Ideal Climate Change Project

Creating a searchable archive of climate change webpages & tweets

**Contact person and their contact information:**

Dr. Edward Fox

Email: fox@vt.edu

Mohamed Magdy Gharib Farag

Email: [mmagdy@vt.edu](mailto:mmagdy@vt.edu)

**Project description (paragraph):**

The IDEAL project needs a way for non-technical users to search through archived tweets and webpages on the topic of climate change. Users will benefit by being able to search by keyword and by tweet frequency, and will use this information for research. Currently. the IDEAL project has archived months of tweets and is continuing to archive. They, the client, need any URLs extracted from these tweets, and the resulting tweets & URLs need to be indexed in a search engine. The search engine’s functionality will be fine-tuned depending on the client’s feedback.

**Project deliverables:**

* Finish extracting web pages from archived tweets
* Have all data extracted from tweets & webpages - ready to index in search engine
* Set up Blacklight & have a *barebones* search engine up and running.
* Configure Blacklight to make it prettier/easier to use.
* Add functionality in Blacklight to search webpages **and** tweets

**Expected impact of the project:**

* Social science researchers will be able to use the climate change search engine we create to quickly and easily:
  + See the change in the frequency of climate change
  + Search for keywords in climate change webpages & tweets
  + Correlate trends in climate change webpages & tweets to marketing campaigns and current events

**Required skills of the team engaging in the project:**

* Python - used to create scripts to manipulate Gigabytes of climate change data
* Ruby & Java - the languages used in the open source search engine & it’s GUI
* Client/server web architecture - search engine will function as web application

**URLs to follow for additional information:**

* <http://hadoop.dlib.vt.edu/>
  + Link to the current climate change archives

**Type of project** (digital library, digital preservation activity, video, construction effort in Second Life, game, web site, database, …):

* Digital library & search engine

**Signup information**

* Email any of us to join this project (pids are listed below)

**List of people interested in this project**

Name : kafley,Somn,

Email/PID : [somkk13@vt.edu](mailto:somkk13@vt.edu)

Phone: (434) 760-4115

Name: Michael Steele

Email/PID: [derek22@vt.edu](mailto:derek22@vt.edu)

Phone: (804) 426-7779

Name: Samyak Singh

Email/PID: [ssingh94@vt.edu](mailto:ssingh94@vt.edu)

Phone: (214) 695-8587

Roles each person would like to play

* We will not be able to decide on roles until we get further into the project.

Timeline plans, giving milestones at least twice per month

* Current plan: get a barebones search engine up and running as fast as possible
  + 2/14 - Finish extracting web pages from archived tweets
  + 3/1 - Have all data extracted from tweets & webpages - ready to index in search engine
  + 3/14 - Set up Blacklight & have a *barebones* search engine up and running. Will be delivered to client.
  + Further deadlines will be created after feedback from the 3/14 project delivery.

Date that client has signed off on this:

Date that instructor has signed off on this: