

Submission Packet:

Eric Ciminelli / SEITZER

2014 Student-Faculty Collaborative Scholarship Program

Name of Student Proposer:

ERIC CIMINELLI, Computer Science A & S, Class of 2015

Name of Faculty Advisor:

DR. JENNIFER SEITZER, Visiting Associate Professor (2013-2014),
Associate Professor (2014-2015); Computer Science

Submitted by:

DR. JENNIFER SEITZER, Visiting Associate Professor (2013-2014),
Associate Professor (2014-2015); Computer Science

Date Submitted:

Thursday, February 27, 2014

Checklist of Items Enclosed in this Packet (in order of Appearance):

1. Faculty-signed cover letter
2. Student written proposal
 - a. Non-technical proposal summary
 - b. Comprehensive literature review
 - c. Description of the proposed work and the expected outcomes
 - d. Project timeline
 - e. Budget
 - f. Reference page
3. 1-2 paragraph personal statement
4. Student-Faculty contract



COMPUTER SCIENCE – ROLLINS COLLEGE – 1000 HOLT AVE., WINTER PARK, FL 32789 – T. 407-646-2432 – F. 407-646-2479

February 26, 2014

Professional Standards Committee
Rollins College
1000 Holt Ave.
Winter Park, FL 32789

Dear Colleagues:

I am pleased for the opportunity to write this letter of recommendation for Eric Ciminelli who is applying for a Student-Faculty Collaborative Scholarship for the period of June 23, 2014 -- August 15, 2014 to work with me in an artificial intelligence (AI) project entitled *Relevant Data Condenser (RDC) Smartphone App*. In this project, we will develop a software system that will use techniques of machine learning, heuristic search, and distributed AI (which encompasses web mining and computer networking) to be used on a smartphone. The user will have the ability to receive intermittent notification messages whenever the user's topic appears on the World Wide Web in the form of an announcement or other news form.

Mr. Ciminelli was a fine student in one of my classes in Fall 2013 (in which he earned a high A) and is currently enrolled in two of my courses. He is an incredibly talented programmer with an unusual amount of programming experience including the study of Visual Basic in eighth grade, learning and mastering AP Computer Science in high school, as well as starting and running his own computer repair business since the age of sixteen. He has been developing apps on his own for five years and has expressed a desire to incorporate some high level artificial intelligence theory into his work thereby raising the overall level of depth and complexity of his programs. I welcome such an enthusiastic student into my work in artificial intelligence research.

This year (2013-2014), I feel fortunate to be serving as a Visiting Associate Professor at Rollins. In November 2013, I was offered a tenure-track associate professorship for the 2014-2015 academic year – which I happily accepted. I am thrilled to be re-starting my research program. I started and directed an undergraduate research program at The University of Dayton for twelve years where I guided students in many different artificial intelligence projects ranging from genetic algorithm applications to stratified multi-agent systems. I am looking forward to the new domain of smartphone applications using the familiar turf (for me) of machine learning and intelligent system implementation.

One of the most interesting aspects of this project is that it will incorporate many sub-disciplines of computer science – each with its own language and set of algorithms. In the 2010s, this *is* the complexion of computer software. Software can no longer be conveniently categorized as “a simple computer program written in one language, running on one computer.” Today's systems are amalgamations of many platforms with many operating systems communicating over the Internet (which is made of many networks). In an academic environment, it is unusual for a student to get the opportunity to develop this type of ubiquitous hybrid. Such a project offers opportunities of design, intercommunication, complex testing, and heterogeneous trouble-shooting that are not present in the development of the typical, reduced-scale, monolithic project assigned in a 16 week course. I believe Mr. Ciminelli's eclectic programming background will allow him to fare very well in the complex nature of the many constituents of the proposed system.

The informational arenas of *Relevant Data Condenser (RDC) Smartphone App* include mastering the high level concepts, vocabulary, programming languages, and algorithms of 1) the AI inference engine (which includes a knowledge base system and language generator), 2) the web miner which includes a web crawler, page ranker, and inverted index, and 3) the smartphone interface and communication server. The implementation languages to be employed include Objective-C, Java, and UNIX shell scripting. The repertoire of algorithms include page-ranking, inverted index, and web-crawler interaction for the web-mining component, underlying hypergraph storage and manipulation for the intelligent knowledge base implementation, and simple graphics and client-server interaction for the smartphone end-user system and graphical user interface.

