

477: Final Project

This year's project theme is genomics. You will choose one among four possible problems, to be elaborated in class, of genomic data compression and processing.

Criteria for project assessment:

1. Experimental: How well does your scheme do on real data? How does it perform compared to current practice?
2. Algorithmic:
 - Complexity: How simple is the implementation?
[running time, required memory size, etc.]
 - Elegance: Is your scheme graceful?
[hard to define formally, you know it when you see it, Ziv-Lempel schemes are an example]
3. Analysis:
 - Modeling: How are you modeling (if you are) the data? How realistic is this model? What features of the problem is it incorporating? What features is it neglecting?
 - Analysis: What kind of performance can you guarantee? How close is it to optimum?
 - Universality: How universal is your scheme? Can it be guaranteed to be making sense, performing well - close to optimally? - regardless of the model, or for a whole family of models? Or perhaps in some appropriate individual-sequence sense?

Grading:

If you score highly under:

- either the first or the third criterion, I would consider that to be a really good project.
- any two of the three criteria, I would consider that a great project.
- all three criteria, I would consider that to be an amazing project.