Boston University
CAS CS 562: Advanced Database Applications
Homework #2
Spring 2018

Due Date: March 16, 2018 at 11:59 PM
Please, use gsubmit to submit your homework.

Problem 1. (Temporal Indexing)

Consider the following Time-Interval query: Given a time period $T = [t_s, t_e]$, find all the records that have lifespan that intersects T. That is, find all the records that were alive at least one time instant during the period T. Provide an algorithm and explain what is the cost of this algorithm in terms of number of I/Os. Hint: Read the Snapshot Index paper.

Problem 2. (Spatio-temporal Databases)

A Continuous Spatio-temporal Query returns a set of $< \mathbf{R}, \mathbf{T} >$ pairs, where each R is a spatial object in the result and T is the validity period for this object. The termination time TT of this query is also specified by the user. So, the answer to this query are all the pairs that satisfy the query from now until TT and can be one or many pairs. Consider now a moving window (moving range) query, where you are given a range query that moves with fixed speed on a specific trajectory (car moving on a road for example). Notice that trajectory can consist of a number of different connected line segments. Find an efficient way to answer this query assuming that the objects are stored in an R-tree. Hint: Check the Time-Parameterized Queries paper.

Problem 3. (Time Series)

The GEMINI approach is used to index time series data for similarity retrieval. One approach to do the transformation to another lower dimensional Euclidean space and use the GEMINI framework is the following: Given a time series of size N you can map it to a vector of K values by creating sub-time series (sequences of consecutive values) of size $\frac{N}{K}$ and take the average of each sub-time series. This approach is called piece-wise aggregate approximation (PAA).

Prove that if the distance function between the original time series is the Euclidean distance and you use the PAA transformation, the lower bound in the transformed space holds. Namely, if X and Y are the original time series of size N and X' and Y' are the transformed time series in the K-dimensional space and D_i is the Euclidean distance in the $i-th$ dimensional space, then $D_K(X', Y') \leq D_N(X, Y)$.