Project Period: June 1, 2014 through May 31, 2015

Title of Project: Science Curriculum Development: A Course on Algorithms for Bioinformatics

GENERAL INFORMATION
Name: Natarajan Meghanathan
Email: natarajan.meghanathan@jsums.edu
Phone Number: 601-979-3661
Address: Mailbox 18839, Jackson State University, 1400 J R Lynch Street, Jackson, MS 39217

ACCOMPLISHMENTS and UPDATES

ACCOMPLISHMENTS

B.1 What are the major goals of the project?
List the major goals of the project as stated in the approved application or as approved by the agency. If the application lists milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion. "Goals" are equivalent to "specific aims."
Limit is 8000 characters, text only

The major goal of this project are to develop a senior-level undergraduate course called "Algorithms for Bioinformatics" as an elective course for Computer Science majors at Jackson State University. The course materials were developed during Summer 2014. The course was taught in Fall 2014. Course assessment was conducted during and at the end of the Fall 2014 semester.

B.2 What was accomplished under these goals?
For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results, including major findings, developments, or conclusions (both positive and negative); and 4) key outcomes or other achievements. Include a discussion of stated goals not met. As the project progresses, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments. "Goals" are equivalent to "specific aims."
In the response, emphasize the significance of the findings to the scientific field.

1) Major Activities: The major activities of the project are to develop the course materials (including the lecture slides for each module, the projects, question
banks), teach the course during the Fall 2014 semester, conduct assessment for the course and submit a proposal to the Jackson State University curriculum committee to list the course as CSC 423 Algorithms for Bioinformatics in the undergraduate catalog.

2) Specific Objectives: The specific objectives of this project are to develop the course materials focusing on specific algorithmic strategies for solving bioinformatics problems and teach them to students. Particular attention was given to developing the course materials based on the foundational knowledge the students had attained in the pre-requisite course (CSC 228 Data Structures and Algorithms) and sufficient review of the background materials were incorporated as part of the lecture materials so that students do not get lost. The course materials were organized in the form of modules: the initial set of modules introduced students to molecular biology and the basics of dynamic programming (a widely used algorithmic strategy for solving bioinformatics problems); the subsequent modules were designed according to the algorithmic design strategy employed and the bioinformatics problems studied. The algorithmic design strategies that were specifically covered in detail include: Brute Force, Greedy, Dynamic Programming, Branch and Bound, Divide and Conquer. The bioinformatics problems that were covered in detail include: Pair-wise sequence alignment, Multiple sequence alignment, Phylogeny, Motif finding, Clustering and String processing using Suffix trees. The hands-on projects were developed in such a way that students were able to relate the algorithms studied in class and their application to analyze large-scale biological data in the NCBI database.

The following were the course-level objectives: Each student who successfully completes this course should be able to:
CO-1: Apply graph theoretic principles and algorithms to analyze sequence data and identify biologically significant information
CO-2: Perform global and local pair-wise sequence alignments using dynamic programming
CO-3: Conduct multiple sequence alignment using efficient heuristic techniques
CO-4: Identify motifs and their starting positions in biological sequences
CO-5: Construct Suffix trees and use them for various string pattern matching problems
CO-6: Identify groups of functionally related genes using efficient clustering techniques
CO-7: Solve the large and small parsimony problems for reconstructing phylogenetic trees as well as sort by reversals to trace evolutionary history.
CO-8: Describe the algorithms behind the working of popular Bioinformatics database search software
CO-9: Use greedy and branch and bound strategies to efficiently solve combinatorial bioinformatics problems

3) Significant results: The lecture slides developed for the course as well as all the relevant course materials are posted online at
The course slides for each topic were developed in an illustrative fashion, with several examples, to make it easy for undergraduate students to understand the topic without much difficulty. We are sure that the materials developed for this course could be adapted at other institutions to teach an algorithms for bioinformatics course for computer science majors.

