A. V. GERBESSIOTIS
FEBRUARY 20, 2006
PS 3 (45 points)

CIS 750
SPRING 2006
Due:March 6

New Jersey Institute of Technology

Problem 1. (40 points)

Read the paper on the **Google Cluster Architecture** by Barroso, Dean, and Holzle. A link to the paper is available in section C3.1 of the Web-page and it is the first item there.

Try to answer after reading the paper the following questions whose answers are in the paper or may need to be further thought upon. You are given 2000 machines (PCs). Using the data available in the paper answer the following questions in a way that are fully compatible with the paper. Read all the questions first since some answers (eg. for (c)) might depend on the answers/solutions that you suggest to some other questions (eg. for (a)).

(a) How would you organize the 2000 machines? How many racks are you going to use? How many network switches? Other resources? Explain. (20 points)

(b) What is the collective memory (RAM), disk space (Gbytes) of these resources? (15 points).

(c) Using data from other papers you have already read for this course, how many of these machines will be file servers, how many index servers, and how many query machines? How many queries can the structure support and how much information can it index (in number of documents, and Gbytes). (10 points)