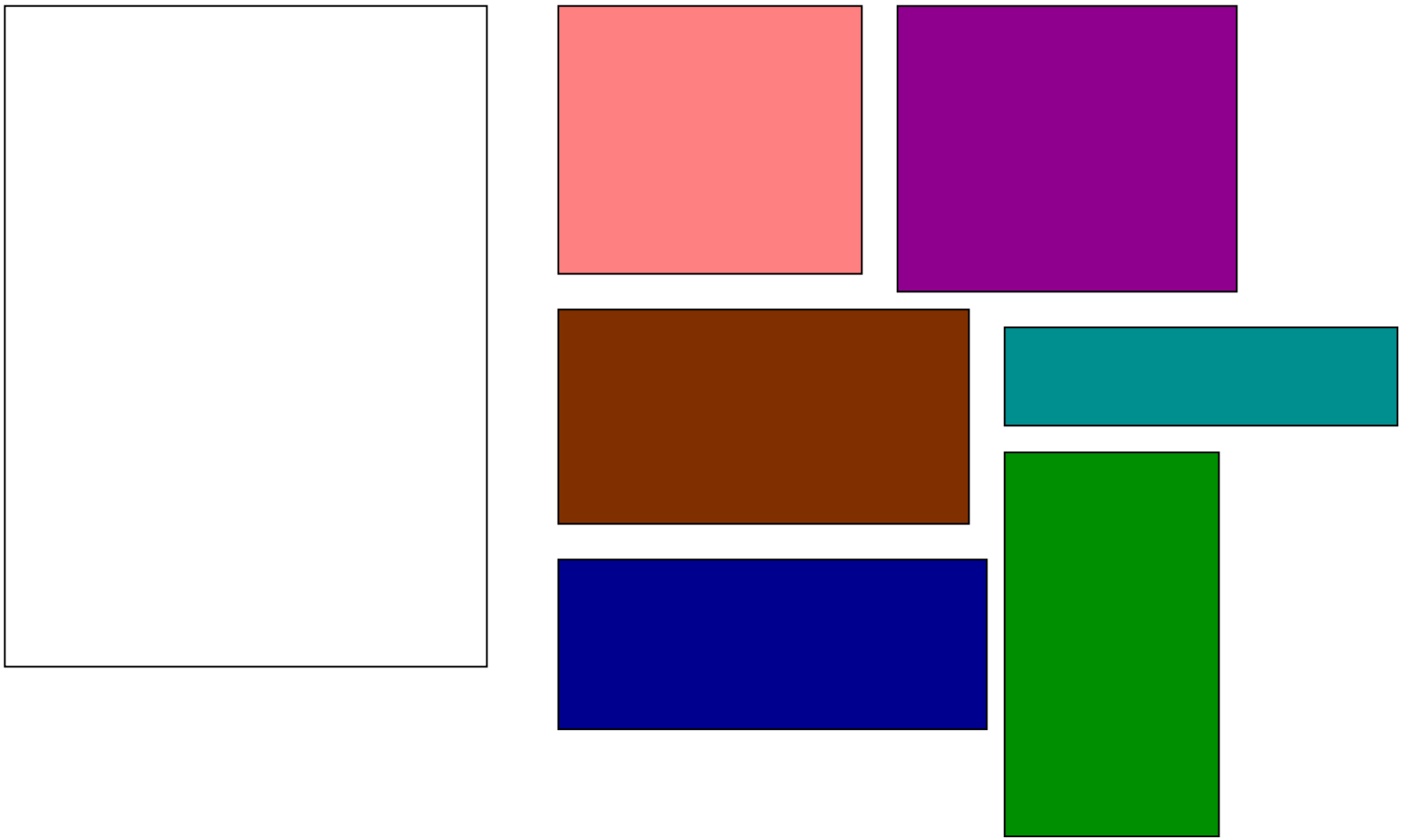


El problema de la mochila

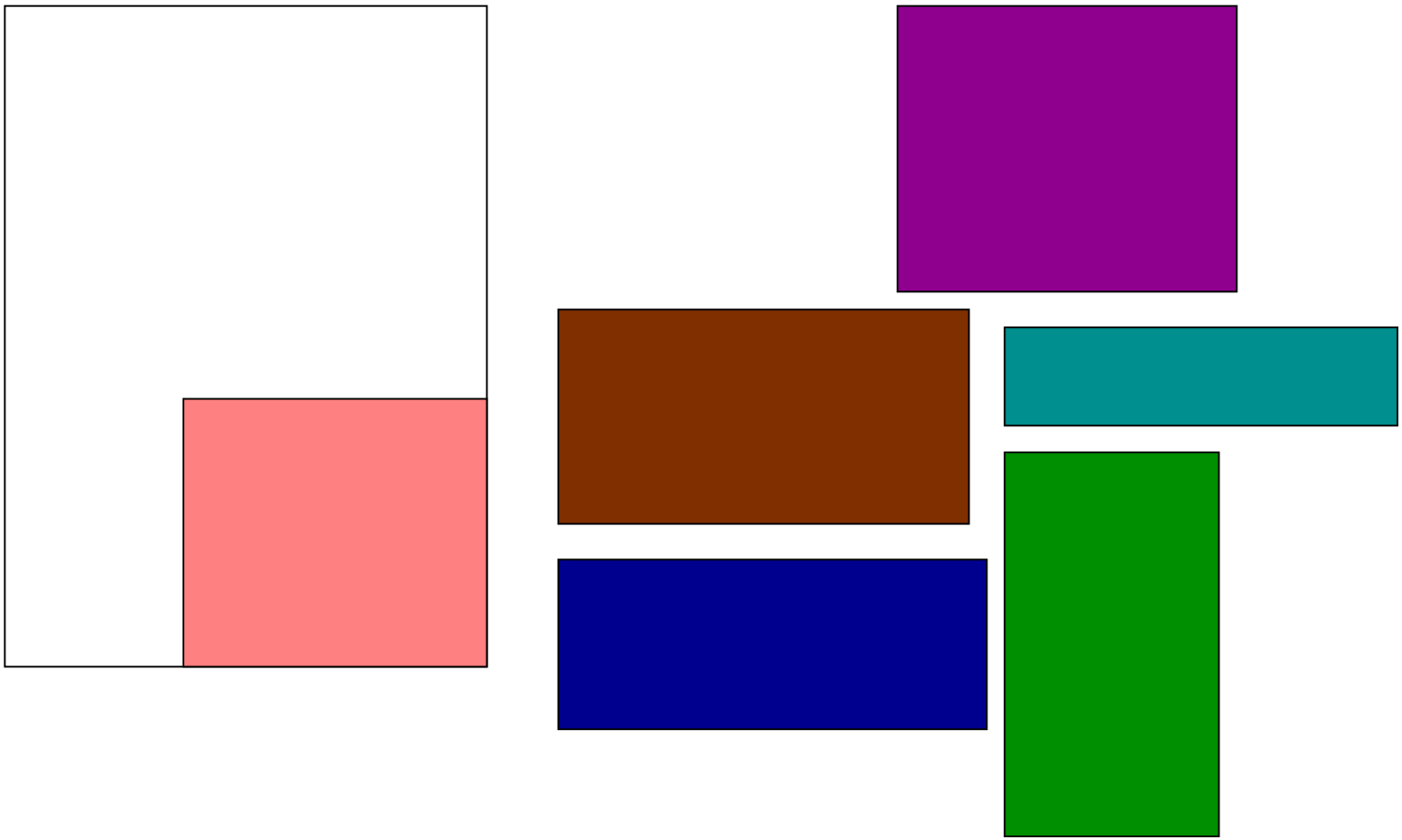




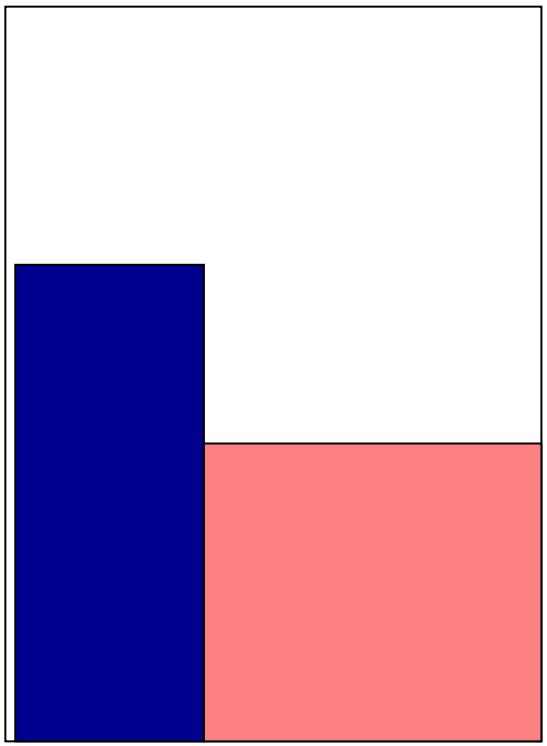
El problema de la mochila

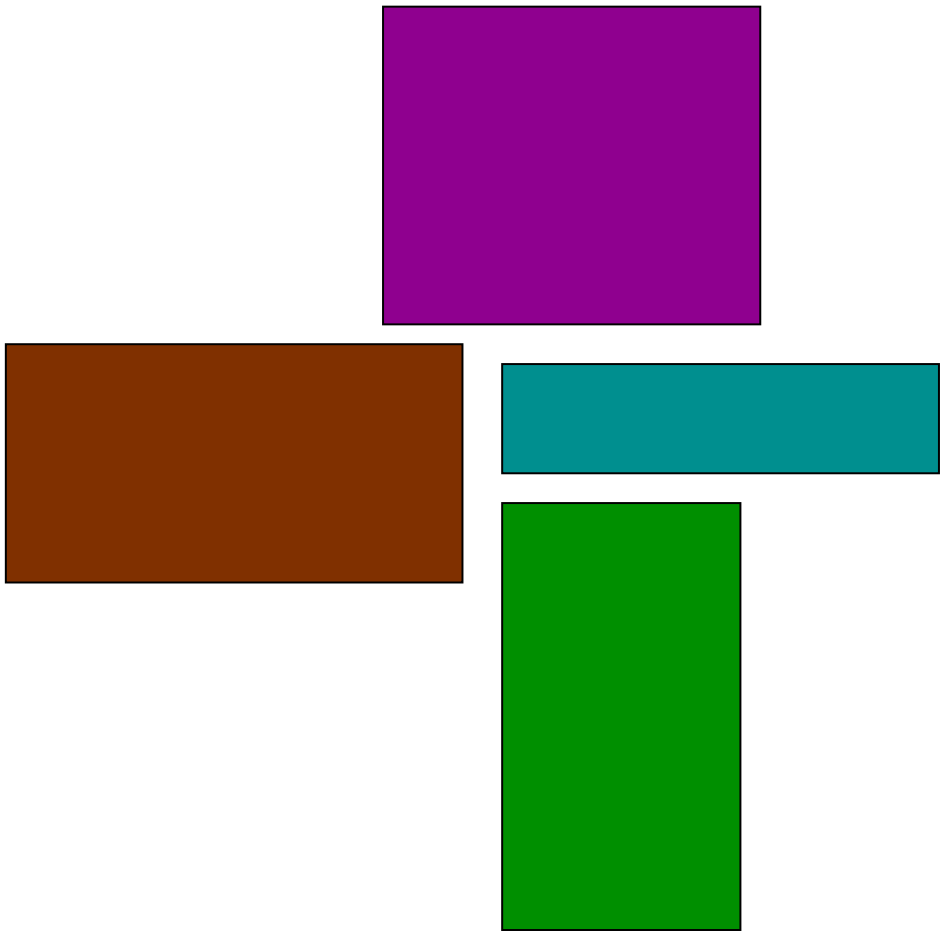


El problema de la mochila

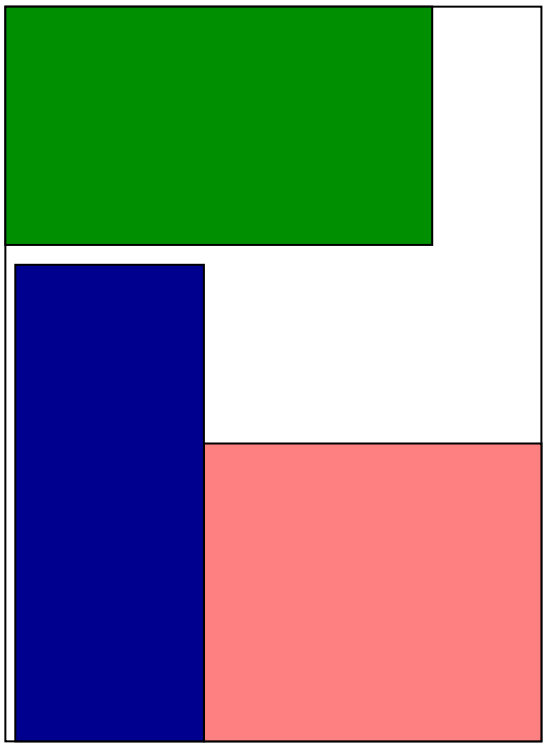


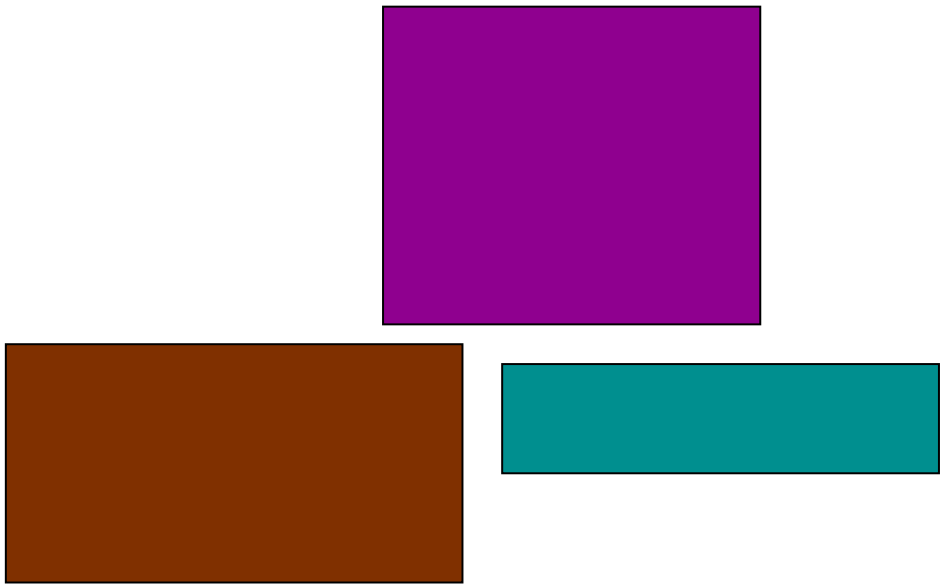
El problema de la mochila



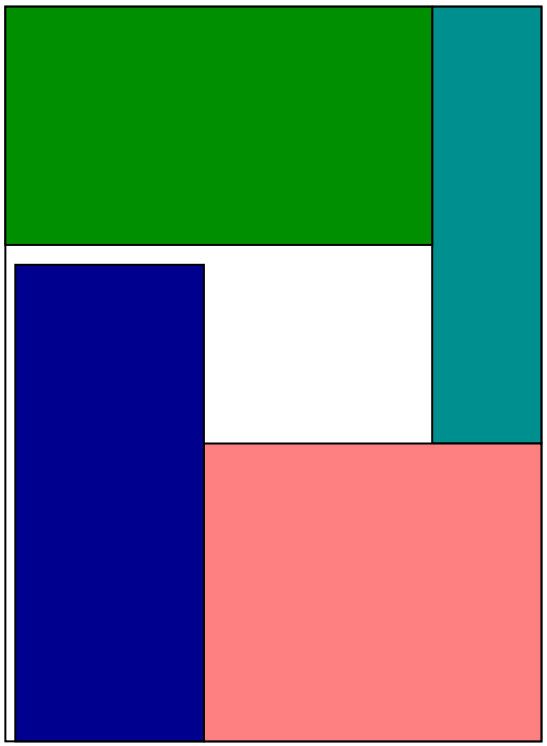


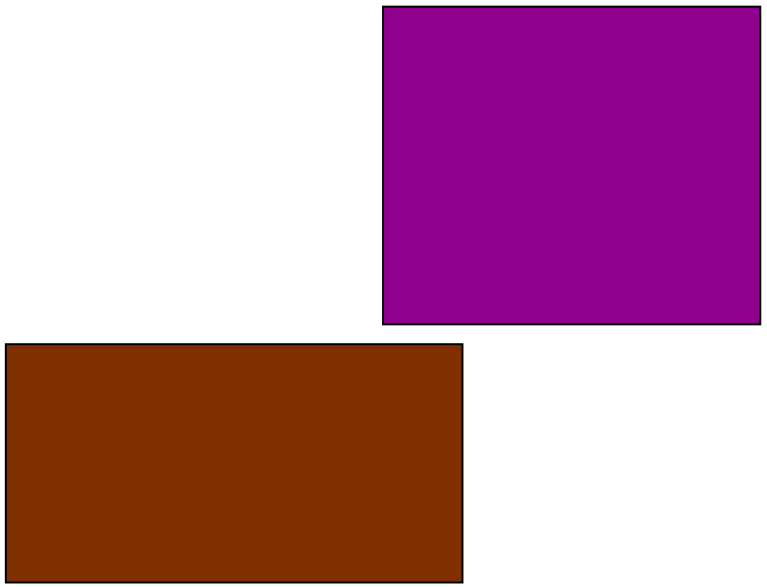
El problema de la mochila





El problema de la mochila





