

Ouafae Debdi
Maximiliano Paredes Velasco
J. Ángel Velázquez Iturbide

Una Evaluación de Eficacia Educativa de GreedExCol

Número 2013-02

Serie de Informes Técnicos DLSI1-URJC
ISSN 1988-8074
Departamento de Lenguajes y Sistemas Informáticos I
Universidad Rey Juan Carlos

Índice

1	Introducción	4
2	Protocolo	4
3	Resultados del Grupo de Control	5
3.1	Análisis de Eficiencia	5
3.2	Análisis de Correlación	7
4	Resultados del Grupo Experimental.....	8
4.1	Análisis de Eficiencia.....	8
4.2	Análisis de Correlación	10
4.3	Prueba T-Student.....	10
5	Prueba de Cohen	11
6	Conclusiones	11
	Agradecimientos.. ..	12
	Referencias.....	12
	Apéndice A: Cuestionario de motivación	13

Una Evaluación de Eficacia Educativa de GreedExCol

Ouafae Debdi, Maximiliano Paredes, J. Ángel Velázquez Iturbide

Departamento de Lenguajes y Sistemas Informáticos I, Universidad Rey Juan Carlos,
C/ Tulipán s/n, 28933, Móstoles, Madrid
{ouafae.debdi, maximiliano.paredes, angel.velazquez}@urjc.es

Resumen. GreedExCol es un sistema para la experimentación interactiva con algoritmos voraces basado en un enfoque CSCL. Este informe presenta los resultados de una evaluación de eficacia educativa bajo diferentes paradigmas de aprendizaje cuyo objetivo es medir la eficiencia educativa de GreedExCol. En el informe se presenta el protocolo de evaluación utilizado y los resultados obtenidos. El informe incluye como apéndice el test de conocimiento realizado.

Palabras clave: Algoritmos voraces, ayudantes interactivos, eficacia educativa, CSCL.

1 Introducción

GreedExCol es una herramienta educativa diseñada para apoyar el aprendizaje de los algoritmos voraces en un enfoque colaborativo. GreedExCol es una ampliación del sistema base GreedEx [1] diseñada para ayudar al alumno a experimentar con diversas funciones de selección para algunos problemas voraces soportados por el sistema GreedEx. El alumno debería apreciar el efecto de las diversas funciones de selección y ser capaz de realizar una elección fundada entre ellas. El método experimental subyacente se explica en [2, 3].

En términos de la taxonomía de Bloom [4], GreedEx debe ayudar en los siguientes niveles de aprendizaje:

- Nivel de comprensión. El alumno comprenderá el problema planteado y el algoritmo voraz que lo resuelve. El algoritmo será independiente de la función de selección elegida, por lo que puede contener fragmentos de pseudocódigo.
- Nivel de análisis. El alumno analizará el efecto de aplicar a unos datos de entrada el mismo algoritmo voraz, pero con diversas funciones de selección.
- Nivel de evaluación. El alumno evaluará el efecto de cada función de selección y seleccionará (empíricamente) las óptimas.

Para evaluar la utilidad de GreedExCol, se realizó un experimento para medir su efecto educativo en dos grupos distintos: un grupo experimental que usó la herramienta y un grupo de control que recibió clases magistrales. En este informe realizamos una evaluación de eficiencia educativa de esta extensión consistente que permite la discusión colaborativa por parte de los alumnos [5]. La estructura del informe es la siguiente. El apartado 2 describe el protocolo utilizado. Los apartados 3 y 4 presentan los resultados obtenidos de los grupos de control y experimental respectivamente. En el apartado 5 comentamos los resultados obtenidos. Finalmente, un apéndice recoge el test de conocimiento.

2 Protocolo

Esta evaluación de GreedExCol se realizó en Abril de 2013 en la asignatura troncal “Diseño y Análisis de Algoritmos” del curso de Grado de Ingeniería del Software en dos grupos diferentes:

- Grupo de control que recibió únicamente clases magistrales y que estaba formado por 46 y 53 alumnos en el Pre y el Post Test de conocimiento respectivamente.
- Grupo experimental que usó la herramienta GreedExCol y que estaba formado por 53 y 41 alumnos en el Pre y el Post Test de conocimiento respectivamente.

