

November 7, 2017

1 Priorities

1.1 Educational mission

This activity will develop infrastructure for distributed processing, that complements the existing support for cloud computing. This infrastructure will enable us to educate the students with hands-on experience on algorithms that exploit parallel processing in a distributed mode. Such algorithms are in common use. (Here I need some numbers from reliable sources.)

We would like to extend our research in the learning of students of computer science about mathematical proofs [SM13, SM14, Smi16] to mathematical proofs about software executing in distributed systems, as are given in Lynch's Distributed Algorithms [Lyn96].

1.2 Visibility

By providing a facility for distributed processing in our department, and making use of it available to any member of the Connecticut State Universities, we increase the visibility of CCSU's CS department. Conceivably in the future we might extend the availability of this facility more broadly.

1.3 Research stature

We hope to use this capability to support our research in computational medicine. We have reached out to the Biomedical Science department, in connection with a planned graduate specialization in computational support for biomedicine. We hope that this facility will support cooperation between our faculty members. Moreover, we hope to support existing projects at the larger scale available with distributed processing, including the search for biomarkers in the relatively restricted contexts of blood tests, urine tests and breath tests. We hope to develop opportunities for commercial activity in the development of these tests, which correspond to manufacturable test kits. Because of the anticipated support for commercial activity, we hope to develop external funding.

2 instructions for proposal preparation and submission / Proposal Narrative

2.1 Cover

Faculty Rank of Principal Contact: Last Name: First Name: University: Department: Funding Request: \$ Is this a Joint Proposal? <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If Yes, please fill in information for co-proposers (add separate sheets as necessary): Name Rank Department Is this a Continuation Project? Yes x No If Yes, you must complete Appendix A.2 E-mail of Principal Contact: Phone Number of Principal Contact: Campus Address of Principal Contact:
--

Please select one disciplinary group category in which this project best fits:

☐ Fine Arts and Humanities ☐ Social Sciences, Business and Education
☒ Life and Physical Sciences, Mathematics, Computer Science, Engineering and Technology

Please select one research focus area in which this project best fits:

☐ Creation of new knowledge ☒ Application of disciplinary/multidisciplinary knowledge, methodologies and/or insights ☐ Production of creative works ☐ Research in student learning

Project Title:

ABSTRACT (Limit: 100 words)

Abstract

We wish to install a distributed programming environment based upon Hadoop, which will serve several purposes. First, we can teach students with hands-on experience of distributed computing. Second, we can support large datasets, and parallelization of suitable algorithms; this includes algorithms for computational medicine. We can provide more extensive support for our statistical calculations (which support biomarker discovery), and carry them out faster, in the distributed programming framework, provided we have multiple machines. Biomarkers are helpful in medical diagnosis. Kits that test for biomarkers are a possible product suitable for manufacture.

IRB/IACUC Statement

(If “yes” to either question please see Section 5, p. 3 of the program guidelines)

YES NO

- | | | |
|-------------------------------------|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Does your research involve human beings as research subjects? |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | Does your research involve vertebrate animals? |

Sign-Off Statement (To be signed individually by each faculty applicant.
Please add separate sheets as needed)

I hereby acknowledge my understand that the lack of compliance with the proposal format and other requirements spelled out in the CSU - AAUP Faculty Research Grant Guidelines for the Spring 2018 Competition may result in the proposal being disqualified without review.

_____ Signature of Permanent, Full-Time Faculty	_____ Date
_____ Signature of Permanent, Full-Time Faculty	_____ Date
_____ Signature of Permanent, Full-Time Faculty	_____ Date

Appendix A.3: BUDGET AND BUDGET JUSTIFICATION FORM

Table 1: **2018 - 2019 CSU - AAUP Faculty Research Grant**

Budget Item	Amount (No Cents)	Brief Justification
Faculty Stipend		
Support Services *		
Supplies and Equipment		
Travel		
Total		N/A

* For definition see Section 9.4 of the “Collective Bargaining Agreement between Connecticut State University, American Association of University Professors and Board of Regents for Connecticut State Colleges & Universities System, August 26, 2016 August 26, 2021”, Section 9.4, pp. 56-57.

2.2 Significance

We have a three-fold purpose:

- Provide distributed processing infrastructure for our own research in computational medicine
- Provide distributed processing infrastructure for teaching students about distributed algorithms
- Conduct research on biomarkers that is intended to lead to manufacturable test kits

2.2.1 Outline of Related Research

Odom et al. [?] developed a breath test for malaria.

2.3 Work Plan

The software infrastructure, a modification of Hadoop, has been developed by Roland DiPratti. The plan is to identify machines onto which we can install this software, to make an operating facility. We plan to use machines obtained separately from the grant. Then we plan to install and verify the installation of the modified Hadoop. Then we plan to install software used by the application, including Python and R, which are generally useful packages. Then we plan to install software used by the application, a survival analysis package of R, which is more focussed in purpose, though of interest to both insurance and medical applications. Then we plan to load and execute our first application, which is expected to be the biomarker software, which uses Python and R. We plan to apply the biomarker software to larger datasets than those to which it has been applied so far; this will be facilitated by the modified Hadoop software.

2.4 Outcomes and Reporting

It is certainly our intention to submit the results of our biomarker research to a journal such as BMC Bioinformatics (<https://link.springer.com/journal/12859>). This research is related to our previous research [AS13].

We intend to submit the results of our research in approaches to teaching distributed programming to ICER and Koli Calling, which have accepted our work previously.

2.5 Budget proposal

2.6 CVs

2.7 Joint proposal individual contributions and level of participation

2.8 Bibliography

2.9 Optional Appendices

3 proposal review criteria

- coversheet abstract
- signoff statement
- proposal narrative

References

- [AS13] Reda Ammar and Therese Smith. Developing time constraints in petri net models of biochemical processes via computation structure modeling. In *Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on*, pages 000034–000039. IEEE, 2013.
- [Lyn96] Nancy A Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.
- [SM13] Therese Smith and Robert McCartney. Mathematization in teaching pumping lemmas. In *Frontiers in Education Conference, 2013 IEEE*, pages 1671–1677. IEEE, 2013.
- [SM14] Thérèse Smith and Robert McCartney. Computer science students’ concepts of proof by induction. In *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*, pages 51–60. ACM, 2014.
- [Smi16] Thérèse Smith. Categories of conceptions of proofs by students of computer science. 2016.