November 7, 2007

Dear Members of the Faculty Development and Personnel Committee,

This is my narrative for my application requesting tenure and promotion to Associate Professor. In this narrative I hope to demonstrate to the members of the committee how I meet and exceed Pacific University's requirements for tenure and promotion.

Throughout this document I reference items that I have placed in my open file.

Teaching

I chose to work at a small liberal arts college because of the focus on student learning. I have always aspired to be an educator, and Pacific University provides me with a great environment for supporting, educating, and mentoring undergraduate students. During the semester, my main focus is on teaching my classes and being there for my students.

Over the last five years I have taught 33 courses (12 different courses), and supervised two internships and three independent research projects. Detailed reflections on all my courses can be found in my annual self-evaluations and my second year review.

I would like to focus here on my overall teaching philosophy, and the rewards and challenges of teaching Computer Science. I will also include references to materials in my file that demonstrate how I have matured as a teacher.

Teaching Computer Science

Teaching Computer Science at a small liberal arts college has many rewards, but is also quite demanding. The biggest reward for me is that I really get to know my students, and watch them grow and mature during their four years at Pacific. Most of the students who major in CS start off not knowing anything about Computer Science. They may have been using computers since they were in kindergarten, but they do not know how computers work, or how to develop computer applications. Then, on Senior Projects Day, they present amazing applications that they have developed, which incorporate knowledge that they have gained from the many different CS classes that they have taken (Databases, Operating Systems, Algorithms, etc.), as well as knowledge from other disciplines.

You see, Computer Science embodies the goals of interdisciplinary knowledge, because the applications developed in Computer Science are developed for or are used by non Computer Scientists. For example, some of the projects that my students in Senior Capstone have developed include:

- Exchange Traded Funds '03: combines finance and economics knowledge with Computer Science [5].
- Biogrammatics '06: combines Biology, Chemistry, and Computer Science [6].

- Automated Music Composition '07: combines Music Theory with Computer Science [7].

The challenge of teaching Computer Science at a small liberal arts college is that we three Computer Science professors need to understand and be able to teach all of the different areas of Computer Science. At a large research school, I would only have to focus on my areas of Computational Linguistics and perhaps Artificial Intelligence and Introduction to Programming; I would never teach Computer Architecture or Algorithms as these do not fall under my area of specialty. At Pacific, I've had to study these areas in order to teach the courses. Next semester, for example, I am teaching a course on Human Computer Interaction, which is a new topic for me.

The other general challenge with teaching Computer Science is that it is fast growing area, and I must stay current in new technologies in order to give the students a high quality education. In my 2005-2006 self-evaluation (page 5) I describe how I taught a Windows Programming course for the first time, which was a new topic for me.

Teaching Style

I would like to explain to you the kind of teacher that I am, and how I run my classes. I am not a lecturer, and do not like to talk at the students; however, the nature of Computer Science does require me to take some amount of time to explain new material. I will generally teach for 10 to 15 minutes, but then I will work through exercises with the students before going back to teaching again. Sometimes we will work through the exercises together on the board, or I will have them spend 3-6 minutes working through the exercises either individually or in groups, while I walk around the room observing the students and helping them when appropriate.

The most important thing for me is that the students are actively engaged during class time, even when I am explaining concepts. There is always a constant stream of dialog between the students and myself, and I will call on students to participate. I try to foster a safe and comfortable environment for all of the students to participate, even if they get things wrong; this seems to work, as I normally will have the full class join in. I tell the students that they are learning more by making mistakes, and so some of my most active students in class are the ones that make mistakes. To illustrate my teaching technique, I've included lecture notes from one of my CS250 (Introduction to Computer Science II) classes in my file along with how I use these notes during a typical class period [1].

I have learnt a tremendous amount in the last five years, and I know I am a much more thoughtful teacher than I was before. With the support of my colleagues in Computer Science I have improved on some of the mechanics of teaching like writing exams, determining suitable assignments, grading, and writing syllabi. I have included in my file examples of how my exams from CS150 (Introduction to Computer Science I) have changed and improved from 2002 to 2007 [2], and a discussion of how I have developed CS490 (Senior Capstone I) over the last five years [3].

