Advanced Data Structures (ADS-MIRI)
Assignment:
Random Binary Search Trees

Amalia Duch

In this assignment you have to implement binary search trees as well as the algorithms to search, insert and delete elements in/from them (or alternatively to look for a working implementation useful to conduct the required experimental study). Then, you will have to conduct some experimental studies on the average-case cost of your implementations (or the implementations that you are using).

For your experiments you will need also a way to build random binary search trees of different sizes from scratch. In order to produce a random binary search tree of size n we propose the following procedure (other possibilities are also acceptable, as for instance random permutations of the n keys). Generate n random points in the unit interval $[0, 1]$ uniformly and insert each point into your tree.

1 Analysis of the Average Cost of Insertions

Let I_n be the expected cost of the insertion of a key x in a random binary search tree T of size n, measured as the number of visited nodes of T.

- Give (reasonedly) a recurrence for I_n.
- Solve it using any method of your choice.
- Show experimentally that I_n fits the theoretical prediction. For your experiments, for each tree of size n generate q random insertions in the same way you generate the random points and take averages.

2 Analysis of Internal Path Length

Let IPL_n be the internal path length of a random binary search tree T of size n.

- Give (reasonedly) a recurrence for IPL_n.

- Solve it using the continuous master theorem.
- Try to show experimentally that IPL_n fits the theoretical prediction. Why is it difficult? Try to find out the reason.

3 Analysis of Interleaved Insertions and Deletions

- Do long sequences of interleaved insertions and deletions destroy the randomness of an initially random binary search tree? Design experiments to find out the answer to this question.

4 Your Report

Once the full suite of experiments has been executed and data has been gathered, you have to prepare a report.

1. Describe briefly your implementations and the program to execute the experiments. Give full listings of the code as an appendix of your report.
2. Describe briefly the experimental setup, how many different combinations of the parameters have you studied, how many runs have you performed, etc.
3. Provide tables and plots sumarising the results of the experiments.
4. Compare the experimental results with the theoretical predictions. Plots combining the theoretical values and the experimental results are useful, but it is also important to quantify the difference between the theoretically predicted values and the empirical values.
5. Write down your conclusions.

We encourage you to use LaTeX to prepare your report. For the plots you can use any of the multiple packages that LaTeX has (in particular, the bundle TikZ+PGF) or use independent software such as gnuplot and then include the images/PDF plots thus generated into your document.

Submit your reports in PDF format using the *Racó*.