Ambos grupos recibieron las mismas clases aunque se diferían en dos sesiones de las clases de los algoritmos voraces. El grupo experimental usó la herramienta GreedExCol mientras el grupo de control recibió únicamente clases magistrales sin ninguna referencia al método experimental. Con respecto al grupo experimental, en la segunda sesión, el profesor había utilizado GreedExCol en clase para dos problemas soportados por GreedExCol, maximizar el número de objetos en una mochila y maximizar el peso introducido en una mochila. Junto a los problemas, había presentado los conceptos de experimentación relacionados y el proceso de experimentación. Después, los alumnos pudieron experimentar en el aula informática con GreedExCol para el problema de la mochila. La práctica que tenían que entregar los alumnos del grupo experimental consistía en la búsqueda y la justificación de funciones de selección óptimas para el problema de selección de actividades que es un problema mucho más exigente que el problema de la mochila. Por otro lado, la práctica del grupo de control consistió en el diseño y la codificación de una función de selección óptima para el problema de selección de actividades.

Los alumnos de ambos grupos completaron la prueba de conocimiento antes y después de recibir las clases sobre algoritmos voraces. El test de conocimiento constaba de cinco preguntas sobre los conceptos básicos de la optimización o algoritmos voraces y la resolución óptima de problemas sencillos. Cada prueba se calificó en una escala que varía entre 0 (el grado más bajo) y 10 (el más alta). El análisis se realizó con el programa paquete estadístico SPSS 20.

3 Resultados del Grupo de Control

En este apartado se presentan los resultados de la eficacia educativa y la correlación en el PreTest y PostTest de conocimiento del grupo de control.

3.1 Análisis de Eficiencia

La tabla 1 muestra los valores descriptivos del grupo de control.

Tabla 1. Valores Descriptivos.

Grupos		Estadístico	Error típ.	
Test est	PreT	Media	3,7659	,26512
	Intervalo de confianza para la media al 95%	Límite inferior	3,2300	
		Límite superior	4,3017	
	Media recortada al 5%	3,6832		
	Mediana	3,3000		
	Varianza	2,882		
	Desv. típ.	1,69759		
	Mínimo	1,00		
	Máximo	8,30		
	Rango	7,30		
	Amplitud intercuartil	2,75		
	Asimetría	,729	,369	
	Curtosis	-,062	,724	
	Post Test	Media	4,3000	,20819
Intervalo de confianza para la media al 95%	Límite inferior	3,8792		
	Límite superior	4,7208		
	Media recortada al 5%	4,2859		

Mediana	4,3000	
Varianza	1,777	
Desv. típ.	1,33304	
Mínimo	1,30	
Máximo	7,70	
Rango	6,40	
Amplitud intercuartil	1,50	
Asimetría	,188	,369
Curtosis	,180	,724

En la Figura 1, se muestra el diagrama de cajas de PreTest y el PostTest de conocimiento del grupo de control.

