```
(12:09:44) David Gucwa: k
(12:45:15) David Gucwa: There's a hitch in the plan to crawl through the
profiles in advance
(12:45:43) David Gucwa: when logged in, I can only seem to access profiles that
are on my friends list or at the same school as I am
(12:47:17) David Gucwa: so I think we're going to have to spider them on the
spot when people import
(12:47:24) David Gucwa: and cache them then
(12:59:32) dr ttol: wait
(12:59:42) dr ttol: atleast we can index entire schools
(12:59:43) dr ttol: with just one email
(12:59:44) dr ttol: right
(12:59:52) dr ttol: so we just need to get you 300 logins
(12:59:55) dr ttol: or so
(13:00:03) dr ttol: to access over 1 mil profiles
(13:00:48) David Gucwa: yeah
(13:01:03) David Gucwa: if you can just give me one big file with all the logins
and passwords then I can use that
(13:05:07) dr ttol: can you use the one we gave you
(13:06:12) David Gucwa: I'm not sure this is going to work actually
(13:06:32) David Gucwa: I just timed the script, and it takes about 5 seconds to
fetch a profile
(13:06:44) David Gucwa: for a million profiles that's 57 days to index them all
(13:06:48) dr ttol: thats fine
(13:07:33) dr ttol: we just want to get as much as we can
(13:07:47) dr ttol: i mean, if we have all of umass
(13:07:51) dr ttol: thats great
(13:07:57) dr ttol: prob only take a few hours
(13:08:33) David Gucwa: well, except I don't just have a list of ids that belong
to umass
(13:08:36) David Gucwa: I have to try each one
(13:08:52) dr ttol: what do you mean
(13:09:34) dr ttol: with one umass login and password
(13:09:38) dr ttol: you're able to see the entire school
(13:09:57) David Gucwa: If I type in the id of a profile, and that person is
from umass, then I can see it
(13:10:05) David Gucwa: but there's no way to know which ids are from umass
(13:10:14) dr ttol: if you click browse
(13:10:21) dr ttol: you get a list of users
(13:10:30) dr ttol: that are from you school
(13:11:17) David Gucwa: where does it say browse
(13:11:27) dr ttol: give me a login and password
(13:11:51) David Gucwa: lbowman@wellesley.edu divya
(13:12:52) dr ttol: social net
(13:12:56) dr ttol: is browse
(13:12:58) dr ttol: it is randomly generated
(13:13:01) dr ttol: so you have to check for duplicates
(13:14:39) dr ttol: see it?
(13:15:05) David Gucwa: yeah
(13:15:09) David Gucwa: let me see how fast I can grab ids from that
(13:15:13) dr ttol: k
(13:15:18) dr ttol: and that "next" link
(13:16:06) David Gucwa: next is just the same as reloading
(13:16:38) dr ttol: k
(13:20:10) David Gucwa: I can get 3 or 4 ids a second
(13:21:07) David Gucwa: How many people from umass would you say are on
thefacebook? A few thousand?
```

Conversation with drttol at 2005-01-27 16:44:12 on David Gucwa (aim)

(16:44:12) dr ttol: status?

(16:44:23) David Gucwa: caching umich

(16:44:36) dr ttol: how fast?

(16:44:37) David Gucwa: I'm going to put the script on the server and see if it goes any faster

(16:44:43) David Gucwa: it's been going for like 10 minutes

(16:44:49) dr ttol: how many profiles

(16:45:11) dr ttol: tfb is run by a techie, so becareful

(16:45:25) David Gucwa: there are two steps, it retrieves all the id's first, then it gets the profile for each id

(16:45:30) David Gucwa: it's still on the first step, up to 6000 ids

(16:45:51) dr ttol: ok, does it get the pictures as well

(16:46:29) David Gucwa: it gets the urls of the pictures

(16:46:44) dr ttol: what about for the importer

(16:48:23) David Gucwa: same

(16:48:42) David Gucwa: it displays the pictures when you're choosing which one you want, but it doesn't store the picture anywhere locally yet

(16:49:02) dr ttol: when they pick the picture, does it store it locally

(16:49:05) dr ttol: to the cu database

(16:49:24) David Gucwa: no

(16:49:30) David Gucwa: I don't know where they're supposed to go

(16:49:37) dr ttol: ask joel

(16:50:04) David Gucwa: yeah that's something I'll need to take care of eventually

(16:50:13) David Gucwa: but I'm trying to get the crawler finished up for now

(16:50:27) dr ttol: ok

(17:27:05) dr ttol: status

(17:27:11) David Gucwa: importing 4 schools

(17:27:16) David Gucwa: it's taking a while

(17:27:17) dr ttol: all profiles?

