Name: Uvarov Michael

E-mail address: freeakk@gmail.com

IRC nickname (if any): freeakk

Any personal websites, blogs, identica, twitter, etc: [Github profile](#)

Biography (tell us a bit about yourself):

I am from Rostov-on-Don in Russia. I am 21 years old. I am studding at Southern Federal University from 2008. I use Open Source software every day.

At 2007, I started to help my school with an Internet gateway and I made a decision to use GNU/Linux for this task. Then I was in Traffpro command as a front-end developer.

My first passion in programming was PHP. It was simple and there were a lot of literature about it. I used PHP for four years with a set of other technologies, such as REST, XML and AJAX. I learned JavaScript and SQL. The university helped me to improve my skills.

Finally, Erlang became my main programming language. I found a noisy problem with advanced Unicode support and solved it. Of course my approach was not so efficient and did not support locales. That's why, I created a binding for ICU library as an alternative.

I decided to participate in GSoC 2012, because I like Open Source software and want to make it better. I want to gain invaluable experience of developing in international command of professionals.

Eligibility: yes

Background Information

Have you taken part in GSoC and/or GHOP and/or GCI before?

No, I have not.

Please tell us about any previous experience you have with Xapian, or other systems for indexed text search.

I used MyISAM full-text indexes. Also, I worked under Unicode Collation Algorithm in Erlang.

Do you have previous experience with Free Software and Open Source other than Xapian?

I am a participant of erlang-question mail list. I forked etorrent torrent client and wrote webui for it. I developed Erlang version of an Unicode collator and a library for handling Unicode strings using Unicode Character Database. I used many applications from Erlang/OTP:

mnesia, eunit, tv, appmon.

And applications from other developers:
- I created few small plugins for rebar.

- I am using lager and SASL for logging.
- I am using gproc as a dictionary of processes.
- I used mochiweb and cowboy as web servers.
- I am an user of few JavaScript libraries: qooxdoo, ExtJS, jQuery and a former PHP enthusiast.

What other relevant prior experience do you have (courses taken at college, hobbies, holiday jobs, etc)?
- Courses of C/C++.
- Coding in Erlang as a hobby.

What development platforms, tools and methods do you prefer to use?

Platform: Debian Linux, Erlang/OTP

Tools: rebar, make, git, vim

Methods: TDD

Have you previously been responsible (as an employee/volunteer/student/etc) for a project of a similar scope? If so, tell us about it.

I implemented ICU binding as NIFs. I used Erlang and C/C++. I learned a lot of things about interoperability of Erlang applications.

What timezone will you be in during the coding period?

GMT+4 / MSK

Will your Summer of Code project be the main focus of your time during the program?

Yes, it will.

How many hours a week will you realistically be able to devote to your project?

20 hours

Are you applying for other projects in GSoC 2012?

No, I am not.

Project name: Erlang binding

Motivations

Why have you chosen this particular project?
- This project is one of the projects that uses Erlang/OTP. It is my passion.
- This project has strong background in information retrieval field.
- I am interested in developing back-end applications.
- I think, it is a good challenge for a developer to participate in this kind of events.
- In my opinion, this work will be useful for community.

- It is fun.

Who will benefit from your project and in what ways?

My project will serve both for the community of Erlang language and for Xapian.

Project Details

Describe any existing work and concepts on which your project is based.

This project will contain both Erlang and C++ code. My main goals are: * to minimize of the C++ part of the driver; * to make calls to a port thread-save.

My project will use an Erlang driver interface. Each driver port will handle queries synchronously and will have their own set of objects.

Ports can be collected into a pool using an Erlang code and can work in parallel. I am planning to shape the project as an Erlang/OTP application.

I will use rebar for building and for dependency management, eunit for unit testing and proper or triq for property testing.

I am planning to use the record syntax for retrieving information from a document and proplists or dict as a dictionary.

I also think, *qlc:table* interface can help to implement iterators.

What is new or different about your approach which hasn't been done or wasn't possible before?

Other bindings use SWIG. My approach is to use only *erl_driver* interface. It provides wide opportunities for customization. But it also requires more work and knowledges.

Do you have any preliminary findings or results which suggest that your approach is possible and likely to succeed?

Yes, I have.

There are few interfaces in Erlang for C/C++ developers: cnode, driver, linked-in driver and NIFs. I will use a linked-in driver to minimize latency. This interface can be replaced by a driver easily.

I will use a *gen_server* for encapsulation, because it is a time tested way to access to a shared object.

How useful will your results be when not everything works out exactly as planned?

Bugs in Erlang code can be easy to find because of both supervising and logging. I will left C++ code as simple as possible to minimize bugs. I will create tests for simple refactoring.

Project Timeline

April - May 21th

There are 4 weeks of "community bonding" after accepted students are announced.

Read more about Xapian and its API. Write examples by hands. Understand how blocks are working, splitting classes on groups.

Creating an OTP application. Make a rebar plugin for easy configuration of building process. Create functions to connect to a database.

Create datasets for testing. Develop and write tests.

May 21th - July 9th

The coding period consists of 7 weeks until the mid-term (July 9th).
  - **1 week:** Writing code for making queries: Simple queries.
  - **2 week:** Writing code for making queries: Parametrized queries.
  - **3 week:** Writing index code: Simple indexes.
  - **4 week:** Writing index code: Indexes with parameters.
  - **5 week:** Writing iterators or replacement for them.

July 10th - August 13th

5 weeks to the "suggested 'pencils down' date" (August 13th).
  - **1 week:** Managing exceptions and error handling. Writing documentation about error codes and exceptions.
  - **2 week:** Writing a pool with help of a pooler application. Testing supervisors.
  - **3 week:** Develop examples: simple index, simple search, console viewer.
  - **4-5 week:** Measuring efficiency, profiling, finding and fixing bottlenecks.

August 14th - August 20th

1 week to the "firm 'pencils down' date" (August 20th).
  - **1 week:** Fixing minor bugs, polish documentation.

Previous Discussion of your Project

http://comments.gmane.org/gmane.comp.search.xapian.devel/1840