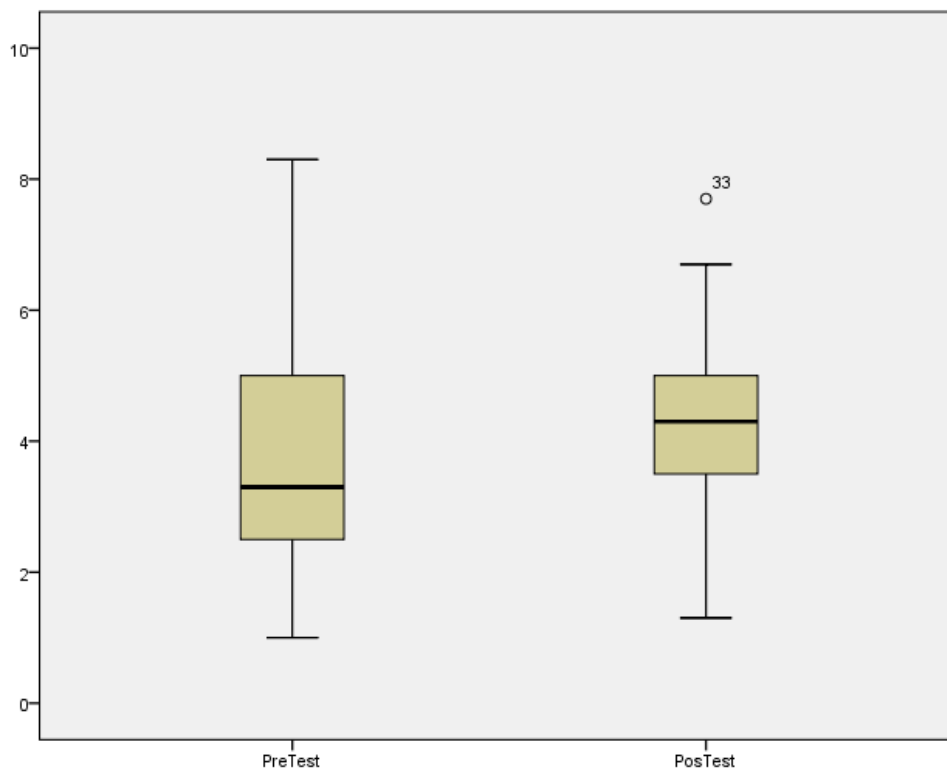


Figura 1. Diagrama de cajas para el grupo de control.

Como puede observarse en el diagrama de cajas, el rango de notas en el PreTest de conocimiento está entre 2.6 y 5.1 mientras en el PostTest conocimiento el rango está entre 3.7 y 5.1. También podemos observar un valor atípico en el PostTest de conocimiento.

Según la tabla 1, la media del PostTest de conocimiento es mayor que la media del PreTest de conocimiento. Para verificar que esta media es poblacionalmente significativa realizamos una comparación de los resultados de los dos Test con un contraste de ANOVA sobre las medias poblacionales, realizando los siguientes pasos:

1. Verificación de la normalidad de las muestras.
2. Verificación de la homogeneidad de las muestras o la prueba de Wilcoxon en el caso de la no normalidad de las muestras

3. Contraste ANOVA.

La tabla 2 muestra el análisis de normalidad para el PreTest y el PostTest de conocimiento del grupo de control.

Tabla 2. Análisis de normalidad.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	Gl	Sig.	Estadístico	Gl	Sig.
PreTest	,169	41	,005	,939	41	,028
PostTest	,086	41	,200*	,991	41	,984

La celda resaltada en negrita en la tabla 2 muestra un sig < 0.05, entonces se rechaza la hipótesis nula, no existe normalidad en la muestra de PreTest del grupo de control, por tanto, se opta por realizar la prueba de Wilcoxon para las muestras no paramétricas y relacionadas.

Tabla 3. Prueba de Wilcoxon.

Resumen de prueba de hipótesis				
	Hipótesis nula	Test	Sig.	Decisión
1	La mediana de las diferencias entre PreTest y PostTest es igual a 0.	Prueba de Wilcoxon de los rangos con signo de muestras relacionadas	,020	Rechazar la hipótesis nula.

Se muestran las significancias asintóticas. El nivel de significancia es ,05.

Según la prueba de Wilcoxon mostrada en la tabla 3, se rechaza la hipótesis nula, existen diferencias entre las notas del PreTest y el PostTest del grupo de control, es decir, la media del PostTest de conocimiento es estadísticamente significativa. En resumen, se puede concluir que el nivel de conocimiento del grupo de control haya mejorado tras recibir clases sobre algoritmos voraces.

3.2 Análisis de Correlación

Según la prueba de normalidad mostrada en la tabla 2, no existe la normalidad en las notas del PreTest, por tanto no podemos realizar un análisis de correlación de Pearson. Pero sí la prueba de Spearman para muestras que no cumplen la normalidad.

Tabla 4. Correlación de Spearman

			PreTest	PostTest
Rho de Spearman	PreTest	Coefficiente de correlación	1,000	,572**
		Sig. (bilateral)	.	,000
		N	41	41
	PostTest	Coefficiente de correlación	,572**	1,000
		Sig. (bilateral)	,000	.
		N	41	41

** . La correlación es significativa al nivel 0,01 (bilateral).

La tabla 4 muestra un sig = 0.00 < 0.05, por tanto existe una correlación significativa de 0.572 entre las notas del Pre y Post conocimiento del grupo de control. Es decir, los alumnos que sacaron una buena nota en el PreTest también lo hicieron en el PostTest y viceversa.

4 Resultados del Grupo Experimental

En este apartado se presenta los resultados descriptivos, el análisis de eficiencia, la correlación entre el Pre y Post conocimiento y la prueba T-Student del grupo experimental.

4.1 Análisis de Eficiencia

La tabla 5 muestra los valores descriptivos para el PreTest y PostTest de conocimiento del grupo experimental.

