

Chemistry 696D – Analytical Informatics

Exam 3 – Due by Friday May 2, 2003

The MySQL server is at `miner.chem.purdue.edu` on port 3306 and you have a user name and password for this system that can be used in connecting to the server using either a GUI (SQLyog) or Perl.

Turn in an electronic copy of all your work: plots, source code and explanations. Perl and R programs have to run on a standard setup (with appropriate drivers and libraries) to obtain credit for a problem. Database can be done locally on your own MySQL instance if you prefer, but the final database has to be created on the class server to obtain credit.

- (5) Describe what is meant by the term “relational data base”. Why are they used in informatics systems, give advantages, disadvantages, and alternatives.
- (5) Write a single Perl regular expressions which can recognize all of the following strings and put the first number past the comma in \$1 and the last word in \$2 (in strings.txt):
 - “1 T, 2000 0 0 4 R”
 - “2 A, 200 0 0 3 R”
 - “3 G, 1000 100 12 1 Tele”
- (10) Create the compound database as described in the `compound.sql` file in your database space on the class MySQL server. Use the `db2sql.pl` program to populate the database. Use the `add_smi.pl` to populate the SMILES field.
- (10) Write an SQL statements which:
 - Count the number of compounds in the database
 - Compute the average number of heavy atoms
 - Delete all the molecules which have BENZ at the start of their name
 - Add a new synonym for acetone in the name table
- (40) Add molecular weights to the database:
 - Extend the database to hold exact and average molecular weight
 - Write a Perl program which computes both molecular weights for each molecule
- (30) Write a Perl program which queries the database for molecular weights and calls an R program which produces a histogram of the distribution of average molecular weight. Have the R program save the plot as a PDF file. [hint: see `help(“device”) in R]`