At a recent All-Faculty Meeting, it was stated that one of our goals is to provide students with exceptional experiences that they would normally not have in any other place but Rollins. At bigger schools, a student *might* get a small part of a *big* project. In this project, I am able to give Mr. Ciminelli a *big* part in (what could grow into) a big project. At Rollins, we are fortunate to have students that study both Shakespeare and smartphones. We have an enormous wealth of multidisciplinary students with a cornucopia of gifts that can be combined and integrated into the artifacts of the day (i.e., our digital footprints and devices). Eric Ciminelli is a socially adroit Rollins student who is well-liked by his colleagues and who recently started the Rollins College Ping Pong team. He is also a student, who, last semester, I witnessed to be a whiz at machine and assembly language programming -- that he just learned -- without working very hard.

I would like to see Eric academically stretch. I would like to see him have to work hard and struggle a bit -- grappling in a research environment where we attempt to answer questions that are hard to even codify. I encourage you to support him in this endeavor.

Very truly yours,

A handwritten signature in black ink, appearing to read "Jennifer Seitzer". The signature is fluid and cursive, with a long horizontal stroke extending to the right.

Dr. Jennifer Seitzer, Visiting Associate Professor (2013-2014)
continuing on as Associate Professor (2014- 2015)
Computer Science Department
Rollins College
407-646-2303

Relevant Data Condenser (RDC) Smartphone App

Eric Ciminelli, A&S Class of 2015

Advisor: Dr. Jennifer Seitzer
Department of Computer Science
Rollins College

*Submitted February 27, 2014 to the
Rollins College Student-Faculty Collaborative Scholarship Program*

I. Non-Technical Proposal Summary

One of the newest genres of software systems is the app for the smartphone. This research will integrate techniques of artificial intelligence into a mobile system that is accessible to anybody who has a smartphone. Artificial intelligence (AI) is the area of computer science that attempts to put human thinking into computer programs. Our application will incorporate AI techniques to eliminate the need for users to constantly “check the net” for new and relevant information on any topic they choose. It will generate “*Alerts*” containing useful information that will come in a succinct text message-like fashion (*phone notification*).

AI techniques employed will include smart web crawling (scanning the Web for valuable patterns) and automatic language generation related to natural language understanding (such as what Siri does). This summer, the app is going to be exemplified using the topic of cryptocurrency (digital money) trading even though any topic could be used.

The main producible from this summer research period (June 23rd through August 15th, 2014) will be a computer system comprised of the following components: 1) a web crawler, 2) an AI inference engine that performs web page ranking, parsing, and language generation, 3) a server that pushes relevant information to 4) a smartphone client app that displays the information.

In summary, the PI is requesting \$4800 (student and faculty stipends or \$6500 including conference funding) to create a mobile app that will contribute on many levels. Personally, this experience will enrich the PI by immersing him in an undergraduate research experience while on his path to graduate school. For the Rollins community, future students can use the software infrastructure to create their own software system requiring real time web *Alerts*. Globally, the system could be applied to applications as important as emergency notification systems.

II. Introduction

Computer science is the discipline that guides the development of information processing and exchange. There are a vast amount of sub-disciplines in computer science including graphics, robotics, machine learning, visualization, computer architecture, and many more. Our proposed summer work is to develop a system drawing from four main sub-disciplines: artificial intelligence, web mining (automatically searching the Web for useful information), language theory (understanding and translating information found), and computer networking (the interconnection of distributed computer systems).

This proposal is requesting funding to build a computer system called RDC (relevant data condenser) that could be used by anyone with a smartphone. The system will be made up of primarily an artificial intelligence (AI) inference engine consisting of a web crawler, parser, and translator, and server that communicates with a smartphone. The purpose of this system will be to automatically incorporate these parts in order to inform users of any fresh news pertaining to their pre-specified topic.

The software will be written in Objective-C, a powerful computer programming language that the student researcher is beginning to learn as part of this project. The PI has experience in the world of app development, with his most popular release, UV Index, written as a high school student having nearly 50,000 downloads. The proposed project is a significantly more difficult app that will enable him to further his knowledge in app development, AI, web mining, and computer networking.

III. Comprehensive Literature Review

The seminal research paper in the sub-discipline of web mining is Brin and Page's presentation of the Google search engine [Brin 1998]. The most useful part of this paper that can be applied to our crawling system is their PageRank algorithm. It works by measuring the amount of relevant links and keywords in a page. In doing so, the importance to the search can be calculated using this simple iterative algorithm. We plan to incorporate this type of thinking to determine which pages to visit next from a set of hyperlinks found on a page. We propose that for a page searched, there will be a maximum number of links that can be visited within that page. The system will determine which of those will be visited based on the amount of relevant information within the page and the publication date (older information is not particularly useful to this app). Only the pages with the most information pertaining to the subject will be used.