Tabla 5. Valores descriptivos

Grupos		Estadístico	Error típ.		
Test	PreTest	Media	3,7237	,23816	
		Intervalo de confianza para la media al 95%	Límite inferior	3,2411	
			Límite superior	4,2062	
		Media recortada al 5%	3,6915		
		Mediana	3,5000		
		Varianza	2,155		
		Desv. típ.	1,46812		
		Mínimo	1,00		
		Máximo	7,50		
		Rango	6,50		
		Amplitud intercuartil	2,50		
		Asimetría	,374	,383	
		Curtosis	-,205	,750	
		PostTest	PostTest	Media	5,1816
Intervalo de confianza para la media al 95%	Límite inferior			4,5784	
	Límite superior			5,7848	
Media recortada al 5%	5,1389				
Mediana	4,9000				
Varianza	3,368				
Desv. típ.	1,83522				
Mínimo	1,80				
Máximo	9,00				
Rango	7,20				
Amplitud intercuartil	2,50				
Asimetría	,454			,383	
Curtosis	-,489			,750	

La Figura 2 muestra el diagrama de cajas para la muestra del grupo experimental.

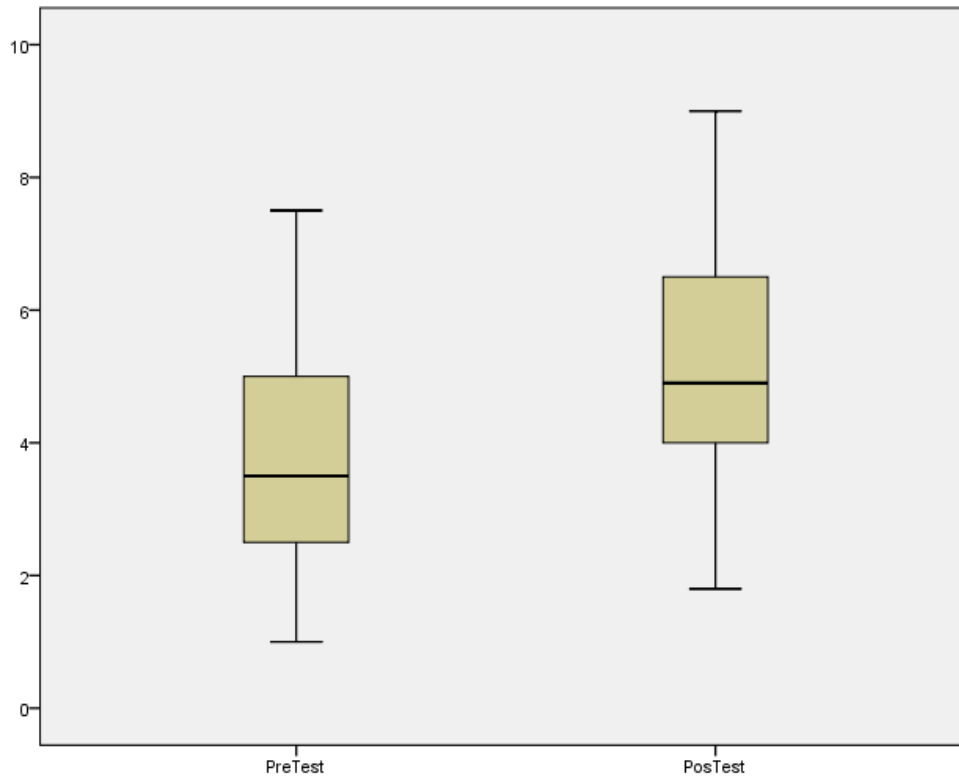


Figura 2. Diagrama de cajas del grupo experimental

Como podemos ver en el diagrama de cajas de la figura 2, el 50 % de los alumnos sacaron entre 2.3 y 5 en el PreTest mientras en el PostTest el rango de notas está entre 3.7 y 6.2

Según la tabla 5, la media del PostTest de conocimiento es mucho mayor que la media del PreTest de conocimiento. Para verificar que esta media es poblacionalmente significativa realizamos una comparación de los resultados de los dos test con un contraste de ANOVA sobre las medias poblacionales, realizando los siguientes pasos:

1. Verificación de la normalidad de las muestras.
2. Verificación de la homogeneidad de las muestras o la prueba de Wilcoxon en el caso de la no normalidad de las muestras
3. Contraste ANOVA.

