

Mailing List Data - Part I

In this assignment you will download some of the mailing list data from <http://mbbox.dr-chuck.net/> and run the data cleaning / modeling process and take some screen shots. All screenshots should use the same data set.

1. Download Content.sqlite

DB Browser for SQLite - C:\Users\mirey\OneDrive\Desktop\content.sqlite

File Edit View Tools Help

New Database Open Database Write Changes Revert Changes Open Project Save Project Attach Database Close Database

Database Structure Browse Data Execute SQL

Table: Messages Filter in any column

	id	email	sent_at	subject	headers	body
	Filter	Filter	Filter	Filter	Filter	Filter
1		1 ggolden@umich.edu	2005-12-08T23:34:30-06:00	call for participation: developers ...	From news@gmane.org Tue Mar 04 ...	At today's most excellent developers ...
2		2 csev@umich.edu	2005-12-09T00:58:01-05:00	report from the austin conference: ...	From news@gmane.org Tue Mar 04 ...	At one of the sessions this week in ...
3		3 kevin.carpenter@rsmart.com	2005-12-09T09:01:49-07:00	cas and sakai 1.5	From news@gmane.org Tue Mar 04 ...	Sorry if I am a bit off topic for the list...
4		4 michael.feldstein@suny.edu	2005-12-09T09:43:12-05:00	re: lms/vle rants/comments	From news@gmane.org Tue Mar 04 ...	Yup, I think this is spot-on. Either/or,...
5		5 john@caret.cam.ac.uk	2005-12-09T13:32:29+00:00	re: lms/vle rants/comments	From news@gmane.org Tue Mar 04 ...	I should chip in her as Dan's PHB :)...
6		6 s gilthens@northwestern.edu	2005-12-09T13:32:31-06:00	re: sakaiportallogin and presense	From news@gmane.org Tue Mar 04 ...	I think some things might only work l...
7		7 john@caret.cam.ac.uk	2005-12-09T13:42:24+00:00	re: lms/vle rants/comments	From news@gmane.org Tue Mar 04 ...	BTW some of the thinking Dan ...
8		8 s gilthens@northwestern.edu	2005-12-09T15:23:09-06:00	re: sakaiportallogin and presense	From news@gmane.org Tue Mar 04 ...	Yuji Shinozaki wrote:...
9		9 ys2n@virginia.edu	2005-12-09T17:54:17+00:00	sakaiportallogin and presense	From news@gmane.org Tue Mar 04 ...	I am trying out Chuck's neat little ...
10		10 ys2n@virginia.edu	2005-12-09T21:16:39+00:00	re: sakaiportallogin and presense	From news@gmane.org Tue Mar 04 ...	The portlet pulls in the "worksite" ...
11		11 slt@columbia.edu	2005-12-09T23:16:40-05:00	re: cas and sakai 1.5	From news@gmane.org Tue Mar 04 ...	Kevin Carpenter wrote:...
12		12 dg cotton@ucdavis.edu	2005-12-10T20:55:54-08:00	memory error with 2.1	From news@gmane.org Tue Mar 04 ...	I keep getting this memory error afte...
13		13 vyu2@yahoo.com	2005-12-10T23:34:43-08:00	re: memory error with 2.1	From news@gmane.org Tue Mar 04 ...	Sakai does not run well in Windows. ...
14		14 winstonw.tw@gmail.com	2005-12-11T03:07:29+08:00	re: sakai ldap authentication	From news@gmane.org Tue Mar 04 ...	Could somebody provide me some ...
15		15 alpoland@iupui.edu	2005-12-11T12:08:15-05:00	re: memory error with 2.1	From news@gmane.org Tue Mar 04 ...	Dan,...
16		16 caseyd1@stanford.edu	2005-12-12T07:02:36-08:00	re: sakai 2.1 provider examples	From news@gmane.org Tue Mar 04 ...	Thanks Aaron, I've been wondering ...
17		17 aaronz@vt.edu	2005-12-12T08:55:09-05:00	sakai 2.1 provider examples	From news@gmane.org Tue Mar 04 ...	I put up the code in contril for the V...
18		18 j.higham@hull.ac.uk	2005-12-12T09:49:16+00:00	re: sakai 2.1 - lone zero block error	From news@gmane.org Tue Mar 04 ...	Thanks everyone. It does appear to b...
19		19 aaronz@vt.edu	2005-12-12T10:13:38-05:00	re: sakai 2.1 provider examples	From news@gmane.org Tue Mar 04 ...	I should mention that we have a ...
20		20 jeffrey.beeman@asu.edu	2005-12-12T11:14:55-07:00	re: sakai logo	From news@gmane.org Tue Mar 04 ...	Not sure if you guys were at the Aust...
21		21 csev@umich.edu	2005-12-12T11:33:29-05:00	sepp library discussion group rename...	From news@gmane.org Tue Mar 04 ...	The group formerly knows as the SEP...
22		22 pgoldweic@northwestern.edu	2005-12-12T13:16:12-06:00	problems accessing the collab site wi...	From news@gmane.org Tue Mar 04 ...	Hi,...
23		23 pgoldweic@northwestern.edu	2005-12-12T14:20:41-06:00	re: problems accessing the collab sit...	From news@gmane.org Tue Mar 04 ...	You might be right in that it may be a...
24		24 andrew@caret.cam.ac.uk	2005-12-12T14:23:12+00:00	file picker for non-legacy tools	From news@gmane.org Tue Mar 04 ...	Hi!...
25		25 john.ellis@rsmart.com	2005-12-12T14:24:48-07:00	re: file picker for non-legacy tools	From news@gmane.org Tue Mar 04 ...	Andrew,...

