Chair for Algorithms
and Data Structures
Prof. Dr. Hannah Bast
Florian Bäurle

Information Retrieval
WS 2012/2013

http://ad-wiki.informatik.uni-freiburg.de/teaching

Exercise Sheet 3

Submit until Wednesday, November 14 at 4:00pm

Exercise 1 (10 points)

Implement a class *Intersect* with two algorithms for the intersection of two sorted lists: the simple linear-time one (which we saw already in Lecture 1), and the asymptotically optimal exponential-binary-search algorithm (introduced and explained in Lecture 3).

Aim for an equally efficient implementation for both algorithms. Also, writing a unit test to check correctness of both algorithms on a non-trivial example is a must. Otherwise, chances are that one algorithm is better than the other simply because it does not compute the results properly.

Exercise 2 (10 points)

Compare the running times of the two algorithms on two sorted lists with numbers chosen uniformly at random *from the same range* (this is important). Run two experiments: one where the ratio of the list lengths is $R = 5$, and one where the ratio is $R = 50$. In both cases, the total number of elements in both lists should be 250 million. Report your results in the table linked from the Wiki, following the instructions given there.

Exercise 3 (optional, good for exam preparation)

Prove that for all positive integers k and n, $\ln n \leq k \cdot \ln(1 + n/k)$ (the running time of the exp-bin intersection is worst when the gaps are all equal), as well as $k \cdot \ln(1 + n/k) \leq k + n$ (the exp-bin intersection is asymptotically better than the linear-time intersection).

Exercise 4 (optional, good for exam preparation)

The entropy (information content) of a probability distribution $p_1, \ldots, p_n$, where all $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$, is defined as $-\sum_{i=1}^{n} p_i \cdot \ln p_i$. Show that the entropy is maximized for the uniform distribution, that is, when the probabilities are all equal.

As usual, commit your code to our SVN, in a new sub-directory *exercise-sheet-03*, along with a text file *experiences.txt* with your feedback. As a minimum, say how much time you invested and if you had major problems, and if yes, where. If you do one of the optional theoretical tasks, commit your solutions in a PDF file in the same folder.