La tabla 6 muestra la prueba de Kolmogorov-Smirnov y Shapiro-Wilk para el Pre y Post conocimiento del grupo experimental.

Tabla 6. Análisis de normalidad.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	G1	Sig.	Estadístico	gl	Sig.
PreTest	,110	38	,200	,973	38	,476
PostTest	,171	38	,007	,957	38	,149

Como se puede observar en la tabla 6, las celdas resaltadas en negrito muestran un sig > 0.05, por tanto no se rechaza la hipótesis nula (las muestras son normales). Lo siguiente es averiguar la homogeneidad de varianzas.

Tabla 7. Prueba de homogeneidad de varianzas (PrePost).

Estadístico de	g1	g2	Sig.
----------------	----	----	------

Levene			
1,703	1	74	,196

La tabla 7 muestra un $\text{sig} = 0,196 > 0,05$ (celda en negrita), entonces no se rechaza la hipótesis nula, es decir, existen homogeneidad de varianzas.

Tabla 8. ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	40,384	1	40,384	14,623	,000
Intra-grupos	204,366	74	2,762		
Total	244,749	75			

La tabla 8 muestra un $\text{sig} = 0,00 < 0,05$, rechazándose la hipótesis nula ($H_0 =$ las medias poblacionales son iguales), lo que implica que los resultados del Pre y Post conocimiento son diferentes.

Según la tabla 5 y la tabla 8, se puede afirmar que las notas del PostTest del grupo experimental se mejoran usando la herramienta GreedExCol.

4.2 Análisis de Correlación

Dado que existe la normalidad en las muestras de Pre y Post conocimiento, podemos realizar la correlación de Pearson entre ambas muestras.

Tabla 9. Correlaciones

		PreTest	PosTest
PreTest	Correlación de Pearson	1	,365*
	Sig. (bilateral)		,024
	N	38	38
PostTest	Correlación de Pearson	,365*	1
	Sig. (bilateral)	,024	
	N	38	38

*. La correlación es significativa al nivel 0,05 (bilateral).

La tabla 9 muestra un $\text{sig} = 0,24 < 0,05$, por tanto se afirma la existencia de una correlación significativa de 0.365 entre las notas del PreTest y PostTest de conocimiento, es decir, los alumnos que sacaron una buena nota en el PreTest también lo hicieron en el PostTest y viceversa.

4.3 Prueba T-Student

Para medir el grado de diferencia entre el PreTest y Post Test de conocimiento del grupo experimental, se ha optado por realizar la prueba de T-Student para muestras relacionadas dado que ambas muestras son normales. Las tablas 10, 11, 12 son el resultado generado por SPSS.

Tabla 10. Estadísticas de muestras relacionadas.

	Media	N	Error típ. de la media	Desviación típ.
PosTest	5,1816	38	,29771	1,83522
PreTest	3,7237	38	,23816	1,46812

Tabla 11. Correlaciones de muestras relacionadas.

	N	Correlación	Sig.
PreTest y PostTest	38	,365	,024

Tabla 12. Prueba de muestras relacionadas.

		PostTest - PreTest
Diferencias relacionadas	Media	1,45789
	Desviación típ.	1,88574
	Error típ. de la media	,30591
	95% Intervalo de confianza para la diferencia	Inferior Superior
T		4,766
Gl		37
Sig. (bilateral)		,000

Interpretando los resultados mostrados en la tabla 12, podemos afirmar que existe una diferencia de media de 1.457 entre el Pre y Post conocimiento del grupo experimental, además la verdadera media de la población se encuentra entre 0,83 y 2,07 con una probabilidad del 95%. En resumen, como T-Student (37) = 4,766 con un sig =0.00, podemos rechazar la hipótesis nula (existe diferencias significativas entre las notas del Pre y Post conocimiento del grupo experimental). Por tanto podemos concluir que se produjo un aumento significativo en las notas de Post conocimiento, es decir, hemos encontrado una fuerte evidencia de que GreedExCol mejora el nivel de conocimiento de los alumnos.

5 Prueba de Cohen

Para poder comparar las diferencias del incremento en el nivel de conocimiento entre el grupo experimental y control, se ha optado por calcular la magnitud o el tamaño del efecto, y así conseguir la relevancia de la diferencia encontrada entre ambos grupos.

