

Calculo empírico del mejor umbral del algoritmo Mergesort

Enunciado del ejercicio

Realizar un estudio empírico para determinar un buen umbral para el algoritmo de ordenación por mezcla o mergesort (que utiliza como algoritmo auxiliar el algoritmo de ordenación por inserción para casos de tamaño menor que el umbral).

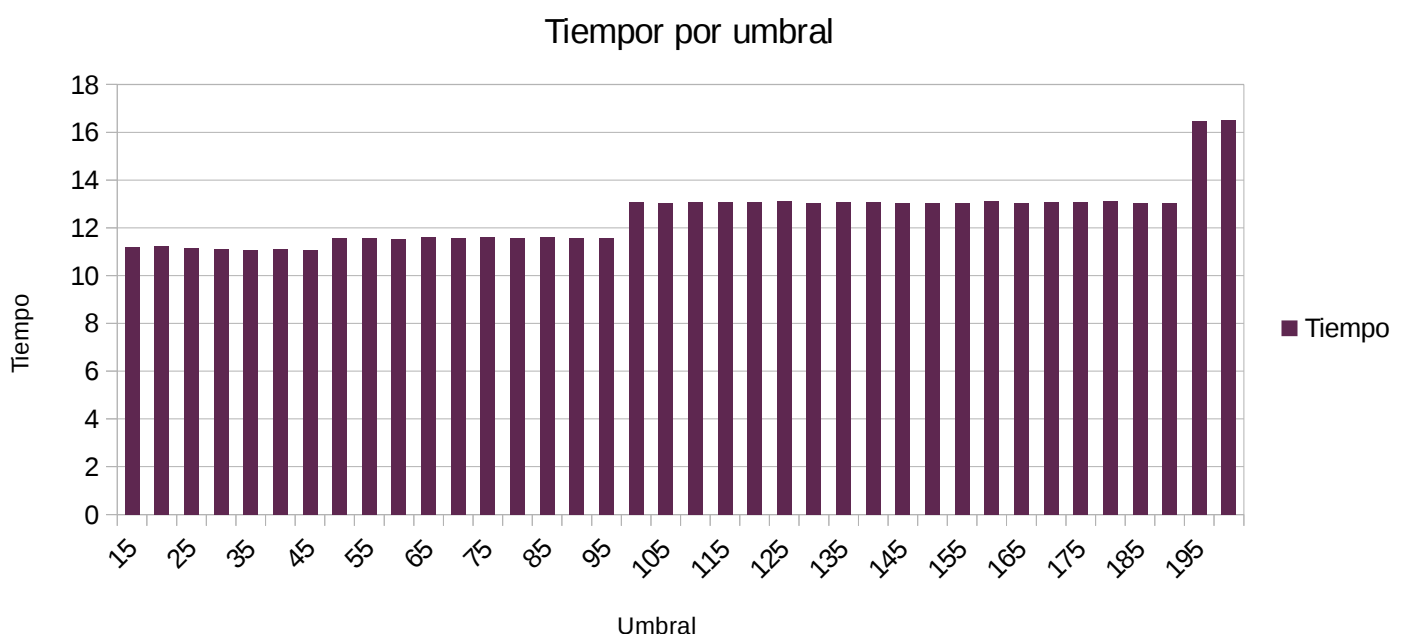
Toma de datos

Para la realización del experimento hemos usado el código del algoritmo mergesort proporcionado para las prácticas de esta misma asignatura. En él hemos modificado el código para que acepte un umbral cuando se ejecute para así facilitar la ejecución mediante un script. También lo he modificado para que imprima tanto el umbral como el tiempo que ha tardado para ese umbral.

El experimento ha sido ejecutado con un tamaño de problema fijo, ya que aquí lo que queremos observar es cómo varía el tiempo en función del umbral. Los datos han sido recogidos con un ordenador que tenía un Intel(R) Core(TM) i5-6600 CPU a 3.30GHz con 16Gb de RAM y el sistema operativo era Linux Mint 18.1.

El tamaño del problema ha sido de 50.000.000. Hemos empezado con un umbral de 10 y hemos ido sumando de 5 en 5 hasta llegar a 200.

Hemos obtenido la siguiente gráfica con los datos obtenidos:



He realizado tres mediciones con los mismo umbrales y el mismo tamaño de problema para luego realizar un media de los tres tiempos obtenidos y poder obtener una medicion un poco mas fiable.

Umbral	Tiempo			Media
15	11,1926	11,2865	11,528	11,3357
20	11,2387	11,2822	11,3093	11,276733333
25	11,1555	11,3036	11,1144	11,191166667
30	11,1282	11,2315	11,1902	11,1833
35	11,0796	11,285	11,1241	11,1629
40	11,1094	11,2746	11,095	11,159666667
45	11,0846	11,2175	11,1466	11,149566667
50	11,5654	11,6742	11,5487	11,5961
55	11,5482	11,6804	11,5851	11,604566667
60	11,5287	11,6999	11,5211	11,583233333
65	11,5898	11,7012	11,5326	11,607866667
70	11,5665	11,6843	11,565	11,605266667
75	11,5968	11,6709	11,5524	11,6067
80	11,5753	11,6832	11,5615	11,606666667
85	11,6013	11,7059	11,5479	11,618366667
90	11,5781	11,6971	11,5593	11,6115
95	11,5841	11,6489	11,6584	11,630466667
100	13,0837	13,1529	13,0507	13,095766667
105	13,0329	13,217	13,0477	13,0992
110	13,0827	13,2483	13,0643	13,131766667
115	13,0696	13,1846	13,063	13,105733333
120	13,0552	13,184	13,0365	13,0919
125	13,1031	13,1888	13,0215	13,104466667
130	13,0278	13,2165	13,3393	13,194533333
135	13,0837	13,2079	13,0088	13,100133333
140	13,0761	13,1582	13,0591	13,0978
145	13,0336	13,1947	13,0451	13,091133333
150	13,0437	13,2346	13,0312	13,103166667
155	13,0328	13,1823	13,0281	13,081066667
160	13,0964	13,203	12,9961	13,0985
165	13,0483	13,1508	13,0623	13,087133333
170	13,0653	13,2076	13,0181	13,097
175	13,0629	13,1623	13,0462	13,090466667
180	13,1129	13,1777	13,0566	13,115733333
185	13,0477	13,3467	13,0586	13,151
190	13,0106	13,1999	13,0264	13,078966667
195	16,4854	16,7426	16,5135	16,5805
200	16,5124	16,7147	16,5766	16,601233333

De los datos obtenidos podemos observar como en el intervalo desde 15 hasta 100 tenemos tiempos muy parecidos, ya que todos están en torno a los 11 segundos. Aunque como buscamos el mejor umbral de todos en nuestro caso, podemos encontrarlo en el 45 ya que es el que menor media tiene de todos, aunque por muy poco. Si yo tuviera que elegir, optaría por cualquier umbral comprendido entre el 35 y el 45, ya que tiene valores tan cercanos los unos con los otros que con otras mediciones nos podrían dar alguno de ellos como el mejor umbral.

Apartir de 100 podemos ver que el umbral se vuelve muy malo y a partir del 195 no sería nada recomendable coger el umbral ya que obtenemos tiempos bastante elevados en comparación.