Connexions Google Summer Of Code Student Application

Name: Saket Kumar Choudhary
University: Indian Institute of Technology Bombay,Mumbai,India
Current Enrollment Information: Third Year Undergraduate at Chemical Engineering Department
Physical Address:
A 19 Hastinapur
Anushakti Nagar
Mumbai
Maharashtra
Pin/Zip - 400094
India
EMail Address:
[1] saketkc@gmail.com, [Preferred]
[2] saket.kumar@iitb.ac.in
Contact : +91-9869649197[Mobile]

Connexions Project You Are Interested In:Slide Importer client using the OERPub publishing API

1. **List your programming skills as related to the project description**

I am a 20 year old developer,a FOSS contributor/supporter and a Python & Ruby enthusiast. I am currently pursuing my undergraduate studies at Chemical Engineering Department of Indian Institute of Technology Bombay,India. My topics of special interest are WebFrameworks , Algorithms and Software-Design.

I have been programming for three years now . I started out with C/C++ in my first year of undergraduate studies and moved ahead to try my hand at PHP, Python and Ruby. I am comfortable with MVC Frameworks(CakePHP,Django,RAILS) and have also worked with microweb frameworks like Sinatra and Flask. I am comfortable with HAML and SASS too.

I was also involved with the IIT Bombay Satellite Project – Pratham [http://www.aero.iitb.ac.in/pratham] where I implemented Analog to Digital conversion (ADC) code using ATMEGA32 microcontroller for the 'Communication' subsytem to convert incoming voltage values to Total Electron Count values. I served as a member of the On Board Computer Subsystem involved with Software and Hardware testing.

I maintain a small blog on Django Techniques: **http://thedjangoway.blogspot.in/**

CV : http://home.iitb.ac.in/~saket.kumar/saket_cv.pdf

2. Have you worked on an open source project before? If so, give the project name and describe the work you did.

I have submitted few patches as listed down here:
http://home.iitb.ac.in/~saket.kumar/bugsandpatches.html

Some of my Open Source Projects so Far have been :

1. **Scilab on Cloud**
[https://github.com/saketkc/scilab_cloud]:
I developed a Python based application that would enable runnign Scilab[http://scilab.org/] a scientific computing software. This app allows user to run his/her Scilab codes online through a browser, thus removing the need to install Scilab Client locally.I soleley developed thi app on my own

2. **IIT Bombay Grading System on SMS**
[https://github.com/saketkc/iitb-library-sms-interface]
In a team of 4 I developed a Flask[http://flask.pocoo.org/] based app to scrape through our institute's Grading interface and send the Grades to a user on his mobile on request. I wrote the Scraping function making use of Beautiful Soup and urlib2 libraries.

3. **Pivotal Tracker Email Bot**
[https://github.com/saketkc/pivotal-tracker-email-wrapper] :
Pivotal Tracker(http://pivotaltracker.com) is an online Agile project Management tool. I sloely developed an interface using their API to create "New Issues" by fetching the emails from a gmail account. This has been recognised as a third party tool by Pivotal :http://www.pivotaltracker.com/help/thirdpartytools.

4. **DropBox on SMS**
[https://github.com/saketkc/dropbox_on_sms]:

This "Sinatra"[http://www.sinatrarb.com/] based app that lets one send a file from your dropbox folder to any email id , just with one SMS. This app stood among the Top 20 at Yahoo! Open Hack India [September 2011]. I was the sole developer of this complete app.

5. **kicad-ngspice**
[https://github.com/saketkc/kicad-ngspice]

Implemented Python Code to convert KICAD generated netlist to ngspice friendly format, the original netlist could not be exported directly to ngspice due to improper node configurations . Also made a GUI using PyGTK for making a file editor.

6. **SlideShare Hacks**

[https://github.com/saketkc/SlideShare-Instants] and [https://github.com/saketkc/SlideShare-Desktop-app]

I had made a few hacks using SlideShare API, one of them made a GUI on the Desktop for being slideshows[was unsuccessful] and the other one would work like Google Instant to fetch slideshows instantly from SlideShare[successful]

7. **Undergraduate Academics Portals:**
[**https://github.com/saketkc/ug_acads**]
Creating a set of portals for IIT Bombay such as a Book Donating Portal(/bookbay) and a Summer Projects Portal(/ispa) for my home institute. This is first time that the websites related to the institute have been open sourced.

Some of my other projects can be found at :
https://github.com/saketkc/

I did my Summer Internship,2011 at SlideShare one of the biggest Ruby on Rails site in the world [a platform for SlideShow uploading and viewing online] wher I designed and implemented Admin Interface to allow super user actions like delete,suspend,view users and the slideshows. I also implemented specs (Behaviour Driven Development(BDD)) for the Ruby code using RSPEC.My expeerinces from the same are featured here :http://engineering.slideshare.net/2011/09/applying-for-a-slideshare-internship-read-this-first-hand-account/

3. **What do you hope to gain from participating in this project?**

I use online resources myself for my own studies , and such I believe Connexions is a great contributor to the society by providing a platform for everyone to refer Classroom notes , presentations and related stuff. Contributing my code towards the betterment of Connexions will be a privilege for me , my way of contributing back to the community

I would gain an indepth knowledge about writing scalable and modular code in Python . Though I have done a lot of projects involving Python and JS , but none of them have taught me to write scalable code. As this project would actually go live on a Web Server and would be millions of people, not only it should be scalable but thoroughly tested to .
presentations and related stuff.

