In as Few Comparisons as Possible

J. Ian Munro

Cheriton School of Computer Science, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada
imunro@uwaterloo.ca

Abstract. We review a variety of data ordering problems with the goal of solving them in as few comparisons as possible. En route we highlight a number of open problems, some new, some a couple of decades old, and others open for up to a half century. The first is that of sorting and the Ford-Johnson Merge-Insertion algorithm [8] of 1959, which remains the "best", at least for the "best and worst" values of n. Is it optimal, or are its extra .028..n or so comparisons beyond the information theoretic lower bound necessary?

Moving to selection problems we first examine a special case. The problem of finding the second largest member of a set is fairly straightforward in the worst case. The best expected case method remains the $n + \Theta(\lg \lg n)$ method of Matula from 1973 [10]. It begs the question as to whether the $\lg \lg n$ term is necessary. The status of median finding has remained unchanged for a couple of decades, since the work of Dor and Zwick [4,5]. $(3 - \delta)n$ comparisons are sufficient, while $(2 + \epsilon)n$ are necessary. So the constant isn't an integer, but is it $\log_{4/3} 2$ as conjectured by Paterson [11]? This worst case behavior is in sharp contrast with the expected case of median finding where the answer has been known since the mid-'80's [3,6].

Finally we look at the problem of partial sorting (arranging elements according to a given partial order) and completing a sort given partially ordered data. The latter problem was posed and solved within n or so comparisons of optimal by Fredman in 1975 [7]. The method, though, could use exponential time to determine which comparisons to perform. The more recent approaches of Cardinal et al [2,1] to these problems are based on graph entropy arguments and require only polynomial time to determine the comparisons to be made. Indeed the solution to the partial ordering problem involves a reduction to multiple selection [9]. In both cases the number of comparisons used differs from the information theoretic lower bound by only a lower order term plus a linear term.

References

1. Cardinal, J., Fiorini, S., Joret, G., Jungers, R.M., Ian Munro, J.: An efficient algorithm for partial order production. SIAM J. Comput. 39(7), 2927–2940 (2010)
2. Cardinal, J., Fiorini, S., Joret, G., Jungers, R.M., Ian Munro, J.: Sorting under partial information (without the ellipsoid algorithm). In: STOC, pp. 359–368 (2010)
3. Cunto, W., Ian Munro, J.: Average case selection. J. ACM 36(2), 270–279 (1989)

S.P. Pal and K. Sadakane (Eds.): WALCOM 2014, LNCS 8344, pp. 5–6, 2014.

4. Dor, D., Zwick, U.: Selecting the median. SIAM J. Comput. 28(5), 1722–1758 (1999)
5. Dor, D., Zwick, U.: Median selection requires $(2 + \epsilon)n$ comparisons. SIAM J. Discrete Math. 14(3), 312–325 (2001)
6. Floyd, R.W., Rivest, R.L.: Expected time bounds for selection. Commun. ACM 18(3), 165–172 (1975)
7. Fredman, M.L.: Two applications of a probabilistic search technique: Sorting x + y and building balanced search trees. In: STOC, pp. 240–244 (1975)
8. Ford Jr., L.R., Johnson, S.B.: A tournament problem. American Mathematical Monthly 66(5), 387–389 (1959)
9. Kaligosi, K., Mehlhorn, K., Ian Munro, J., Sanders, P.: Towards optimal multiple selection. In: Caires, L., Italiano, G.F., Monteiro, L., Palamidessi, C., Yung, M. (eds.) ICALP 2005. LNCS, vol. 3580, pp. 103–114. Springer, Heidelberg (2005)
10. Matula, D.W.: Selecting the the best in average $n+\theta(\log\log n)$ comparisons. Washington University Report, AMCS-73-9 (1973)
11. Paterson, M.: Progress in selection. In: Karlsson, R., Lingas, A. (eds.) SWAT 1996. LNCS, vol. 1097, pp. 368–379. Springer, Heidelberg (1996)