The Sixth DIMACS Implementation Challenge:

Near Neighbor Searches

`http://www.dimacs.rutgers.edu/Challenges/Sixth/`

Call for Participation

In conjunction with its Special Year on Massive Data Sets, the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) invites participation in an Implementation Challenge focusing on **Near Neighbor Searches**. The Implementation Challenge will take place between February 1998 and January 1999. Participants are invited to carry out research projects related to the problem area and to present research papers at a DIMACS workshop to be held early in 1999. A refereed workshop proceedings will be published as part of the AMS-DIMACS book series.

Finding near neighbors according to various distance measures arises in many application areas, including but not limited to finding similar web documents, images, DNA sequences, lines of text, or audio and video clips, and more generally in areas of statistics, data mining, information retrieval, pattern recognition and data compression. The goal of the Implementation Challenge is to determine how algorithms depend on the structure of the underlying data sets, as well as to determine realistic algorithm performance where worst case analysis is overly pessimistic and probabilistic models are too unrealistic. We invite research projects in the following two directions.

Providing Interesting Data Sets and Distance Measures. We invite researchers from various application areas to provide interesting data sets for near neighbor problems. Contributions could consist either of sample data sets from a true application or of realistic instance generators resembling practical data sets. Our goal is to construct a library of test problems that will be available for study both during and after the Challenge.

Developing Near Neighbor Algorithms. Neighbor algorithms may be developed either for specific application domains or for more general spaces. Algorithms may aim to find the nearest neighbor, several nearest neighbors, or approximate neighbors. Projects may involve either public domain or proprietary codes.

DIMACS Support. DIMACS facilities will provide a clearing-house for exchange of programs and data and for communication among researchers, including a collection of benchmark instances and evaluation criteria for algorithms. A steering committee offers the participants advice and support in their projects throughout the duration of the implementation challenge. DIMACS can provide neither financial support for research projects nor machine cycles for the experiments.

How to Participate. All information for the Challenge, including a more detailed "General Information" file, will be available on the Challenge web site (above), as well as through ftp. To register for the Challenge or for more information, please send mail to **challenge6@dimacs.rutgers.edu**.

Steering Committee. A committee of DIMACS members will provide general direction for participation in the Implementation Challenge. Committee members are Jon Bentley, Bell Labs, Ken Clarkson, Bell Labs, Michael Goldwasser (coordinator), Princeton University, David Johnson, AT&T Labs, Cathy McGeoch, Amherst College, and Robert Sedgewick, Princeton University.