Student Observations

One of the main joys of working at a small liberal arts college for me are the friendships that I make along they way. I am in touch with most of my graduates, and many of them are now close friends of mine. I take a lot of joy in their successes in both their careers and their lives. I have been invited and attended many of my former students' weddings and baby showers, and have visited them in their places of employments. They also return to Pacific to recruit current students for employment; alumni events where they

give advice to students on interviewing skills; graduate school survival; and insights into the latest technologies.

I feel that I have a good rapport with all of my students, regardless of whether they are CS majors or not. My student evaluations are evidence of this. I consistently get high student evaluations in all of my courses, and students describe all of the new skills that they learn in my classes. With regards to me as an instructor, students time and again say that I am nice, helpful, fair, patient, available, and explain things well for different learning styles.

That is not to say that I haven't had some bumps along the way, such as trying various teaching styles, some that worked better than others; fairness in grading; unclear assignments; long exams; and some respect issues. However, this is all part of the learning process of becoming a good teacher, and I believe that I have addressed these issues to the committee's satisfaction in my annual self-evaluations, and I will continue to address them and other issues that might come up as I strive to continue to be an excellent teacher.

Finally, the biggest indication of my success as a teacher is the success of the CS graduates. Most of the CS graduates, regardless of their academic abilities, have gone on to be successful in their careers. Students that I've taught for three or more years have gone on to renowned Computer Science graduate schools (University of Indiana, University of Washington, Oregon State University), have started their own successful companies (Merrill Interactive), and have found employment at good companies (Intel, WebMD).

What Next?

The Computer Science program at Pacific University is a great environment to work in, and I feel lucky to have the colleagues that I do. We get on famously, and I believe that this is reflected in the quality of the program that we offer. We have established a system of class visitations to consider different pedagogies, and learn from our very different teaching styles. We are also creating a repository of teaching materials, standardising syllabi, and clarifying the learning expectations for each of our courses. The whole process is invigorating and enlightening, and keeps the courses fresh and exciting for our students.

As a program, we continue to evolve our curriculum, which is necessary for a young field such as Computer Science. Last week we had a meeting where we discussed the programming languages that we teach, and whether it is time for a change. If we move to a four-credit system we will need to modify our service oriented classes since they will probably be 2-credit courses and we will not be able to cover all of the material that we currently cover.

I plan to continue to evolve and grow as a teacher. The following are some of my plans for the next few years:

- I am teaching CS315 (Human Computer Interaction) next semester, which is a course that I have never taught before. I discuss how I plan to teach this course and the new teaching strategies that I will use in my file [4].
- Professor Doug Ryan, Dr. Chadd Williams and I have all taught CS150 (Introduction to Computer Science I) and have observed that some students

compile[1] their programs dozens of time during a lab period, while others will write the complete program before compiling. We were wondering if this made a difference in how quickly the students became good/better programmers, so we are planning to run an experiment next year, where we will limit the number of compiles that a student can make in a given lab period. If we get interesting results we plan to write and submit a paper to the Northwest Consortium for Computing in Small Colleges (CCSC)[2], which is an annual conference that focuses on teaching Computer Science at small colleges.

- I plan on attending the national Association of Computing Machinery (ACM) conference on Computer Science Education, which will be held in Portland in March 2008[3].
- Finally, I am working with the other faculty in the program to submit a bid to host the 12th annual CCSC Northwest conference at Pacific University in 2010.

Professional Development/Scholarship

To help the committee evaluate scholarship in Computer Science, I am providing you with a document published by the Computing Research Association (CRA) describing scholarship in Computer Science and how it is evaluated[4]. This document is attached to the end of my narrative.

My field of research is Computational Linguistics or Natural Language Processing, which is an interdisciplinary field that draws from both Computer Science and Linguistics. The focus of my research is on developing linguistic tools for the Arabic language that can then be used in Automatic Machine Translation, Information Retrieval, and Linguistic Analysis.

In my professional career so far, I have produced two tools: (1) An Arabic Stemmer, and (2) An Arabic Part-of-Speech Tagger.