El problema de la mochila (PM)

- Descripción:
  - Tenemos una serie de objetos ($O_1, O_2, ..., O_n$) que se desea introducir en una mochila
  - Cada objeto O_i tiene un valor v_i y un volumen w_i
  - También la mochila tiene un volumen W
  - Hay que introducir en la mochila los objetos de más valor, hasta que se llene
- Esto quiere decir que:

$$V = \sum_{i=1}^{n} v_i \cdot x_i$$

$$x_i = \begin{cases} 1 & \text{si el objeto } O_i \text{ ha sido seleccionado} \\ 0 & \text{en caso contrario} \end{cases}$$

Hay que hacer V máximo, pero sin sobrecargar la mochila:

$$\sum_{i=1}^{n} w_i \cdot x_i \leq W$$

- Es un problema de los llamados np-completos. Si el número de objetos aumenta, el tiempo necesario para resolver el problema aumenta exponencialmente
- Cuando la dimensión del problema es muy grande, es imposible obtener la solución de forma exacta
- Se necesita utilizar algún método aproximado

ALGORITMO CODICIOSO

Se suele resolver con un algoritmo codicioso

- Introducir los objetos en la mochila por orden decreciente de valor específico, comenzando por el objeto de mayor valor específico, hasta que se llene la mochila
- Proporciona óptimos aceptables cuando la distribución de valores específicos es uniformemente variada, pero no cuando es dispersa

Representación del problema

- Hay que:
  - Buscar una buena codificación de las soluciones
  - Encontrar una función de evaluación para cada solución. Es la manera de definir la competición por los recursos de los individuos
  - Incluir las restricciones del problema de alguna forma
- Aquí, un individuo es una solución, es decir un grupo de objetos seleccionados
- La evaluación es el valor total de los objetos seleccionados. Hay que maximizarla

Codificación del problema

- Si es posible es mejor utilizar codificaciones binarias
- Si hay diez objetos diferentes, el tamaño del individuo es precisamente esta cantidad
- El genotipo es binario. Por ejemplo: (1001110000)
- El fenotipo correspondiente será introducir en el saco los objetos: uno, cuatro, cinco y seis