Two abstracts were published based on the work developed in this project. One abstract is more of a research work that originated from an idea that evolved during the development of a course module. This abstract is titled: A Greedy Algorithm to Determine the Longest Suffix Prefix Overlap of two DNA Strings and is now published in the proceedings of the ACM-Mid Southeast Conference, p.74, Gatlinburg, TN, November 2014. The second abstract summarizes the key aspects of the project and the development of the course materials. This abstract is titled: Design of an Algorithms for Bioinformatics Course for Computer Science Majors and is now published in the proceedings of the 79th Annual Meeting of the Mississippi Academy of Sciences, vol. 60, no. 1, p. 131, Hattiesburg, MS, February 2015.

A major finding was that students needed sufficient time to work through the solutions for the algorithms and the bioinformatics application problems covered in the course. Hence, the instructor developed take-home exams and quizzes that had questions that were detailed enough to test the comprehensive understanding of the subject matter. Moreover, the questions were individualized (each student had a different instance of the same question) to avoid any copying. As a result, each student has to thoroughly understand the subject matter in order to answer the questions.

4) Key Outcomes: The key outcome is that the course created a significant level of awareness among the Computer Science majors about the application of the algorithmic strategies and the algorithms (they had previously studied in the traditional Computer Science courses) to the Bioinformatics problems. Some of the students who took this class are now intending to pursue graduate studies in Bioinformatics. Also, the course now serves as one of the theoretical senior-level electives for the undergraduate computer science curriculum at Jackson State University. The development of the course materials also paved the way for the PI (who is also the instructor for the course) to expand his research base in the area of Bioinformatics. The research abstract published in the 2014 ACM Mid southeast conference is a pointer in this direction.

B.5 How have the results been disseminated to communities of interest?
Describe how the results have been disseminated to communities of interest. Include any outreach activities that have been undertaken to reach members of communities who are not usually aware of these research activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.
Reporting the routine dissemination of information (e.g., websites, press releases) is not required. If the Component is not designed to disseminate information to the public or conduct similar outreach activities, enter “Nothing to Report.” If the Component is designed to disseminate information or conduct outreach activities, report those activities here. Note that scientific publications and the sharing of research resources will be reported under Products.

Limit is 8000 characters, text only

We have developed a website http://143.132.8.23/cms/courses/CSC423-Fall2014/ to host all the course materials. Since the time of offering of the course, the course has created a significant level of awareness among the undergraduate and graduate students about the application of the traditional algorithms and algorithmic strategies to study bioinformatics research problems.

B.6 What do you plan to do during the next reporting period to accomplish the goals?
   Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.
   Limit is 8000 characters, text only

All the stated goals of the project have been already accomplished.

PRODUCTS

C.3 Technologies or Techniques
   Identify technologies or techniques that have resulted from the research activities. Describe the technologies or techniques and how they are being shared.
   If none, enter “Nothing to Report.”
   Limit is 8000 characters, text only

Nothing to report.

C.5b Other Products and Resource Sharing - Resource Sharing
   If the initial research plan addressed, or the terms of award require, a formal plan for sharing final research data, model organisms, Genome Wide Association Studies data, or other such project-specific data, describe the progress in implementing that plan. For sharing model organisms, include information on the number of requests received and number of requests fulfilled during this reporting period. If the sharing plan is fully implemented, provide a final statement on data sharing.
   If none, enter “Nothing to Report.”
   Limit is 8000 characters, text only

Nothing to report.
CHANGES

F.2  Actual or Anticipated Challenges or Delays and Actions or Plans to Resolve Them

Describe challenges or delays encountered during the reporting period and actions or plans to resolve them. Describe only significant challenges that may impede the research (e.g., accrual of patients, hiring of personnel, need for resources or research tools) and emphasize their resolution.
If none, enter “Nothing to Report.”
Limit is 8000 characters, text only

Nothing to report.

AWARDS, HONORS, SPECIAL RECOGNITIONS, AND OTHER INFORMATION

A. Awards, honors, and special recognitions received during the reporting period.

Nothing to report.