Tabla 13. Prueba de Cohen (Tamaño del efecto)

	Cohen's d	Effect size r	Porcentaje
Grupo de control	0.34	0.17	14%
Grupo experimental	0.87	0.40	39%

La tabla 13 muestra un cambio de 0.34 para el grupo de control que corresponde a un efecto pequeño (se considera medio para d igual a 0.5) con un efecto de cambio de 0.17 y un porcentaje de 14%. Sin embargo, el grupo experimental ha tenido un valor de d de cohen de 0.87, lo que corresponde a un efecto grande (se considera grande para d 0.80 o superior) con un efecto de cambio de 0.40. Del mismo modo, un porcentaje de cambio de 39% que muestra un nivel de cambio alto para el grupo experimental.

6 Conclusiones

En resumen, los resultados de nuestra evaluación muestran que la uso de la herramienta GreedExCol tuvo una influencia positiva en el rendimiento de los alumnos para aprender algoritmos voraces, obteniendo diferencias significativas en la notas de Pre y Post conocimiento después de usar

GreedExCol. Sin embargo, los alumnos que recibieron enseñanza tradicional tuvieron una mejora pequeña en su nivel de conocimientos tras las clases magistrales.

Por otra parte, el resultado del tamaño del efecto es muy alto en el grupo experimental como consecuencia de la utilización de GreedExCol frente a un efecto pequeño para el grupo de control.

Agradecimientos. Este trabajo se ha financiado con el proyecto TIN2011-29542-C02-01 del Ministerio de Economía y Competitividad.

Referencias

1. Velázquez Iturbide, J. Á., Debdí, O., Esteban Sánchez, N., Pizarro, C.: GreedEx: A visualization tool for experimentation and discovery learning of greedy algorithms. *IEEE Transactions on Learning Technologies*, 6(2):130-143, Abril-Junio 2013
2. Velázquez Iturbide, J. Á., Pérez Carrasco, A.: Active learning of greedy algorithms by means of interactive experimentation. En: *Proceedings of the 14th Annual Conference on Innovation and Technology in Computer Science Education – ITiCSE 2009*, ACM Press (2009) 119-123
3. J Velázquez Iturbide, J. Á.: Refinement of an experimental approach to computer-based, active learning of greedy algorithms. En: *Proceedings of the 17th Annual Conference on Innovation and Technology in Computer Science Education – ITiCSE 2012*, ACM Press (2012) 46-51
4. Bloom, B., Furst, E., Hill, W., Krathwohl, D. R.: *Taxonomy of Educational Objectives: Handbook I, The Cognitive Domain*. Addison-Wesley (1956)
5. Debdí, O., Paredes Velasco, M., Velázquez Iturbide, J. Á.: Una evaluación de usabilidad de GreedExCol. *Serie de Informes Técnicos DLSII-URJC*, 2012-05 (2012). Departamento de Lenguajes y Sistemas Informáticos I, Universidad Rey Juan Carlos

Apéndice A: Cuestionario

Nombre y apellidos: _____

Titulación/grupo: _____

Número de Expediente: _____

El test es voluntario y permite aumentar tu nota de prácticas si has rellenado completamente todos sus apartados.

IMPORTANTE: No basta con que contestes sí o no. Debes completar cada respuesta o, al menos, decir claramente que no la encuentras.

1. Los problemas de optimización buscan aumentar beneficios o reducir costes mediante la toma de unas decisiones entre todas las posibles. Por ejemplo, se puede querer encontrar el camino más corto entre dos ciudades. Dado un problema de optimización, ¿sólo hay una solución óptima o puede haber varias en general?

¿Por qué?

2. Y sobre el método o algoritmo que pueda resolver un problema de optimización, ¿sólo hay un algoritmo posible o puede haber varios en general?

¿Por qué?