A direct quote that beautifully explains the main technique of their search engine:

PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. And, the damping factor is the probability at each page the "random surfer" will get bored and request another random page.¹

Denis Shestakov introduced some problems that web crawlers face when searching the Internet. He explains that, "In order not to be banned by a web server, a crawler has to avoid sending too many URL requests to a server within a short time period. Since the distribution of pages over web servers is non-uniform, a crawler faces a problem of downloading a large number of pages from only a relatively small number of web servers." [Shestakov 2013] In our work, we will take into account the number of requests allowed by servers in a certain amount of

¹ Brin, Sergey, and Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks & ISDN Systems* 1998 (107-117).

time. For each site crawled, a limit will have to be determined in order not to overflow a website's server. We also propose using slightly varied random request times within said limit to increase security. In addition, keeping track of previously cached pages and knowing not to visit them again will ensure that the crawler does not overflow a site.

Other sources include work by Dmitriy Meyerzon and Sankrant Sanu where they introduce the idea of address mapping within web browsing [Meyerzon 2000]. Our system will build on their work incorporating primary and secondary address mapping to aid in determining the importance of links and pages.

An important feature of our system will be parsing and translating text into concise messages for the user. This process comes from language theory. Specific techniques include finite state machines, parsing, and code generation [Hopcroft 1979]. In Russell and Norvig's Artificial Intelligence book, the PI has studied techniques of search, heuristics, and intelligent mining to use in the web crawler and AI inference engine (see Figure 1) [Russell, Norvig 2010]. Computer networking connects a system by allowing parts to communicate with each other. Kurose and Ross demonstrate how to do this through their presentation of the client-server programming paradigm and communication on the TCP/IP stack [Kurose 2013]. This is crucial to our application as our system is going to function through the Internet.

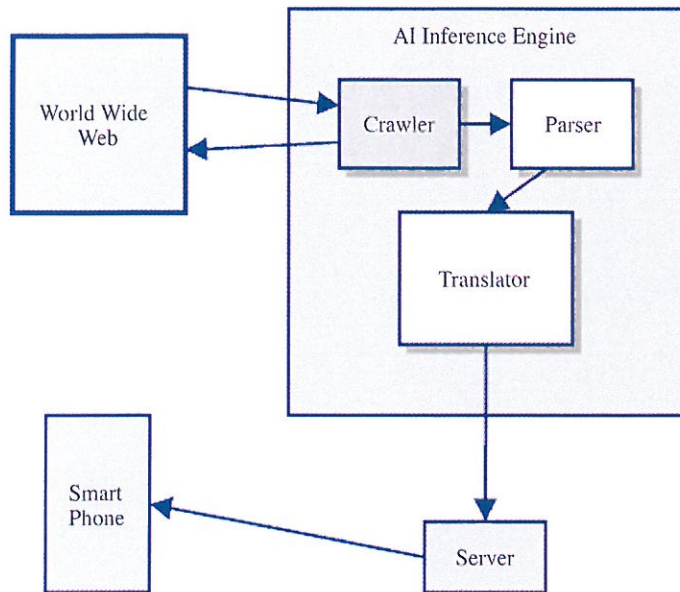
IV. Description of the Proposed Work & Expected Outcomes

In any research endeavor, a question should be “What is new in this research?” For this work, some of the individual ideas are not necessarily new. However, their combination and implementation is. The work being proposed for the funded period June 23rd through August 15th, 2014, is to generate an intelligent app for the smartphone. This app will automatically provide up-to-date information on a topic chosen by the user. It will relieve the user of the necessity to constantly check the Web for relevant information on their subject of interest. It will be able to be used as a notification system that informs the user of important and fresh information. The research team will develop the entire system from scratch from the crawling algorithms to the interface which the user interacts with.

A. System Specifications

The application is going to utilize an internal engine that constantly scans through a repository of pre-specified web sites for keywords. Once those keywords are found, the page that contains the keyword is stored locally. Then, using a number of AI methods including a context-based ranking algorithm, the usefulness of the keyword is determined. Then, a brief message summarizing the information is generated and sent to the smartphone user. This is called “pushing” information to the user.

In Figure 1 below, the overall system framework of RDC is presented. There are five basic steps that stay in a repetitive loop.