(17:27:26) David Gucwa: I think it's actually going slower on the server

(17:27:28) David Gucwa: than on my computer

(17:27:35) David Gucwa: still just grabbing id's

(17:28:00) dr ttol: hm

(17:28:03) dr ttol: which server are you using

(17:28:19) David Gucwa: 69.44.59.182

(17:28:27) dr ttol: it could e that the schoools have more ids than you think

(17:28:40) dr ttol: so the crawler works?

(17:28:47) David Gucwa: yeah

(17:29:22) dr ttol: which universities

(17:29:24) dr ttol: do we have umass

(17:29:29) David Gucwa: no I don't have a umass login

(17:29:38) dr ttol: rdegutis@student.umass.edu

(17:29:40) dr ttol: ruth1783

(17:29:49) David Gucwa: I'm getting umich, cornell, yale and harvard

(17:30:06) dr ttol: ok

(17:30:07) dr ttol: get umass next

(17:30:40) David Gucwa: k

(17:31:00) dr ttol: we want to import the top i2hub schools first

(17:31:48) David Gucwa: ok

(17:31:54) David Gucwa: well I've used all the logins you gave me

(17:32:09) David Gucwa: so if you give me more logins I can start importing

(17:32:13) dr ttol: ok

(17:32:14) dr ttol: you mean

(17:32:17) David Gucwa: also I'm waiting for wellesley to start showing emails again

```
(17:32:19) dr ttol: you've imported all the ones i gave you
(17:32:33) David Gucwa: yes
(17:32:37) David Gucwa: in the process of, at least
(17:32:41) dr ttol: ok
(17:32:43) dr ttol: yee@bu.edu
(17:32:45) dr ttol: ab1036
(17:34:00) David Gucwa: ok I started that one
(17:34:22) dr ttol: check to make sure the logins are valid
(17:34:23) dr ttol: im not sure
(17:34:35) dr ttol: barroseu@bc.edu
(17:34:39) dr ttol: eutychius
(17:34:50) dr ttol: beelzebub@brown.edu
(17:34:53) dr ttol: asdfgh
(17:36:50) dr ttol: hoganc@alum.mit.edu
(17:36:52) dr ttol: csfb
(17:36:58) dr ttol: porcella@princeton.edu
(17:36:58) dr ttol: angelo
(17:37:31) dr ttol: carlsson@stanford.edu
(17:37:35) dr ttol: fake_pwd
(17:37:46) dr ttol: maria.rocha_oliveira@tufts.edu
(17:37:47) dr ttol: maria
(17:37:54) dr ttol: Stinab8@ucla.edu
(17:37:55) dr ttol: divya
(17:38:15) dr ttol: Sen4@georgetown.edu
(17:38:20) dr ttol: sexyrena
(17:38:34) dr ttol: rl6@duke.edu
(17:38:35) dr ttol: nacho
(17:38:38) dr ttol: some may not work
(17:41:04) David Gucwa: k
(17:44:45) David Gucwa: yeah 3 or 4 didn't work
(17:44:47) David Gucwa: I started the rest
(17:45:05) dr ttol: which ones didnt
(17:46:39) David Gucwa: rl6@duke.edu Sen4@georgetown.edu porcella@princeton.edu
hoganc@alum.mit.edu
(17:47:59) David Gucwa: it's got 46,000 id's so far and still going
(17:48:08) David Gucwa: in total from all of them
(17:48:48) dr ttol: we need the profiles
(17:48:52) dr ttol: not just ids
(17:49:45) David Gucwa: I know
(17:50:10) dr ttol: singer.d@neu.edu
(17:50:12) dr ttol: 1211
(17:50:39) dr ttol: mj405@nyu.edu
(17:50:41) dr ttol: rajubros
(17:52:09) dr ttol: how many proiles?
(17:52:21) David Gucwa: none yet, I'll start that
(17:52:39) dr ttol: ok