1. Stemming is the process or removing prefixes, suffixes, and in the case of Arabic, infixes from words to produce their roots. Arabic is a Semitic language, and by knowing the root and pattern of the word, both of which can be derived from the stemming process, information on the word's grammatical part-of-speech and meaning can be obtained.

2. Part-of-speech tagging is the process of defining a word's grammatical part-of-speech as it occurs in running text. Tagging is the first step in many linguistic and Computer Science applications as described above.

The Arabic Stemmer

Since working at Pacific University, I have re-implemented the stemmer that I developed for my M.Sc., and provided it with a functioning user interface. I have

1 Use the computer to test their programs for errors.

2 http://www.ccsc.org/northwest/

3 http://www.cs.duke.edu/sigcse08/

4 Best Practices Memo: Evaluating Computer Scientists and Engineers for Promotion and Tenure, Computing Research Association, 1999, http://homepages.dcc.ufmg.br/~bigonha/Bigonha/ullman.pdf.

packaged the stemmer and released it under the GNU General Public Licence[5], which allows anyone to use the stemmer for academic, non-profit purposes [8].

Since providing my stemmer for academic research, I have received 179 emails from researchers requesting the stemmer, dating from 10/20/2002 to 11/2/2007. The stemmer has also been integrated into various academic projects and has been analysed against other Arabic stemmers. These results, by other researchers, have been published in around 20 peer-reviewed publications, which I have included in my file [9].

In the summer of 2003 I licensed my stemmer to MitoSystems, Inc. for $10,000 [10]. This included my spending a week at the company's headquarters in Santa Monica, California, and integrating my stemmer into their decision support system for international government and multinational corporate clients. MitoSystems mention my stemmer on their website [11].

To highlight the importance of my work, I'd like you to consider the fact that Arabic Computational Linguistics is still far behind the work on English Computational Linguistics; ten years ago it was practically non-existent. It has always been a goal of mine to share all of my work with the community to encourage more research in the field. I would not be so egotistical to say that my work spurred on research in Arabic Computational Linguistics (I'm sure that the events of 9/11/2001 had a lot to do with it); however, I do know that my work on Arabic stemming and tagging has been used in a variety of projects.

I am immensely proud at how much attention my stemmer has received, and the number of publications that cite my work. In one publication the author states, "The two most successful approaches to Arabic stemming have been a root-extraction stemmer developed by Khoja…."[6], while in another the authors state: "We adopted Khoja's algorithm for root-based retrieval. The reason behind this adoption is that Khoja's algorithm showed superiority over previous works in root detection algorithms". They go on to say "The root algorithm based on the work of Khoja, which is considered as an aggressive stemmer, has shown performance superiority over surface-based (no stemming) approach"[7]. I have been providing the stemmer to the community all of the time that I've been at Pacific, and publications as recent as 2007 cite the stemmer.

Under the Boyer model, these activities can be categorised as "Discovery of Knowledge" since the stemming algorithm is a unique and new algorithm for the stemming of Arabic text. The stemmer itself could be categorised as "Application of Knowledge" as the algorithm has been integrated in a stand-alone application for distribution, and has been integrated into a much larger application at MitoSystems.

The Arabic Part-of-Speech Tagger

The Arabic Part-of-Speech tagger represents the bulk of the work for my Ph.D. dissertation. The tagger uses both language rules and statistics to automatically assign parts-of-speech to Arabic words in running text.

---

[5] http://www.gnu.org/copyleft/gpl.html

[6] Arabic Stemming Without A Root Dictionary Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) Volume I Pages: 152 - 157 Year of Publication: 2005 ISBN:0-7695-2315-3

[7] Proceedings of the eleventh international conference on Information and knowledge management McLean, Virginia, USA Pages: 340 - 347 Year of Publication: 2002 ISBN:1-58113-492-4

I completed the work on my dissertation during the summer after my first year at Pacific University, and defended my Ph.D. that September. My Ph.D. work has resulted in two international, peer-reviewed conference publications (papers and presentations), and one of these papers was revised and published as a book chapter [12, 13].

These papers have been cited in 8 or so peer-reviewed publications by other researchers, and have highlighted the intricacies of the Arabic language that add to the difficulty of developing Arabic linguistic software. I have included in my file the research papers that cite my work on Arabic tagging [15].