1. Every 5 minutes crawl the repository (frontier of URL's) on the web
2. Collect pertinent information using our context-based ranking algorithm
3. Parse the pages and extract useful information
4. Translate possibly verbose info into succinct messages
5. Server delivers messages to subscribers on their smartphones

Fig.1 Left: A diagram of the RDC system

B. Real World Example: Cryptocurrency (aka Digital Money)

Traders are taking advantage of how volatile the cryptocurrency market is right now. Cryptocurrencies are just beginning to be used in the world. There are even ATM's popping up where you can withdraw your cryptocurrency for actual cash [Jervis 2014]. But until they become more widely accepted and recognized, their exchange rates will continue to fluctuate greatly on a daily basis.

Our system will be helpful because of its ability to forecast these fluctuations by being the first to know about any news that goes on with cryptocurrency. One must realize that the

exchange rates for cryptocurrencies normally change after news gets out of a certain event that can affect the price. For instance, if a popular business begins to accept a cryptocurrency, as soon as news spreads, the price is going to go up. Those who are the first to know the news are the ones who are most likely to succeed in the trading world.

An important part of this work is that the topic of interest can be changed. We choose to exemplify our software system using the topic of cryptocurrency. However, as in many artificial intelligence applications, the underlying software infrastructure can be extracted and applied to virtually any other subject. For example, intelligent expert systems originally written for medical diagnosis were decomposed and then applied to computer system configuration [Russell 2010]. Even though the initial topic will be cryptocurrency trading, the purpose of this summer research is to exemplify system RDC and the framework that it is built from. Some other ideas for applications beyond trading include a disaster alert system, “Rollins specific” data such as ‘deadline to drop classes’ messages, and any other uses where new information about a certain subject is needed.

C. The User Experience

The users of the application will interact with the web-crawler via their smartphones. Inside the application, the user will be able to pick the type of cryptocurrency they want to monitor from a list. Each currency selected will have its own tab where the current price along with a graph displaying the price over time can be seen. At the top there will be a news feed where any relevant news that the web-crawler finds will be displayed. If the user clicks on the news *scroller* it will bring them to a page where a list of relevant articles found will be displayed.

Selecting an item will open a browser and bring the user to the source page where the data was found.

The most powerful feature of the app will be the notifications it sends to its users. Depending on the type of cryptocurrency the user selects to monitor, alerts discovered by the crawler will appear on the home screen in a form such as, “*PC World says that Litecoin is expanding. Now is a good time to buy.*” or “*LiteCoin hit a low for the week at \$742.*”

The app will also include notifications for significant price drops or gains. The user can determine at what price values they want to be alerted and the app will do so accordingly.



Fig. 2 Right: A screen mockup of the ticker page in the application. The top shows the news scroller followed by the different types of currencies. Under that is the current price in green along with other useful information including a price graph.

D. Work Finished to Date

We have already begun work on the RDC design and composition of algorithms. The PI has also begun learning Objective-C and has successfully coded some small apps for learning purposes. The PI has also created a number of screen renderings including the one in Figure 2 to get a better idea of the user interface of the app. In Figure 3, below, the newly developed pseudo-code for the ranking algorithm is shown:

- Every k time units, do the following:
- Crawler visits a webpage, caches the page and gathers a list of URL's
 - If the date for each URL is younger than a certain set amount of time
 - Cache that page locally
 - Search for relevant keywords
 - Store relevant information in knowledge base
 - Rank the pages: only the top number n of pages containing valid information will be saved
 - Delete dead-ends, pages with no relevant information, from the cache

Fig. 3 The ranking algorithm

V. Project Timeline

Work to Date

- Developed the modified PageRank algorithm
- Created mock-ups of the user interface for the smartphone

March – April 2014

- Continue to learn Objective-C
- Continue to design and improve the crawling algorithms and AI engine

Summer 2014

Week 1 – 3

- Implement the AI used in the system in Objective-C
- Code the web crawler

Week 4 – 5

- Design and code the smartphone app used to control the system

Week 6 – 7

- Program the server that communicates between the system and smartphone app

Week 7 – 8

- This time will be used for beta testing

Fall 2014

- Prepare a presentation of the work accomplished
- Write a conference paper for submission
- Write and submit a journal article

VI. Budget

Ciminelli (student) stipend.....	\$3000
Seitzer (faculty) stipend.....	\$1800
Conference travel expected costs	
Student registration.....	\$200
Student travel and meals.....	\$500
Faculty registration.....	\$500
Faculty travel and meals.....	\$500
*Total without conference funding.....	\$4800
Total with conference funding.....	\$6500

*The faculty advisor was notified on 2/24/2014 that she has been awarded a Rollins Individual Development Grant to cover the conference expenses which can be used if conference expenses are not able to be funded by the SFCS program.