3. Di si conoces el nombre de los algoritmos siguientes:

o Algoritmo de Dijkstra:

Sí No

En caso afirmativo, indica qué problema resuelve:

o Algoritmo de Prim

Sí No

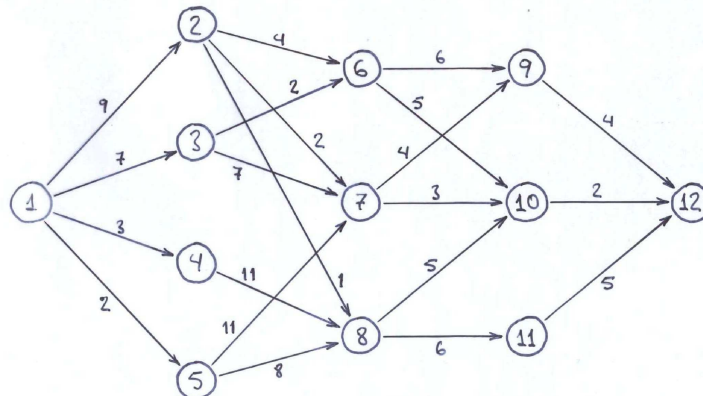
En caso afirmativo, indica qué problema resuelve:

o Algoritmo de Kruskal

Sí No

En caso afirmativo, indica qué problema resuelve:

4. Sea el siguiente grafo cuyos nodos se encuentran dispuestos en cinco grupos verticales:



El problema consiste en encontrar un camino de longitud mínima entre el nodo origen (nodo 1) y el nodo destino (nodo 12). Por la forma del grafo, dicho camino óptimo constará de cuatro “fases”, siendo cada fase un arco que une un nodo de un grupo a otro nodo del grupo situado inmediatamente a su derecha. En general, para un grafo de este estilo, ¿crees que debe haber un único camino óptimo o que podría haber varios?

¿Por qué?

Dicho camino óptimo, ¿crees que puede encontrarse en una sola iteración de izquierda a derecha, tomando un arco de cada fase, o que debe analizarse todo el grafo de alguna forma más exhaustiva?

5. Tenemos n trabajos, donde cada trabajo i tiene una fecha tope de realización $f_i > 0$ y un beneficio $b_i > 0$. El beneficio b_i del trabajo i se gana si y sólo si se empieza a realizar antes o coincidiendo con su fecha tope f_i . Cada trabajo necesita una unidad de tiempo para realizarse; además, en un instante de tiempo sólo se puede ejecutar una tarea. Queremos determinar los trabajos a realizar (y en qué orden) para que el beneficio total sea máximo.

Por ejemplo, sean 4 actividades con las fechas tope y los beneficios contenidos en la siguiente tabla. Tres soluciones posibles serían $\{2,1\}$, $\{1,3\}$, $\{3,1\}$; sin embargo, la solución de mayor beneficio sería $\{4,1\}$.

Núm. trabajo	1	2	3	4
Beneficios	100	10	15	27
Fechas tope	2	1	2	1

¿Cuál o cuáles de las siguientes estrategias crees que encuentra soluciones óptimas (índice es el número de trabajo)?

- Selección por orden de menor a mayor índice
- Selección por orden de mayor a menor índice
- Selección por orden de menor a mayor beneficio
- Selección por orden de mayor a menor beneficio
- Selección por orden de menor a mayor fecha tope
- Selección por orden de mayor a menor fecha tope

Respalda tu respuesta con algún ejemplo o razonamiento.

6. Sean dos vectores a y b de n números naturales cada uno, de los cuales a está ordenado. Se desea obtener una reordenación de la secuencia b tal que maximice la suma de los elementos de a que no superen al elemento correspondiente de b , es decir:

$$\text{maximizar } \sum a_i \mid a_i \leq b_i$$

Por ejemplo, dados:

$$a = \{ 2, 6, 7, 10, 22 \}$$

$$b = \{ 15, 1, 7, 5, 2 \}$$

una reordenación óptima de b es:

$$b = \{ 5, 2, 7, 15, 1 \}$$

con un beneficio $2+7+10 = 19$. Sin embargo, no es óptima la reordenación:

$$b = \{ 1, 2, 5, 7, 15 \}$$

con un beneficio 0.

Un algoritmo sencillo consistiría en coger los elementos de a en orden decreciente. A cada elemento de a se le hace corresponder el mayor de los elementos que quedan de b , si éste es mayor, o el menor de los elementos que quedan de b , en caso contrario.

En el ejemplo anterior, en el primer paso no se hace corresponder al valor 22 de a con el mayor valor de b (15) porque éste es menor, sino con el valor menor de b (1). Sin embargo, en el segundo paso, se hace corresponder el siguiente valor de a (10) con el mayor valor de b (15) porque éste es mayor.

¿Crees que el algoritmo es óptimo siempre?

Sí

No

Respalda tu respuesta con algún ejemplo o razonamiento.
