

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, May 06, 2003
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Arnab Chakraborty
Stanford University

An attempt at Optical Character Recognition for Bengali

In this talk I shall discuss my attempts to perform Optical Character Recognition for machine printed Bengali characters. I shall go over the general algorithm quickly and focus on certain particular features in more detail. These will include looking closely at the noise model of flatbed scanners (like the ones we have in our dept), and how a Markov-like model works (or rather fails to work) using bigram probabilities. The discussion will hardly involve any mathematical treatment.