VII. Conclusion

We have proposed to create a computer app for a smartphone during the period from June 23rd to August 15th that will provide *Alerts* to users on their chosen topic. This summer, we will build the system and then exemplify it using the topic of cryptocurrency. This topic, however, is fully replaceable with any topic of interest. The benefit here is that the user doesn't have to actively search. Our system, RDC (Relevant Data Condenser), will automatically inform the subscriber of any news about their topic.

VIII. References

- [Brin 1998]. Brin, Sergey, and Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *The Anatomy of a Search Engine*. Computer Networks & ISDN Systems, 30 (1998, pages 107-117). Web. 13 Feb. 2014.
- [Hopcroft 1979]. Hopcroft, John E. *Introduction to Automata Theory, Languages, and Computation*. New York: Addison-Wesley, 1979. Print.
- [Jervis 2014]. Jervis, Rick. "Bitcoin ATMs Come to USA." *USA Today*. Gannett, 20 Feb. 2014. Web. 22 Feb. 2014.
- [Kurose 2013]. Kurose, James F., and Keith W. Ross. *Computer Networking: A Top-Down Approach*. Boston: Pearson, 2013. Print.
- [Meyerzon 2000]. Meyerzon, Dmitriy, and Sankrant Sanu. "Method of Web Crawling Utilizing Address Mapping." *Google Patents*. Microsoft Corporation, 7 Nov. 2000. Web. 12 Feb. 2014.
- [Russell 2010]. Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall, 2010. Print.
- [Shestakov 2013]. Shestakov, Denis. *Intelligent Web Crawling*. *IEEE Intelligent Informatics Bulletin*, Dec. 2013. Web. 12 Feb. 2014.

Personal Statement: Eric Ciminelli, Computer Science A&S, Class of 2015

Ever since my very first computer programming class in 8th grade, I have had great interest in programming. I learned Visual Basic in middle school and then went on to learn C++ and Java in AP Computer Science in high school (Charlotte Latin High School, Charlotte, NC) where I was also the president of the Technology Club. Along side of classes, I explored computer hardware by building my own computers. At the age of 16, I started my own computer repair business (“Computer Repair and Construction of Greater Charlotte”) where I would make house calls to do a wide range of possible things (optimization to virus removal).

I came to Rollins in 2011 as a Computer Science major. I have done well in the computer science curriculum and have been involved in other ways at Rollins including just starting the Ping-Pong club. I am excited to be doing undergraduate research in app development and artificial intelligence because this is something beyond the normal curriculum. I believe participating in the Rollins Student-Faculty Collaborative Scholarship Program will allow me to be immersed in a project that I am passionate about. I plan to go to graduate school and feel that doing undergraduate research will help me get admitted as well as figure out the specific area of the discipline that I want to go into.

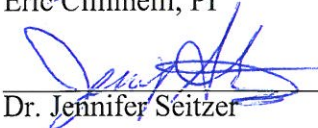
Student Contract - 2014

As a participant in the 2014 Student-Faculty Collaborative Scholarship Program I, Eric Ciminelli, will contribute intellectually and physically to my chosen scholarly work to the greatest extent possible. I agree to be present for work at the times and places specified by my faculty mentor, Dr. Jennifer Seitzer, and I understand that I will be paid \$3,000 as a stipend for working between June 23, 2014 and Aug 15, 2014. I understand that the purpose of the Student-Faculty Collaborative Scholarship Program is to allow me to participate in original scholarship, and that the success of the program depends upon my dedication to it. With this in mind, under the guidance of my faculty mentor I will accomplish the following goals prior to the final date listed above.

1. Daily meetings Monday through Friday unless otherwise agreed upon (e.g., for contingent situations such as out-of-town meetings, etc.)
2. Attendance at special talks and events that would help the Student researcher in his studies as directed by Dr. Seitzer
3. The completion of development of the parts of the RDC software system as assigned
4. The software development of the AI Inference Engine to be used for web mining and pattern extraction for the RDC system
5. The testing of the RDC system with several smartphone "subscribers"
6. The composition of the final report documenting outcomes and results (to be completed Fall 2014)

I understand that while participating in the program I will be subject to the student rules of conduct and those infractions of these rules will be dealt with by the existing student judicial process. I further understand that violation of the College's non-toleration policy for drug use and paraphernalia will result in immediate expulsion from the program.


Eric Ciminelli, PI


Dr. Jennifer Seitzer


Feb. 26, 2014


Feb 26, 2014