Summary of Key Readings
DJ Oneil  Esten Rye  (G01)

Approach:
Our project focus (at this point) is on the nearest neighbor problem. We are currently surveying the various methods used and some of the other topics related to our project. The papers approach the problem by finding the real solution or an approximate solution. Because spatial data is very large and computationally expensive and new technologies and applications, such as wireless sensor devices, provide huge amounts of streaming data with little computational capacity, the approximation techniques seem to be the most immediately relevant. Approximation techniques, for the most part, use statistical methods and heuristics. At this point, we haven’t found any papers that examine the possibility of combining these techniques. Our tentative approach is to review the various approaches proposed, identify the shortcomings of these individual approaches, and see where/whether synergetic relationships can be established to mitigate these shortcomings. We’d also like to examine how well some specific geographical heuristics would decrease the search space of GIS problems.

Key Sources:
Our key sources for this are mainly papers and articles from various online libraries such as the ACM Digital Library and the IEEE Explore Library. We also searched for papers and articles on CiteSeer and Wikipedia. We have also searched Amazon for books, but have found many of them to be very expensive and are currently searching local libraries for copies. The list of our currently compiled sources is listed below with brief summaries of the main ideas addressed.

Algorithms and Data Structures:

Nearest Neighbor Problem

Comments:
Given n points and a query point, the nearest neighbor problem finds the closest point to the query point (O(n)). This problem is usually extended to find All-k-nearest-neighbors where the k nearest neighbors of every point in the set are found. In spatial database problems, the nearest neighbor problem is usually used to find the k-nearest-neighbors of a single query point. There is also some research in finding the nearest neighbors over time where the elements are moving.

Sources:
  - Provides problem definition.
  - Presents the basic algorithm.

- Describes an algorithm for ranking spatial objects relative to some query object.
- The algorithm can be incremental in that it can be reported one by one.
- NOTE: in this algorithm, k must be fixed in advance*


- Provides a concise problem definition.


- The multi-step algorithm is intended for cases where there is a high-dimensional search space. The idea is to create a smaller (more accurate) set of candidates during the filter step, and then finish the search in a refinement step.
- The shortcomings of the state-of-the-art knn search (1996) are discussed.

Chavez, et al… “A Fast Algorithm for the All k Nearest Neighbors Problem in General Metric Spaces”

- Provides background research information on the topic; brute force, using Voronoi diagrams, and Delaunay triangulation.
- Proposes faster method that uses an index that can satisfy range queries. Using a range query, it find the \( n \) elements performing \( \gamma n \) distance computations, with \( 0 \leq \alpha \leq 1 \) depending on how good the index searches the space. An \( \alpha \) close to 0 is a costly and an \( \alpha \) close to 1 is cheap. The authors’ method becomes \( O(n*\sqrt{kn}) \) when \( \alpha \) is close to 1 with \( O(kn) \) space required.


- Proposes a method for finding the k nearest and reverse k nearest neighbors of moving objects. The motivation of the paper is the increasing popularity of new wireless technologies such as remote sensor networks. It uses previous positions of an object to estimate future positions.


- Proposes the first approach to the Rknn problem on arbitrary metric spaces where k is specified at query time.


- Compares two algorithms for using an R-tree to get the k nearest neighbors.

o Describes an algorithm for the aggregate nearest neighbor problem.
o Given a set of n locations, the aggregate nearest neighbor problem finds a location (of some type) that minimizes the aggregate distance the n locations would have to travel to get to the location.

**Spatial Range Query**

**Comments:**
Given an n-point set and a set of points defining a query geometric shape, a spatial range query finds all points within the defined shape.

**Sources:**
o Briefly defines the problem.
o Briefly describes attempts to model spatial problems using existing DBMS implementations and how they are inadequate.
o Examines the common principles and needs of spatial problems and proposes a small set of constructs.
o Shows how these constructs can be easily implemented in existing DBMS implementations.
o Investigates the data types commonly used for range queries.
o A survey of the algorithms commonly use for spatial queries.
o Proposes a new data structure for range searching in secondary memory.
o Also describes data structures for use in main memory and shows that it is faster than previously used data structures.
o Proposes a way to perform spatial queries using sensor network information and devices within the low power and bandwidth constraints of the devices.
o Proposes an approximation technique for spatial queries.
o Discusses some of the drawbacks of current uses of histograms and/or sampling.
**Heuristics:**

**Heuristics and Spatial Databases**

**Summary:**
The search space for a spatial database can be huge. The purpose of spatial database heuristics is to significantly decrease this search space.