In 2004, I received a faculty development grant to rewrite the part-of-speech tagger for distribution to other researchers, similar to what I did for the stemmer [14]. This grant enabled me to attend the International Association for Computational Linguistics (ACL) conference in the summer of 2005, and visit researchers at the University of Leeds and Lancaster University. The tagger is in the process of being rewritten and is not yet in a form that is ready for distribution; however, there is a lot of interest in the tagger and I have received numerous emails from researching requesting the tagger.

Under the Boyer model, my research on tagging could be considered "Discovery of New Knowledge" since there is no Arabic part-of-speech tagger available, "Integration of Knowledge" as it integrates the linguistic properties of Arabic with Computer Science techniques such as Hidden Markov Models, and "Application of Knowledge" since it is being developed into an application.

Other Scholarship Activities

My scholarship activities have recently moved beyond stemming and tagging to research on Internet based forms of communication.

In the fall of 2006 I applied and received a $3,500 Berglund Fellowship Grant to research web logs in the Middle East [16]. I have used this grant for course release in the fall of 2007 and I am currently compiling a corpus of Arabic texts for my analysis. The results of this research will be written up and submitted to the Berglund Center journal "Interface on the Internet"[8] for publication. I will also submit a paper to the International World Wide Web Conference (WWW2009 – Madrid, Spain)[9].

My other activities include:

- Receiving a $2,000 research grant in 2003 from the Computing Research Association's Committee on the Status of Women in Computing Research (CRAW) to fund two undergraduate students to work with me on a neural network part-of-speech tagger [18].

- Translating children's Math software (Kali) to Arabic for the Mathematician and MacArthur fellow Dr. Jeffrey Weeks. I met Dr. Weeks when he came to Pacific University to present at the Whiteley Distinguished Lectureship program. The software is available for download [17].

- Presenting my research at the University of Leeds in 2004 [19].

[8] http://bcis.pacificu.edu/journal/

[9] http://www.iw3c2.org/conferences/future/

- Presenting at the Pacific University Faculty Forum in 2005 [20].

What Next?

Here are some of my plans for the future:

- Jim Thomas, director of Homeland Security's National Visualization and Analytics Center (NVAC)[10] at the Pacific Northwest National Laboratory (PNNL) in Richland, WA, has invited me to apply for a fellowship at the center. The fellowship would support me for a year during my sabbatical, should I successfully get tenure and the sabbatical. This would be an amazing opportunity for me as I would work with top researchers in the field of information technology.

- I have also been invited to spend a year at MitoSystems to work with them on their information analysis system. Although this does sound appealing, I am concerned about the cost of living in Santa Monica, CA.

- I am submitting a paper on "Blogging in the Middle East" to the Twenty-Second Annual Symposium on Arabic Linguistics to be held at the University of Maryland, College Park in March 2008.

- I would like to provide my Arabic Tagger to the community. I am considering applying for a summer research grant to work with one or two undergraduate students to develop a user-friendly front-end to the tagger. The tagger would then be made available to the community.

- Finally, I am working with other researchers on writing a proposal for a workshop on Semitic Languages for the 22nd International Conference on Computational Linguistics to be held in August 2008.

Summary

I believe that I meet and exceed the expectations for tenure and promotion in the area of Professional Development. I believe that the quality of my scholarship speaks for itself, and since starting at Pacific University I have received three grants, given two presentations, published a book chapter, and licensed my software. I have also moved beyond my doctoral work and am now pursuing new venues for research.

Service

Service is an integral part to being a university professor, and because of that I will do whatever is necessary in service to my department, university, and community. I have been involved in many service activities over the last five years and I would like to highlight to you that I received the highest rating from FDPC for my service in 2004-2005, which I understand is unusual for a probationary faculty member.

Here are a few examples of some of the activities that I've been involved in.

Within the department, I participate in all discussions on the curriculum; I have been the faculty advisor to the CS club for the last four years, which has now become a very

[10] http://nvac.pnl.gov/index.stm

active club [21]; I wrote up half of the CS program's assessment report; and participated in four search committees for both Math and Computer Science.
Within the division, I have been on the Division of Natural Sciences strategic planning committee since February 2006, I am on the Natural Sciences facilities planning committee, and I am the director-in-training to Dr. Juliet Brosing for a summer science camp for girls in the summer of 2008.