(17:55:33) dr ttol: let me know when we hit 100 profiles
(17:56:54) David Gucwa: we've probably hit that by now
(17:57:19) dr ttol: we're prob setting up huge alarms at thefacebook
(17:57:30) dr ttol: how can we go faster
(17:58:22) dr ttol: procella@princeton.edu
(17:58:23) dr ttol: frankmaria
(17:58:56) David Gucwa: I'm not sure we can go any faster
(17:59:14) dr ttol: what if we had the scripts running from other computers at
te same time
(17:59:32) David Gucwa: do we have the proxies available yet
(17:59:36) dr ttol: yes
```

```
(17:59:54) dr ttol: unix15.dmbhosting.com
(17:59:58) dr ttol: i2hub / Importer123
(18:00:05) dr ttol: 66.96.217.229
(18:00:13) dr ttol: i2hub / TX6ADGXIAHR
(18:00:19) dr ttol: 67.131.250.102
(18:00:23) dr ttol: i2hub / 4it2AXqi
(18:00:28) dr ttol: 67.18.33.226
(18:00:29) David Gucwa: do these have ssh access
(18:00:33) David Gucwa: I can't seem to get into the first one
(18:00:34) dr ttol: i2hubtes / Importer123
(18:00:41) dr ttol: 69.56.226.102
(18:00:44) dr ttol: i2hub / importer12
(18:00:51) dr ttol: i dont know about ssh
(18:00:52) dr ttol: some may
(18:00:53) dr ttol: some may not
(18:00:55) dr ttol: all have ftp
(18:01:04) David Gucwa: hm
(18:01:09) David Gucwa: ok
(18:07:07) David Gucwa: procella@princeton.edu doesn't work
(18:10:29) dr ttol: porcella
(18:10:32) dr ttol: cameron typo'd
(18:10:37) David Gucwa: ok
(18:12:32) dr ttol:  cmg25@georgetown.edu
(18:12:35) dr ttol: top59gun
(18:16:47) dr ttol: how many profiles so far?
(18:19:07) David Gucwa: 5300
(18:19:27) dr ttol: what schools?
(18:20:00) dr ttol: or is that combined
(18:20:46) David Gucwa: it was all from one school, but I'm checking them now
and the emails aren't there
(18:20:55) David Gucwa: they must have just gotten turned off
(18:21:04) dr ttol: fuck
(18:21:30) dr ttol: what school
(18:21:52) David Gucwa: umich
(18:22:16) dr ttol: they just got turned off?
(18:23:25) David Gucwa: wait not umich
(18:23:36) David Gucwa: yale
(18:23:58) David Gucwa: maybe they were never turned on, I'm not sure I checked
(18:24:01) David Gucwa: I thought I did though
(18:24:02) dr ttol: ok
(18:24:13) dr ttol: try the ones with emails
(18:24:16) dr ttol: not the ones without
(18:25:22) David Gucwa: ok I'm going to try bu
(18:25:26) David Gucwa: I just made sure the emails are showing
(18:25:26) dr ttol: ok
(18:26:25) dr ttol: see how many we can get
(18:26:37) dr ttol: did you try the other shells
(18:27:16) David Gucwa: yeah I've got one importing from princeton
(18:27:22) David Gucwa: most of them didn't have ssh open
(18:27:30) dr ttol: you need ssh?
(18:27:35) dr ttol: you said ftp is ok
(18:28:13) David Gucwa: yeah ftp will work
(18:28:19) David Gucwa: I just need to modify some things
(18:28:31) David Gucwa: ssh is just easier
(18:35:36) dr ttol: status
(18:35:38) dr ttol: on bu caching
(18:37:00) dr ttol: do the pton and gtown emails first
```