**Sources:**
  - Combines research in combinatorial algorithm development and geographic databases.
  - Outlines the shortcomings of combinatorics and optimization (C & O) and GIS, and explains how/why it is good to combine methods.
  - Discusses why adding heuristics to GIS is difficult (regional/national differences in classification, developing areas/countries, etc…)
  - Uses pieces of the polygons to decide whether they intersect.
  - NOTE: not sure if this is useful yet.
  - Address issues of using heuristics to reduce the response time of spatial queries.
  - Gives a method of approximating the query.

**Zoning**

**Summary:**
Zoning is a system of land-use regulation. It is often used in urban planning to separate residential and commercial areas (or zones). It defines the acceptable activities and densities of those activities that can be performed on specific lots. In regards to nearest-neighbor queries zoning could be used to narrow the search area.

**Sources:**
- Wikipedia Sept 30, 2006 "Zoning"
  - Provides a general description of Zoning
  - Provides examples of various zoning schemes.
**Minimum Bounding Box**

**Comments:**
Minimum Bounding Boxes are often used in algorithms as a heuristic to improve the runtime. A Minimum Bounding Box is a closed volume containing the set of the unions of its contents. It is often used to eliminate, with a large amount of error, chunks of the data from being considered in an intersection.

**Sources:**
- Barequet and Har-Peled, "Efficiently approximating the minimum-volume bounding box of a point set in three dimensions", Proceedings of the 10th annual symposium on Discrete algorithms, Baltimore, MD, 1999, pgs 82-91
  - Presents two efficient algorithms that approximate a minimum bounding box of n points in R3
  - Analyzes the performance of bounding box heuristic in terms of aspect ratio and scale factor.
  - Evaluates the validity of the performance increase due to the use of a bounding box heuristic.
  - Brief description of what a bounding box is.
  - Provides an O(n4) time combinatorial algorithm for computing the smallest spatial configuration of n dimensions determined by n+1 points in a space of dimension equal to or greater than n.
  - Provides a general definition of bounding boxes and their uses
  - Paper sets out to prove bounding boxes improve performance in the worst case.
  - Provides an overview of approximation-based query processing and the role of minimum bounding boxes in that endeavor.
  - Defines and describes several bounding volumes of various shapes.
  - Introduces a new Minimum bounding rectangle (MBR) based spatial access method (SAM): QSF
  - Demonstrates current MBR-based SAMs are unable to cope with the increasing dimensionality of data.
  - Uses Experimental data to show that QSF is an improvement to MBR-based SAMs.
http://en.wikipedia.org/wiki/Bounding_volume
  ○ Good General definition of the topic.
  ○ Describes the uses and common types of bounding volumes.
• Papadias, Sellis, Theodordis, and Egenhofer. "Topological relations in the 
world of minimum bounding rectangles: a study with R-Trees". Proceedings 
of the 1995 ACM Sigmod international conference on Management of 
data. 1995. pgs 92-103.
  ○ Paper explains the topological relations conveyed by the objects held 
inside Minimum Bounding Rectangles (MBR)
  ○ Explains how to construct an R-Tree to support query operations for 
MBRs

Zip Codes

Summary:
Zip Codes are the postal codes used by the United States Postal Service. Zip 
codes as we know them today were made mandatory in 1967. The first digit of a 
five digit zip code describes the general region of states. The second and third 
digits describe a more specific region of those states. The final 2 digits describe a 
location at the town or city level. In regards to the Nearest-Neighbor problem, zip 
codes can be considered as labels for bounding boxes that contain all the point 
addresses in the United States.

Sources:
• Wikipedia Sept 30, 2006 "Zip Codes"
  • Describes the history of Zip Codes
  • Describes the structure and allocation of Zip Codes
  • Gives a general description of other uses and pop culture references