Welcome to CMSC 828U

- Administrivia
- Overview
- Tentative syllabus and schedule
  - Topic 1: Data Modeling
- Glossary of terms
- Project 1: Navigational Queries on NCBI data sources.
Administrivia

• Who am I? Why do I want to teach this course?
• Class meeting time and place.
• Unofficial GA – Adam Woei-Jyh Lee adamlee@cs.umd.edu
• Who are you and what do you want from this course?
• Format
  • Lectures lead by me (first half).
  • Discussions lead by the Ph.D. students (second half).
  • Programming projects and exercises (3 or 4).
  • Final project - individual or group.
• Grading
  • Ph.D. - Mid-term (30), Final Project (30), Presentation (20), Programming projects (20).
  • Undergrad - Programming projects (40), Final project (30), Mid-term (15), Class participation (15).
• Questions?
Overview

• What will be covered in depth?
  – Data modeling (focus on genomics).
  – Data integration (focus on genomics).
  – Why the focus on genomics?

• What will be covered (superficially)?
  – Workflows, ontologies, biological pathways, protein data resources, text analysis and text mining, …

• What will not be covered in this class?
  – A LOT!!!

• What if you don’t have domain knowledge?
• What if you don’t have data management knowledge?
• What if you don’t have programming skills?
Syllabus and Schedule

• Week 1:
  – Introduction.
  – Project 1 Assigned.
  – Due at the end of Week 4? Individual or group?
• Week 2:
  – Introduction to bioinformatics courtesy of Dr. Nathan Edwards.
• Weeks 2 - 4: Data Modeling
  – Readings are on the course site.
  – What should be captured by a data model?
  – Importance of biological data modeling and templates/standards.
  – Review of data modeling using ER, relational, XML (graph).
  – GMOD/Chado and GUS.
  – Case studies of data modeling – BIP-Splice and BIP-Marker.
  – Other case studies contributed by students.
  – Pros and cons of the two approaches.
  – Exercise on modeling – at least two groups; class presentation of results at the end of Week 5?
Syllabus and Schedule

- **Weeks 5-8 Data Integration:**
  - Readings will be posted on the Web site.
  - Modeling/Programming exercise; due at the end of week 10?
- **Week 9**
  - Mid-term.
- **Weeks 10 - 14**
  - Ph.D. students lead discussions.
  - Ontologies; Workflow; Information retrieval and text mining; …
  - Please post topic and readings on the Web site (Adam) by the end of Week 6.
- **Week 12**
  - Proposal for final project.
- **Week 15? 16?**
  - Class presentation of final project.
  - Final report.
Glossary of terms (genomics)

- Genome
- Base pairs
- DNA, mRNA
- Sequence/transcript
- intron, exon, codon
- Alignment (local and global)
- SNPs
- Amino acids, proteins,
- Genes and gene families
- Alternative splicing
- Gene expression
Project 1: Navigational queries