The Collector
(a.k.a. libcollector)

- Aditya Kumar Praharaj (121030012)
Team : Deus Ex Machina

Wikipedia, OCW, Coursera, course-pages and internet in general serves as a great tool for us apart from the general textbooks and class notes to learn a certain subject and clear related doubts. For doubt clearing, there also exist different sites and forums (e.g. stackoverflow.com and physicsforums.com) which are huge databases of questions ranging from Computer Science to Chemistry. However till date there is no such website or tool available which unifies the benefit of all these websites and resources into a single structured readable resource. The goal of this project is to create such an online backend library which can allow users to do so.

This online backend will collect the available online educational resources such as slides, video lectures, wiki articles, quiz papers and other stuff and somehow structure it in the form of an easily readable book, as well as a database which will enable users to directly search for doubts or similar questions. Obviously heavy usage of the search technologies and machine learning will be used, both of which I am familiar with already. This is what is planned to be achieved on the client side: Say we are reading Calculus and are stuck on a particular part of the stokes theorem. This tool will take a structured input containing Stokes Theorem and then search the web to actually spit out a structured webpage (like a PokeDex, if you have seen Pokemon, or Something like Wolfram Alpha) which will contain detailed information about Stokes theorem, direct links to sites such as StackOverflow and other helpful sites which will help him understand the stuff more better. As far as the UML class diagram is concerned, it will be planned in the project itself, however the features are listed.

NOTE : This is supposed to be a library which can be used to create educational softwares later. The sole goal of this project is to build such a library and a demo software which exhibits the capabilities, not any fancy software which will be thrown away after this project.

Contents of the library

1. An online backend which will utilize services such as Dropbox and Google Drive as the storage space. An API to both of them will be used.

2. This backend will be powered by a collection of machine learning modules (I plan to use shogun-toolbox, a leading machine learning library for this project) and NLP libraries (such as Wordnet) for finding relevant terms for searching. My current background with shogun will help in easing the integration.

3. As far as web crawling and search APIs are concerned, open source web crawlers and processors such as DBPedia and Xapian will be used. Implementation specific details are omitted for the sake of simplicity of the proposal.

Timeline

- First Week : Getting acquainted with required APIs and libraries (already mentioned earlier). Also in this week itself, the UML framework designs will be planned and finalized. Only study this week.
- Second - End of Fourth Week : Start implementing the skeleton of the framework. The exact details can only be discussed after the framework design which will be done later. This week will include debugging the library and building various unit tests for testing.
- Fifth Week : Start implementing the frontend. This will not take much time owing to the library itself. However I have to learn PhotoShop and stuff (which I hate!!!).
- Sixth Week - End : Documentation and presentation.

Components required?

Guts and patience to help convert lots of chinese and snacks to useful code!!!

Why am I doing this project and what I expect to get out of it?

I like tinkering with open source libraries and in developing them as well. I have contributed in a few open source projects, namely PoDoFo, SDL and shogun-toolbox. Till now I only have tinkered with stuff and not built something credible. With the help of this promising project, I will not only sharpen my programming skills, but also get a knowledge about other APIs and basics that I am unaware of right now.The reason I am doing this project alone is I wanted to test my own development skills. Apart from that, if this project is successful I also plan to open source it and gather other developers who may be interested.