Within the university, I worked with the English Language Institute (ELI) and International Programs to help the 35 new Saudi students adjust to life on campus and in the United States; gave two guest talks to Dr. Erica Kleinknecht's Cognitive Science classes in 2004 and 2006 [22], two of Dr. David Boersma's Linguistic classes in 2004 and 2006 [23], and to Dr. Kazuko Ikeda's Intercultural Communication class in 2006 [24]; was the advisor to the undergraduate student's annual yearbook for three years [25]; and I served on the Pacific Undergraduate Community Council (PUCC) for two years.

For the community, I have given a lecture on the status of Women in Computer Science to a meeting of the Association of University Women (AAUW), I am a member of an AAUW committee organizing a STEM (Science, Technology, Engineering, and Math) award event for local junior high school girls, and I represented Pacific University at the 2003 convocation and annual meeting of the Council of Higher Education for the United Church of Christ in New Orleans.

Furthermore, I have just been elected as a faculty senate member for 2.5 years.

Final Words

My five years at Pacific University have been wonderful. I look forward to coming into work every day and enjoy working with my colleagues here. I hope that I have demonstrated to you my dedication to Pacific University, and I hope that I will continue to serve Pacific University for many years to come.

Sincerely,

Shereen Khoja

References:

1. Lecture notes from CS250 (Introduction to Computer Science II), along with a description of how I would use these notes in a typical lecture.
2. Sample exams for CS150 (Introduction to Computer Science I) from 2002 and 2007 along with a description of how I have changed the way that I write exam.
3. Sample syllabi, rubrics, and assignment descriptions for CS490 (Senior Capstone I), along with a discussion of how I have developed these over the years.
4. Discussion of my plans for CS315 (Human Computer Interaction).
5. "Exchange Traded Funds" Senior Project by Nicole Pearson and Stephen Farnsworth '03.
6. "Biogrammatics" Senior Project by Adam Herr '06.
7. "Automated Music Composition" Senior Project by Dan Uhl '07.
8. Screenshots of my Arabic Stemmer, which uses the UI I developed while at Pacific University.
9. Publications that reference my Arabic Stemmer. I have highlighted the sections n these publications that mention my stemmer.
10. Licence agreement with MitoSystems Inc.
11. Screenshots of MitoSystems Inc. website. http://mitosystems.com/mito-multilingual.htm.
12. "APT: Arabic Part-of-speech Tagger" by Shereen Khoja. Proceedings of the Student Workshop at the Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL2001), Carnegie Mellon University, Pittsburgh, Pennsylvania. June 2001.
13. "An Arabic Tagset for the Morphosyntactic Tagging of Arabic" by Shereen Khoja, Roger Garside and Gerry Knowles. Chapter in "A Rainbow of Corpora: Corpus Linguistics and the Languages of the World", edited by Andrew Wilson, Paul Rayson, and Tony McEnery; Lincom-Europa, Munich. January 2003.
14. "An online part-of-speech tagger for Arabic". 2005-2006 Faculty Development Grant, Pacific University.
15. Publications that cite my research on Arabic Part-of-Speech Tagging. I have highlighted the places in the publications where my tagger is mentioned.
16. 2007/2008 Summer Institute Fellow for the Berglund Center for Internet Studies for a research project entitled "Blogging in the Middle East".
17. Kali Geometry Game. http://www.geometrygames.org/Kali/.
18. Application for CREW grant, and outcomes.
19. Slides from presentation at the University of Leeds in 2004.
20. Slides from presentation at the Pacific University Faculty Forum in 2005.
21. List of CS club activities.
22. Notes from a lecture I gave to the Cognitive Science course in 2006.
23. Notes from a lecture I presented at the Linguistics course in 2006.

24. Notes from a lecture I presented to the Intercultural Communication class in 2006.
25. CDs of the College of Arts & Sciences annual yearbook for the years 2003, 2004, and 2005.