Google Summer of Code 2016: Ahmia.fi - Search engine for hidden services

Organization: The Tor Project

skrish13
IRC Handle: skrish13 on #ahmia,#tor,#tor-dev on OFTC
Email: skrish1113@gmail.com
Github: https://github.com/skrish13
OTR Fingerprint: 0F826351 7B848155 A0735793 72E493AB B3DDF34D

Mentor:

Juha Nurmi (IRC Handle: numes on OFTC)

Email: juha.nurmi@ahmia.fi

What project would you like to work on?

Improving the hidden-service search engine ahmia.

1. Improving the UX/UI
    a. Switch many of the pages which are not in sync with the new design.
    b. Fix broken features
    c. Make an ergonomic design for general look and usability
    d. Add a page to provide more information regarding tor, hidden services and general FAQ.
    e. Auto completion of search query.
        i.   Use of jquery or javascript
        ii. Determine the pros and cons of the above versus existing plugins for the same like awesomplete.
        iii. Determine pros and cons of using a minimal search completion feature without the use of jquery vs the otherwise.
    f. Think and come up with a design implementation without the use of javascript.
    g. Write code for the above mentioned implementation if time permits
    h. Estimated time: 3+(1) weeks

2. Review Code and build automated testing
    a. Automated tests for the Django
        i. Use of modern test-suites to effectively encounter unknown bugs in the future
        ii. Use class based approach , unittest, a standard python library module.
    b. Run the test before the deployment of code
    c. Improve documentation.
        i. Current level of documentation is very less and one has to take more efforts to understand how it works, to build etc than required.
        ii. The aim is to provide beautiful documentation regarding every process which will also help us in getting valuable inputs from people as they can understand easily and can provide feedback more easily.
    d. Estimated time: 3 weeks

3. Automated recovery from broken processes
    a. Currently one has to manually kill and restart some of the broken processes which occur randomly.
    b. The aim is to automatically look for broken processes and kill them and reboot them without causing any other changes to the system.
    c. Estimated time: 1 week.

4. Show statistics and visualization.
    a. Understand the current setup in terms of gathering statistics
    b. Design ways to gather statistics without intruding with the privacy of the users.
    c. Implement the design.
    d. Build HTTP REST JSON API to share the statistics.
    e. Share the statistics in creative visualizations.
        i. Gather ideas from community regarding what they would like to know.
        ii. Publish statistics based on those feedback.

      f.   Estimated time: 2 weeks

5. More Ideas (If time permits, on the side/simultaneously)
    a. Gather feedback from community regarding search engine
        i. This is pertinent in the long run and is not limited to the summer project.
    b. Discuss ways to improve the search results, since Juha feels it is already giving good results and the scope for improvement is less. So, gathering ideas from many people will be a 1st step in this.
    c. Depending on the feedback, the format of the search results can be modified in the long run.
        i. Ie. The way in which search result is being displayed etc.
    d. Estimated Time: Simultaneous with other tasks.

Timeline:

April 1 - April 21:

    Minor bug fixes and getting to know more about the codebase

April 22-May 22:

    Community bonding period. Get feedback from community on the implementation of the ideas, get to know mentors, read documentation, get up to speed to begin working on the project.

Week 1:

    Fix the old pages which are not in sync with new design.
    Start asking feedback on the general usability of the website as mentioned above.

Week 2:

    Work on auto completion of search queries.

Week 3:

    Getting started with automated test cases

    Extra page for useful information as mentioned above

Week 4:

    Writing test cases

    Getting started with writing documentation.

Week 5:

    Continue with documentation.

    Collect feedback and brainstorm on the responses for the design of js free website.

Week 6:

    Bugfixes and catch up

    Getting ready for Mid term evaluation.

    Continue the design implementation.

    Mid term report on the initial feedback on the usability of the website, focussing on the long term improvement areas and small short term bug fixes or enhancements.

Week 7:

    Running automated test cases

    Getting started with fixing broken processes.

Week 8:

    Finishing up fixing broken processes.

    Getting started with collecting statistics, ideas and ways to do so

Week 9:

    Building the HTTP JSON API

    Gather feedback from community on what kind of statistic needed.

Week 10:

    Publish statistics

    Provide creative visualization.

Week 11:

    Catch up incase of any backlogs.

    If everything is finished start coding on the js free website design.

    Provide the final report on the feedback received on the usability and enhancement requests etc.

    Start planning, coding on bugs or feasible enhancements from the feedback

Week 12:

    Catch up and final touches. Final mentor evaluations.

Experience Related to the Project:

I have sound knowledge of search engines, mainly through one of my courses at the university - Information Retrieval. I got introduced to elasticsearch only recently, I am currently using it in one of my projects, implementing Cloudshark, where i have to configure the API service of elasticsearch to provide dynamic responses to the queries. I am fairly familiar with writing unit tests and will improve my ability on relevant parts by working on small bugs during the pre-GSoC phase. I've been using linux for more than 3yrs and won't have any problem in working in that environment.

Point us to a code sample:

Although i haven't done much work in ahmia, I spent most time in familiarising myself with their codebase. You can check out my minor PR here.

Why do you want to work with The Tor Project in particular?

I don't use tor a lot but I have huge appreciation for the work that you guys do. I've watched many of your talks, by Roger, Mike Perry, Jacob and others at Defcon, C3 and other conferences. I've been a spectator at some of your IRC meetings on #tor-project and was one of

the main reasons I became interested in developing for tor. The openness and respect among the members is huge and i would love to be one among them some day soon.

Tell us about your experiences in free software development environments.

I am relatively new to contributing to open source organisations but i have **always used linux** (even though i bought a gaming laptop :P) and the various free software that comes with it. I go out of my way to find and use free and open source software instead of commercial ones, just for the sake of supporting them. I have collaborated with peers on projects at my university, so I do have experience in working within a group.

Will you be working full-time on the project for the summer

Yes, full time

Will your project need more work and/or maintenance after the summer ends? What are the chances you will stick around and help out with that and other related projects?

Yes, I believe ahmia will need developing even after the summer. Since my interested field is related to information retrieval i hope to keep in touch with any new technologies and incorporate them in ahmia with the help of Juha, whenever needed. I'm also very much interested in contributing to one of the core tor projects at some point of time.

What is your ideal approach to keeping everybody informed of your progress, problems, and questions over the course of the project? Said another way, how much of a "manager" will you need your mentor to be?

I will share the status report bi-weekly on tor-dev mailing lists. I expect my mentor to be as a guide to clarify my doubts and review my ideas and would always keep him informed of the status of the project. Needless to say, i'll be using the github platform and git as version control.

How can we contact you to ask you further questions?

I'll be available on IRC but best way would be email.

Are you applying to other projects for GSoC?

Nope

Is there anything else that we should know?

Even if i don't get selected for GSoC, i would love to contribute to this project (assuming no one else got selected) or any other project which really needs developing and didn't get any candidates as part of GSoC. In short, I request you guys to make a **small notice** on **tor-dev** mailing list **or** on **the site,** mentioning which all projects got selected and along with that a list of projects that **you strongly feel** the need of volunteers. I would love to contribute if i can, to any of those projects. I would also like to thank Juha, the mentor of this project, who helped me a lot in writing this proposal.