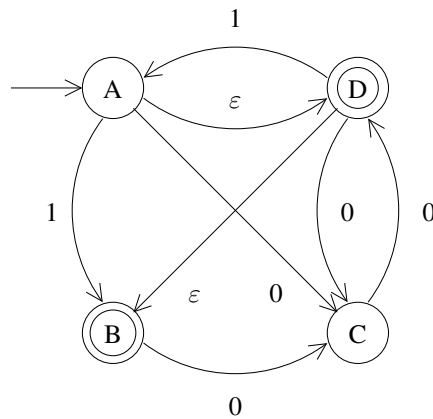


58093 String Processing Algorithms (Autumn 2012)

Exercises 1 (1 November)

Solve the following problems before the exercise session and be prepared to present your solutions at the session.

1. The number of comparison operations required to sort n items is $\Omega(n \log n)$. Recall the proof of this fact and be prepared to present the proof.
2. The worst case time complexity of the standard quicksort algorithm is $\Omega(n^2)$, but by a suitable pivot selection one can achieve $\mathcal{O}(n \log n)$ time. Explain how to achieve
 - (a) average time complexity $\mathcal{O}(n \log n)$
 - (b) expected time complexity $\mathcal{O}(n \log n)$
 - (c) worst case time complexity $\mathcal{O}(n \log n)$.
3. Transform the following nondeterministic finite automaton into a deterministic finite automaton.



4. Write a program to find the ten most frequent characters in the english text file <http://pizzachili.dcc.uchile.cl/texts/nlang/english.50MB.gz> (Note that the file is compressed with `gzip` and must be decompressed before processing.) Be prepared to show the code and the output of your program.
5. Write a program to find the ten most frequent *words* in the file of the previous exercise. Here you may consider the space character to be the only word separator, i.e., words can contain punctuation, newlines etc., but using a more sophisticated word parsing method is allowed too.