Seminar on Advanced topics in data structures

Fall 2002/2003

Haim Kaplan
phone: 03-6408814
email: haimk@math.tau.ac.il

October 12, 2002

We shall focus on the following topics in data structures

1. Answering lowest common ancestor queries
2. Suffix trees and suffix arrays
3. Applications of suffix arrays to data compression

Depending upon the number of students participating we may touch other topics as well.

How to get in: To participate in the seminar you must choose a paper and register it with me by Thursday (17/10) of this week. You can get copies of the papers to examine either by coming to my office or at the library. Some may be available through links from the course web site. Once you decide that you want to participate and picked a paper please come to register the paper at my office (its also possible to do that by phone or email if you cannot come in person). Please note that as soon as you register your chances to get in with the paper that you want are higher, in particular if there are a lot of students in class. Papers in the list below are ordered more or less in the order in which I would like them to be presented so you can estimate the approximate week of your presentation.

Once registered you must do the presentation you commited to!

Your grade: You'd have to attend all meetings and actively participate in the discussion so as to make sure you understand the material. The papers are strongly related. You should follow the material covered before your presentation and connect your presentation to it. (That does not necessarily mean that you do not have to repeat crucial definitions etc.) You are encouraged to ask questions and you get credit for it. To encourage you to ask everything you like I guarantee not to take any credit from the speaker even in case that he or she does not answer well. So there is absolutely no way you can hurt your friend by asking, and you may hurt yourself if you refrain from asking.

Participation worth up to 15% of your final mark. The other 85% will be determined by the quality of your presentation.

Guidelines for preparing your presentation:

1. All the papers on the list are theoretical and you may have to spend quite a bit of time in order to understand your paper. Good understanding of the paper is the key for a good presentation. So start working early and try to gain a good understanding of the results.
2. Some of the papers on the list are related to others. So attending the other lectures would certainly help you when you work on your paper.
3. You are not restricted to use any particular media for your presentation. Clear conventional blackboard talk certainly can gain maximum credit. Remember that sometimes a picture worth 1000 words. Media such as slides or computer slides is particularly useful when you want to show pictures and animate operations since doing that on the board may take too much time. Your time is limited so you have to use it carefully.
4. Its likely that in the time you have you will not be able to cover everything that is written in your paper. Depending on how crowded the seminar would be, it may not be possible to give you extra time. So plan the talk carefully, you need not cover everything in the paper but what you cover should be transfered clearly. Don't cover stuff that is in your paper but has already been covered by previous speakers.

References

[1] M. Farach-Colton M. A. Bender. The LCA problem revisted. In 4^{th} *LATIN*, pages 88–94, 2000.

[2] M. Farach-Colton, P. Ferragina, and S. Muthukrishnan. On the sorting-complexity of suffix tree construction. *Journal of the ACM*, 2000.

[3] U. Manber and G. Myers. Suffix arrays: a new method for on-line string searches. *SIAM Journal on Computing*, 22(5):935–948, 1993.

[4] E. Fiala and D. H. Greene. Data compression with finite windows. *Communications of the ACM*, 32(4):490–505, 1989.

[5] M. Rodeh, V. R. Pratt, and S. Even. Linear algorithm for data compression via string matching. *Journal of the ACM*, 28(1):16–24, 1981.

[6] G. Manzini. The burrows-wheeler transform: Theory and practice. In *Proc. 24th Int. Symposium on Mathematical Foundations of Computer Science (MFCS '99)*, volume 1672 of *LNCS*, pages 34–47. Springer, 1999.

[7] J. I. Munro and V. Raman. Succint representation of balanced parentheses, static trees and planar graphs. In *Proc 38th Symposium on Foundations of Computer Science (FOCS)*, pages 118–126. IEEE, 1997.

[8] J. I. Munro, V. Raman, and S. S. Rao. Space efficient suffix trees. *J. of Algorithms*, 39:205–222, 2001.

[9] C. R. Aragon and R. G. Seidel. Randomized search trees. In *Proc. 30th FOCS*, pages 540–545. IEEE, 1989.

[10] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, pages 37–49, 1998.