LANE DEPARTMENT OF COMPUTER SCIENCE AND ELECTRICAL ENGINEERING

CS 591A/791B: String Algorithms[1]

Fall 2009

Instructor: Don Adjeroh, Room ESB 937; Tel: 304-293-9681; email: don@csee.wvu.edu
Office Hours: Tuesdays 11:00 am– 12:30pm; Thursdays 11:00-12:30pm
Course Schedule: Tuesday and Thursday 5:00 – 6:15pm, Venue: ESB Rm 449
Course homepage: http://www.csee.wvu.edu/~adjeroh/classes/cs791b

Purpose

A string is basically a sequence of symbols from a defined alphabet. Analysis and manipulation of strings have become an essential part in various aspects of our day-to-day activities, from web search engines to computational biology to analysis of music sequences, to efficient IP classification and lookup in fast networks. An understanding of the special nature of strings and the algorithms and data structures required for their efficient analysis is therefore important for various applications. Such algorithms and data structures are also of independent theoretical interest in their own right.

In this course, we will study algorithms on strings, from traditional combinatorial pattern matching to recent problems, such as suffix sorting, and string embeddings. Our emphasis will be on the data structures and algorithms required, their analysis, and optimal constructions. We will also consider some modern applications of string algorithms in different fields, such as in computational biology, computational geometry, computer vision, web search engines, music synthesis and analysis, speech recognition, data compression, IP lookup and classification in large networks, natural language processing, intrusion detection, image-based pattern matching, text mining, etc.

We will look at string algorithms from four major perspectives:

String pattern matching
String index structures
Compressed index structures
Periodicities and symmetries in strings

Expected Learning Outcomes:

At the end of the course, students are expected to:

- Have understood the concept and properties of a string,
- Have understood, and be able to construct simple algorithms for string pattern matching
- Be able to prove complexity results on classical string algorithms
- Be able to construct compact data structures for string representation and analysis
- Have understood the fundamental algorithms used in compressed string representations and analysis.
- Have understood periodicities and symmetries in strings

References

1. Required Text: Gusfield D, *Algorithms on Strings, Trees and Sequences*, Cambridge University Press, 1997.
2. Recommended Text: Smyth W., *Computing Patterns in Strings*, Addison-Wesley, 2003
3. Others: Materials to be handed in class

Pre-requisite: Undergraduate courses in Algorithms, and Data Structures (CS 221 or CS 420; and CS 210); or Consent

Assessment:

	Worth	Release Date
Weekly quiz	20%	
3 Assignments	15%	Late September
	15%	Late October
	15%	Mid November
Project & Presentation	35%	Mid October
Class Participation	5% extra on Quiz	

Grading
A ≥ 85
B 75-84
C 60-74
D 50-64
F < 50

[1]CS791B and CS591students will attend the same lectures. However, CS791B students will be given additional (and more advanced) problem sets and homework.

Others

Academic Honesty

Students are encouraged to discuss class topics and analyze problems among themselves. However, copying assignment solutions or written reports (or parts of) is strictly forbidden. Also, while the Internet could be used as a research tool, copying materials verbatim from the Internet is plagiarism, and will not be tolerated in this class. Please, be aware that your submitted materials may be compared with each other, or with materials from the Internet during the evaluation.

Social Justice Statement

West Virginia University is committed to social justice. I concur with that commitment and expect to foster a nurturing learning environment based upon open communication, mutual respect, and non-discrimination. Our University does not discriminate on the basis of race, sex, age, disability, veteran status, religion, sexual orientation, color or national origin. Any suggestions as to how to further such a positive and open environment in this class will be appreciated and given serious consideration. If you are a person with a disability and anticipate needing any type of accommodation in order to participate in this class, please advise me and make appropriate arrangements with Disability Services (293-6700).

LANE DEPARTMENT OF COMPUTER SCIENCE AND ELECTRICAL ENGINEERING

CS 591/791B: String Algorithms

Fall 2009

Instructor: Don Adjeroh
Room ESB 937
Tel: 293-0405 ext. 2567
email: don@csee.wvu.edu

Weekly Course Schedule

Week	Starting	Topic	Notes
1	Aug. 24	Introduction **String Pattern Matching**	
2	Aug. 31	Exact pattern matching	
3	Sep. 7	Exact pattern matching	
4	Sep. 14	Inexact pattern matching	
5	Sep. 21	Multiple pattern matching	**Assignment 1 ready** (due 2 weeks after)
6	Sep. 28	Multidimensional pattern matching Pattern matching with don't cares	Project topics ready (due week of Dec. 1)
7	Oct. 5	**String Index Structures** Suffix trees	
8	Oct. 13	Suffix trees Suffix arrays	
9	Oct. 20	The LCP preprocessing	**Assignment 2 ready** (due 2 weeks after)
10	Oct. 27	The LCA Problem	
11	Nov. 2	The LCA Problem Order-Preserving Hashing	
12	Nov. 9	**Compressed Data Structures** Compressed pattern matching	
13	Nov. 16	Compressed index structures Index structures in external memory	**Assignment 3 ready** (due after 2 weeks)
14	Nov. 23	**THANKSGIVING BREAK**	
15	Nov. 30	**Embeddings, Periodicities, Applications** Symmetries and periodicities in strings Applications	
16	Dec. 7	**DEAD WEAK** In-Class project presentations Course review	**Project reports due**
17	Dec. 14	**FINALS WEEK**	

Note that the above represents only an estimate of the weekly schedule. The actual date/week that a particular topic is discussed, and the specific topic sequence could vary during the semester.