Función de evaluación para el PM

- Cada individuo tiene que tener una evaluación que indique hasta qué punto dicho individuo es bueno. O sea hasta qué punto es capaz de resolver el problema
- En el caso del PM es el valor total de los objetos introducidos en el saco
- Pueden existir otras consideraciones, como por ejemplo las restricciones del problema
- En este caso hay que expresar que el saco no puede soportar más de un peso determinado

Función de evaluación para el PM

- Hay diferentes maneras de expresar dicha restricción:
  - Penalizar a todos los individuos que no respeten la restricción. Se puede valorar negativamente el exceso de peso. Por ejemplo si la restricción es de 55kg y el individuo representa un peso de 65kg, la evaluación es de -10
  - Eliminar de la población los individuos que no respeten las restricciones

- Funciones de penalización: Se modifica la función de evaluación

$$f(x) \leftarrow f(x) - P(h(x))$$

donde $h(x)$ es el grado de violación de la restricción, definido como:

$$h(x) = \begin{cases} w \cdot x - W & Sii \quad w \cdot x > W \\ 0 & Sii \quad w \cdot x \leq W \end{cases}$$

- Posibles funciones de penalización:

$$P_{log}(x) = log_2(1 + v_{max} h(x))$$

$$P_{lin}(x) = v_{max} h(x)$$

$$P_{sq}(x) = (v_{max} h(x))^2$$

  siendo $v_{max} = max_{i=1,\ldots,m} \frac{v_i}{w_i}$
- La función de penalización logarítmica proporciona excelentes resultados para una amplia gama de casos

Evaluación por reparación

- Algoritmos de reparación: se extraen objetos de la mochila cuando está demasiado llena
  - Reparación aleatoria: Se construye x_f extrayendo elementos de la mochila al azar hasta que se vuelva a verificar la restricción de volumen $w \cdot x_f \leq W$
  - Reparación codiciosa: Se extraen los objetos en orden creciente de valor específico, comenzando por el menos valioso y terminando cuando se vuelva a verificar la restricción de volumen
- El algoritmo codicioso proporciona mejores resultados

Evaluación por decodificación

- Algoritmos de decodificación: Se cambia la codificación de los individuos de manera que cada uno pueda ser interpretado como una ”estrategia de incorporación de objetos en la solución”, de tal manera que sea imposible construir una solución no factible
- Para ello se ordenan primero todos los objetos en una lista dinámica de referencia según cierto criterio de prioridad; por ejemplo, si son seis objetos $1,\ 2,\ 3,\ 4,\ 5$ *y* 6 se pueden ordenar así $4, 1, 2, 5, 3, 6$
- Se construye cada individuo como un vector de componentes en el que la i-ésima componente es un puntero con el que se referencia a la lista dinámica (codificación ordinal)

Evaluación por decodificación

- El individuo $\{4, 3, 4, 1, 1, 1\}$ codifica la siguiente colección de objetos: $5, 2, 6, 4, 1, 3$ (la i-ésima componente es un entero positivo no mayor que $m + 1 - i$
- La ventaja de la codificación ordinal: los operadores genéticos conservan la factibilidad de los individuos (si se construye una población inicial de puntos factibles el AG trabajará en todo momento con puntos factibles)
- Hay que tener en cuenta que, dada la codificación, los operadores genéticos se aplican en el nivel de los genes, no de los bits. Así pues, la mutación del i-ésimo gen sólo puede proporcionar un entero del rango $[1..m + 1 - i]$

Evaluación por decodificación

- La elección del criterio de prioridades con el que construir la lista dinámica de referencia L influye decisivamente en la rapidez de la búsqueda (si no hay información $\rightarrow$ orden aleatorio)
- En este caso se crea una lista L de objetos en orden decreciente de sus valores específicos (decodificación codiciosa ordinal)
- Resultados experimentales similares a los obtenidos con los algoritmos de reparación
- Se ha violado el requisito del alfabeto mínimo (el AG está operando con cadenas de enteros)
- Esta representación es mucho menos natural que la anterior

Temario