Besides that here are my own visions for the project.:

1. **Conversion Of Hand Written Notes:**

A lot people using tablets make their notes i nform of Windows Journals or scanned papers of original handwritten notes. Right now the Module Upload API doesn't support PDFs and OCR. A lot of content can be generated of the API can be made open for reading PDFs and Images[

scanned papers].

2. **Using EPubs format for uploading :**

A lot of people have shifted entirely to epubs for reading and writing notes. The API needs to be extended to allow for EPUB Conversion to say a .doc and then upload it as a ".doc" . Though this is already part of another GSoC project in the form of rendering epubs, uploading is also an essential feature.

3. **Generating more content from SlideShare:**

We can search SlideShare for a particular tag and for every slideshow that is tagged say "science and technology" , fetch the user contact information from slideshare and then send him/her a request form asking to upload his/her content on cnx.org.
This can be automated in the sense that the user agrees to the CC license by replying to the request mail from cnx.org and then the uploading takes place on its own.

4. **Add a short outline of how you will approach your selected project. You could include things like a timeline or a design if needed.**

Action Plan:

Before May 21st:

1. Discuss ideas about introducing Facebook/Twitter connect with the mentors.
2. This would make sharing more easy and more content oriented
3. Read and get acquainted with the Pyramid Framework
4. Discuss with the mentors a testing framework/method to be used for testing the functionalities of the website
5. Test Driven or Behaviour Driven /
6. Study the existing code of the OERPPub Publishing API
7. I have already forked the code on Github and have built it locally. I am currentyl studying the Pyramid framework before I start twaeking the code
8. Try fixing the pre-existing bugs

Week 1 May 21st - May 28th:

1. **Read the Google Docs API and CNXML documentation**
   I am well acquainted with the SlideShare API , but Google Docs API would be new for me. I would familiarise myself with Google Docs API documentation.

2. **Discuss and set expectations of code – dependencies, packages required code style, testing framework , and documentation.**

Week 2 May 29th - June 5th and Week 3 June 6th- June 13th :

1. **Develop a module for uploading slides into cnx.org with introductory**

paragraphs

This module will be a basic uploader implemented using the Pyramid framework . This will enable one to upload pdfs from one local PC to our servers. The text extraction can be done using SlideShare's APIs.

2. **Write test cases for the module**

Week 4 June 14th - June 21st

1. **Use the SlideShare and Google Docs API to make modules can be made embeddable**

   Using the SlideShare and Google Docs API and upload the documets to these services. These services have an oEmbed API which can be used to make the slideshows embeddable, and Slidehsres API can be used to extract the text from the slides

2. **Write test cases for the plugin**

Week 5 June 22nd - June 29th

1. **Create the link box to download the original slides**

   The download link can either point to cnx.org servers or to the slideshows uploaded pubically to SlideShare/Google Docs API. The structure of the url can be something like this :

   username + slideshowname + slidenumber +".pdf"
   where slidenumber will be any number from 1- total slides

2. **Using the publishing API, write code to publish the module on cnx.org**

3. **Write test cases for the code**

Week 6 June 30th - July 6th

1. **Implement slide by slide printing option for printing to either .pdfs or .epubs**

   User should be able to print out individual slides

2. **Write test cases for printing**

Week 7 July 7th - July 14th and Week 8 July 15th- July 22nd(optional to be done only if the above timeline is met on time)

1. **Implement Audio/Video toolbox along with the slide player to go along with the slides**

   For Video/Audio Slides we need to extract the .mps/.flv and develop a toolbox to track and sync the audio /video with the slideshow

2. **Write tests for the above model**

Week 9 July 23d - July 30 (OPTIONAL to be done only if the above timeline is met on time)

1. **Implement the sites mobile version**

It is essential that the site looks good in its mobile version to.. This would involve some CSS fixes

2. **Test slide rendering in mobile version**

Week 10 July 31st - August 6th.[OPTIONAL]

1. **Implement OCR to allow uploading of images/scanned notes**
   OCR will give people the liberty to upload their handwritten notes
2. **Test OCR code**

Week 11th August 7- August 13th

1. **Document code**
2. **Bug Fixing**

Week 12 August 14th - August 20

1. **Test the functionality of the website**
2. **Complete Documentation**
3. **Bug Fixes**

Deliverables

At the end of My GSoC Timeline I will be able to deliver the following things to Connexions:
1. A module for uploading slides into cnx.org with introductory paragraphs

2. A box with download links to the original slides

3. A module that publishes the uploadee slides on cnx.org

4. Make modules embeddable using SlideSHare/Google API

5 Slide by slide printing option for printing to either .pdfs or .epubs

Following are the things that I will try to implement additionally incase I am successful with the above 5 main objectives:

1. Implement an Optical Character Reader to facilitate importing of handwritten notes

2.Implement Audio/Video toolbox along with the slide player to go along with the slides

3. Implement a mobile version of the website

Obligations

I have no obligations during the summers and would be able to contribute y full time to GSoC. I will be available full time on GTalk and IRC