B. Service in Peer-review Panels.  List the names of key personnel who served in review panels during the reporting period.

a. Inside the Institution/Network

<table>
<thead>
<tr>
<th>Name</th>
<th>Name of Review Panel</th>
<th>Date (MM/DD/YY)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total # Investigators who served in Review Panels: 0

b. Outside the Institution/network

<table>
<thead>
<tr>
<th>Name</th>
<th>Review Panel, Agency</th>
<th>Date (MM/DD/YY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natarajan Meghanathan</td>
<td>HBCU-UP (ad hoc reviewer), NSF</td>
<td>02/02/2015</td>
</tr>
</tbody>
</table>

Total # Investigators who served in Review Panels: 1
Nothing to report.

Non-Federal Support
Provide information for other non-federal support that were newly obtained during the reporting period only. Do not include awards that were obtained prior to and remained active in the reporting period.

Source Types: FDN = Foundations, RI = Research Institutes, IND = Industry, PVAS = Professional and Voluntary Associations or Societies, SCC = State, County, City Funds, and OTH = Other

<table>
<thead>
<tr>
<th>Investigator Name (eRA Commons Name)</th>
<th>Source Type</th>
<th>Organization</th>
<th>Grant Title and Budget period</th>
<th>Total Award $</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Federal non-PHS Support
Provide information for other Federal non-PHS support that were newly obtained during the reporting period only. Do not include awards that were obtained prior to and remained active in the reporting period.

Source Types: DOD = Department of Defense, DOE = Department of Energy, DVA = Department of Veteran Affairs, NASA = National Aeronautics and Space Administration, NIST = National Institute of Standards and Technology, NSF = National Science of Foundation, OTH FED = Other Federal Agency

<table>
<thead>
<tr>
<th>Investigator Name (eRA Commons Name)</th>
<th>Source Type</th>
<th>Organization</th>
<th>Grant Title and Budget period</th>
<th>Total Award $</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Federal-PHS Support

Provide information for other Federal PHS support that were newly obtained during the reporting period only. Do not include awards that were obtained prior to and remained active in the reporting period.

Source Types: NIH-XXX = National Institutes of Health (where XXX are the initials of the Institute), FDA = Food and Drug Administration, HRSA = Health Resources and Administration Services, CDC = Centers for Disease Control and Prevention, SAMHSA = Substance Abuse and Mental Health Services Administration, OASH = Office for the Assistant Secretary of Health, and AHRQ = Agency for Healthcare Research and Quality

<table>
<thead>
<tr>
<th>Investigator Name (eRA Commons Name)</th>
<th>Source</th>
<th>Grant Title and Budget period</th>
<th>Total Award $</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PUBLICATIONS and PRESENTATIONS

I. PUBLICATIONS

Provide the following information for Published and In Press manuscripts. All relevant publications should be associated with the grant in your MyNCBI publication list.

A. Published Work
Publication type: PR = peer-reviewed manuscript, R = Review article, BC = Book chapter, B = Book, A = Abstract, OTH = Other

<table>
<thead>
<tr>
<th>Publication type</th>
<th>Did the publication cite the grant? Y/N</th>
<th>PMCID</th>
<th>Publication Date</th>
<th>Bibliographic Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. In Press
II. PRESENTATIONS

1) 
Author: Meghanathan, Natarajan*
Subject/Title of Presentation: Design of an Algorithms for Bioinformatics Course for Computer Science Majors
Event: 79th Annual Meeting of the Mississippi Academy of Sciences
Location: Hattiesburg, MS
Date: February 26, 2015

2) 
Author: Meghanathan, Natarajan*
Subject/Title of Presentation: A Greedy Algorithm to Determine the Longest Suffix Prefix Overlap of two DNA Strings
Event: ACM Mid Southeast Conference
Location: Gatlinburg, TN
Date: November 13, 2014