Title : Nearest Neighbor Searching
Thematic : Computational Geometry
Institute: INRIA
Center : Centre de Recherche de Sophia Antipolis Mediterranée
Location: Sophia Antipolis
Team: Geometrica, http://www-sop.inria.fr/geometrica
Adviser: Mariette Yvinec, Mariette.Yvinec@inria.fr
Team leader : Jean-Daniel Boissonnat, Jean-Daniel.Boissonnat@inria.fr

Context :
Given a set of point P in a metric space and a query point q, the nearest neighbor $nn(q)$ of q in P, is the point of P that is closest to q. If ϵ is a small positive constant, an ϵ-approximate nearest neighbor of q in P, is any point of P whose distance to q is at most $(1+\epsilon)$ times the distance between q and $nn(q)$.

Given a set of point P in a metric space, the nearest neighbor searching problem or its approximate version, consists in building a data structure such that for any query point q, the nearest neighbor or an approximate nearest neighbor of q in P can be retrieve efficiently, i.e. more quickly than through an extensive visit of all points in P.

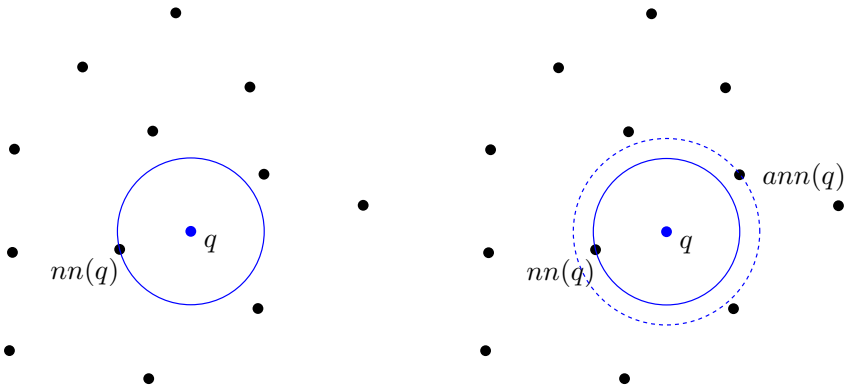

Figure 1: Nearest neighbor and approximate nearest neighbor.

The nearest neighbor searching problem is of major importance in a variety of applications where data are represented as points in a high dimensional metric space. The metric is designed such that the distance between two points measures the similarity (or rather the dissimilarity) between the corresponding data. Examples of such applications are databases, data mining, image and video databases, machine learning, clustering, quantification and data compression, pattern recognition, statistic and data analysis.

Work program:
The proposed work consists in a theoretical and experimental study of different

methods for nearest neighbor and approximate nearest neighbor searching. The set of tested methods will include at least the famous ANN algorithms for approximate nearest neighbor searching and a method based on a hierarchy of Delaunay graphs which provide exact nearest neighbors. Time permitting the internship student will experiment about coupling these methods with dimension reduction methods such as the Jonhson Lindenstrauss random projection method.

Tools:
PC Linux or Mac,
C++ language,
CGAL library `http://www.cgal.org`

Prerequisite that would be highly appreciated:
Basic notions of computational geometry, especially Delaunay triangulations and Voronoi diagrams.

Références:
1. *Computational Geometry: Algorithms and Applications*, M. de Berg, O. Cheong, M. van Kreveld, M.H. Overmars, and O. Schwarzkopf, third edition, Springer-Verlag 2008.
2. *An optimal algorithm for approximate nearest neighbor searching in fixed dimensions*, S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman and A.Y. Wu.