1. Course number and name
   CptS 471: Computational Genomics

2. Credits and contact hours
   3 credits, 3 lecture hours

3. Instructor’s or course coordinator’s name
   Ananth Kalyanaraman

4. Textbook, title, author, and year
   There is no required textbook for this course. All course material will be based on
   lecture notes, handouts, and classroom scribes. All materials will be made available to
   the students over the course of the semester, on the course website:
   <https://eecs.wsu.edu/~ananth/CptS571/>.
   Other supplemental materials
   The following textbooks will serve as reference:
   1584884061.
   Durbin et al. 1999. Biological Sequence Analysis: Probabilistic Models of Protein

5. Specific course information
   a. Catalog description: Computational Genomics
   b. Prerequisites or corequisites: CPT S 223 with a C or better or CPT S 233 with a
      C or better; CPT S 350 with a C or better; certified major in Computer Science,
      Computer Engineering, Electrical Engineering, or Software Engineering.

6. Specific goals for the course
   By the end of the course, students will be able to
   - Demonstrate a fundamental understanding of the core algorithms, techniques and
     data structures used in the specialized area of computational biology and
     bioinformatics (1a, 1b, 1c, 6a).
   - Formulate and/or model a real-world biological problem into a well-defined
     computer science problem (1b, 1c).
   - Design and analyze efficient algorithms and data representations for problems in
     computational biology and bioinformatics (1a, 1b, 1c,1d, 1e, 2a, 2b, 2g, 6a).
   - Implement programs, and test and evaluate real-world applications in
     computational genomics (2a, 2b, 2e, 2g, 6a, 6b, 6c, 6d).
   - Effectively document and communicate empirical results in a manner that is
     consistent with scientific practice, including providing reasoning and rationale in
     written documents (3a, 3b, 3c, 3d, 3e, 6b, 6c, 6d).
   - Apply and extend the algorithmic concepts/techniques and data structures learned
     in the context of computational biology problems, into other scientific domains -
     e.g. text mining, pattern matching, speech recognition (1e, 7a).
• Demonstrate an interdisciplinary vocabulary that encompasses basic and emerging research themes within bioinformatics and the life sciences (1e, 6a, 7f).
• Function effectively as part of a team (5b, 5c, 5g).

7. Brief list of topics to be covered
• Approximate string matching: Dynamic programming, sequence alignment, edit distance
• Exact string matching: Compacted tries, Suffix trees and suffix arrays
• Probabilistic modeling for biological sequence analysis: Likelihoods, Markov chains, Hidden Markov Models
• Applications: Genome sequencing, genome annotation, read mapping, gene identification, clustering