

Google Summer of Code 2011 Project Proposal

Open Bioinformatics Foundation

Mocapy++Biopython

Vita Smid
me@ze.phyr.us, <http://ze.phyr.us>

Motivation. I am thrilled by the opportunity to work on this interdisciplinary project, since I am sure that my skills *and* my interests are a perfect match for it.

First of all, I am studying mathematics at my school (Charles University in Prague, Faculty of Mathematics and Physics) with strong bent towards application. In particular, I have taken courses in graph theory, probability, statistics, modeling, etc. Applied mathematics is very attractive for me.

Second, my previous (secondary level) education and interests were purely in computer science. I have 9+ years of experience programming in various languages, most notably PHP, C++, and Python. Despite being a student of mathematics, I still frequently work as a programmer.

And to wrap it all up, many of my friends are students of (bio-)chemistry and through them I became interested in bioinformatics. Although I do not have much solid knowledge of the field yet, it is intriguing for me and I have access to excellent resources.

Programming experience and skills. As mentioned above, I have been programming for more than 9 years. Apart from my specialization in web development, I have written a great deal of software in C, C++, and Python and placed well in several nation-wide contests with it. This included physical simulations, distributed computations, and networking. I have done even more work in the commercial sector. My CV¹ lists all the details.

Projects and contribution. Several of my projects are available as open-source at <http://projects.vitasmid.com>. There have been many more, most of them web applications, whose source must remain closed by the request of my employers.

In 2007 I participated in Google Highly Open Participation contest as a high school student.

Project plan. Before the actual program starts in May 24, I would like to learn more about directional statistics and Bayesian network training. I have full access to my university's library and I can consult the topics with teachers from the Department of Probability and Mathematical Statistics and the Department of Applied Mathematics at my faculty.

I expect to spend May 24 – June 14 studying the papers based on Mocapy++'s models to get a good understanding of the toolkit's use cases. I can get a lot of help with molecular biology from my friends.

June 14 – June 21 would be spent on me getting familiar with Mocapy++ code. I have glanced over it quickly and I think the code is high-quality and easy to understand.

I think it is very important to plan the bindings interface carefully, and I would accordingly devote at least 10 days (June 21 – June 30) to design it and to discuss my ideas with developers of Mocapy++ and Biopython.

Implementing and testing the actual bindings should take about three weeks, or July 1 – July 21.

Since writing example applications will involve a lot of molecular biology study on my part, I think the remaining time till August 22 will be just enough to develop a few of them. I will certainly need advice and suggestions from the community of Biopython and Mocapy++ users. Building the example applications will also be a great opportunity to test the bindings in practice and hunt down some remaining bugs.

¹<http://ze.phyr.us/cv.pdf>

Obligations, vacations, and other plans for the summer. I do not have any specific plans for the summer. I might spend a few weekends hiking in the mountains or making a trip to Germany, but that would be about it. However, end of term exams at my university take place during June, so it can be expected my performance will be impacted during this period.