COMPUTER SCIENCE COLLOQUIUM

Gianni Franceschini Department of Computer Science University of Pisa

Thursday, November 22, 2007, at 10:15 IMADA's Conference Room (former NIS-lab Conference Room)

Order Statistic Problems on Suffixes

ABSTRACT

In an Order Statistic Problem we are given a set S of n elements drawn from a total order (U,<) and a rank set R of integers drawn from {1,2,...,n}. The objective is to find the k-th smallest element in S for any rank k in R (according to the total order (U,<)). Given their basic importance in Computer Science, various instances of Order Statistic Problems have been widely studied in the past. Usually, such studies have been carried out within the settings of the comparison model and assumed simple unidimensional objects (i.e. comparable in O(1) time) as input elements. Probably the most renown results on the subject are the two classic papers [Blum, Flovd, Pratt. Rivest, Tarjan, STOC 1972] and [Blum, Floyd, Pratt, Rivest, Tarjan, JCSS 7, 1973] where the authors established O(n) worst-case time complexities for the median selection and generic selection problems. In this talk we focus on Order Statistic Problems on suffixes, that is where the input set S contains the suffixes of a given sequence T of n elements (again, drawn from (U,<)). Problems on suffixes have practical and theoretical relevance in Computer Science in connection with various fields e.g. Text Indexing, Bioinformatics etc. The Suffix Sorting problem has been widely studied in the past and its asymptotic complexity is well-known (O(n log n) time in the worst case). This is not the case for other fundamental Order Statistic Problems on suffixes, like Generic Suffix Selection, whose asymptotic complexity has been unknown until this year. Or Maximum (or Minimum) Suffix Selection, whose precise complexity (i.e. considering the constant factors) has been improved only recently from the (3/2)n upper bound of twenty years ago. These two last studies will be presented in this talk. More precisely, we will describe the first O(n) time worst-case solution to the Generic Suffix Selection Problem (appeared in the paper [Franceschini Muthukrishnan, STOC 2007]) and a (4/3)n time worst case solution to the Maximum Suffix Selection Problem (based on the paper [Franceschini, Hagerup, 2007]).