1 - 25 of 59823

Go to: 1

2. Run gmodel.py to create index.sqlite

```
Command Prompt

C:\Users\mirey\OneDrive\Desktop>gmodel.py
Loaded allsenders 1803 and mapping 29 dns mapping 1
1 2005-12-08T23:34:30-06:00 ggolden22@mac.com
251 2005-12-22T10:03:20-08:00 tpamsler@ucdavis.edu
501 2006-01-12T11:17:34-05:00 lance@indiana.edu
751 2006-01-24T11:13:28-08:00 vrajgopalan@ucmerced.edu
1001 2006-02-02T08:27:30-07:00 john.ellis@rsmart.com
1251 2006-02-15T19:13:17-05:00 ggolden22@mac.com
1501 2006-02-23T16:49:00-05:00 csev@umich.edu
1751 2006-03-13T13:48:36-05:00 csev@umich.edu
2001 2006-03-27T13:34:32-06:00 swgithen@mtu.edu
2251 2006-04-05T12:38:43-04:00 ccount@mit.edu
2501 2006-04-17T13:28:54-07:00 jholtzman@berkeley.edu
2751 2006-04-25T14:26:03-07:00 vrajgopalan@ucmerced.edu
3001 2006-05-03T13:49:11-04:00 cheryl.wogahn@yale.edu
3251 2006-05-10T17:14:01+00:00 nuno@ufp.pt
3501 2006-05-17T10:37:49-04:00 ys2n@virginia.edu
3751 2006-05-23T14:52:13-04:00 azeckoski@unicon.net
4001 2006-06-06T12:51:32-04:00 ajpoland@indiana.edu
4251 2006-06-14T08:28:01-04:00 lance@indiana.edu
4501 2006-06-21T14:31:06-05:00 mustansar@rice.edu
4751 2006-06-29T15:52:15+01:00 antranig@cam.ac.uk
5001 2006-07-12T11:50:55-04:00 markjnorton@earthlink.net
5251 2006-07-20T17:24:14-07:00 sinou@etudes.org
5501 2006-07-31T18:45:12-04:00 ggolden22@mac.com
5751 2006-08-08T07:44:09-04:00 dave.ross@gmail.com
6001 2006-08-14T15:32:02-04:00 daisy.flemming@gmail.com
6251 2006-08-23T02:02:38-12:00 woleraymond@enownow.com
6501 2006-08-31T13:53:24-04:00 ryaneyby@gmail.com
6751 2006-09-12T17:52:18-05:00 zach@airplanesoftware.com
7001 2006-09-19T08:34:39+01:00 b.steele@blueyonder.co.uk
7251 2006-09-27T10:56:27+02:00 wilfrid.labeta@gmail.com
7501 2006-10-06T10:03:19+02:00 b.toeter@uva.nl
7751 2006-10-17T10:09:17-07:00 sfischbein@ucdavis.edu
8001 2006-10-27T12:40:07+01:00 nuno@ufp.pt
8251 2006-11-08T10:16:19+00:00 harriet@cam.ac.uk
8501 2006-11-19T17:08:08+01:00 jcreyesbe@ya.com
8751 2006-11-30T16:03:54-05:00 markjnorton@earthlink.net
9001 2006-12-13T15:35:15-05:00 jmpease@syr.edu
9251 2006-12-27T17:22+00:00 nuno@ufp.pt
9501 2007-01-12T12:22:47+00:00 matthew.buckett@ox.ac.uk
9751 2007-01-24T22:26:10-05:00 ys2n@virginia.edu
10001 2007-02-03T10:14:31-05:00 jimeng@umich.edu
10251 2007-02-14T10:04:45+00:00 liam@k-int.com
10501 2007-02-23T10:19:58-05:00 ajpoland@indiana.edu
10751 2007-03-05T09:01:35-08:00 caseydl@stanford.edu
11001 2007-03-14T08:12:18-04:00 cheryl.wogahn@yale.edu
11251 2007-03-21T11:34:36-04:00 csev@umich.edu
11501 2007-04-02T15:28:54-07:00 mallika@etudes.org
11751 2007-04-12T10:54:42-07:00 sinou@etudes.org
12001 2007-04-24T17:28:22-04:00 dave.ross@gmail.com
```

DB Browser for SQLite - C:\Users\mirey\OneDrive\Desktop\index.sqlite

File Edit View Tools Help

New Database Open Database Write Changes Revert Changes Open Project Save Project Attach Database Close Database

Database Structure Browse Data Execute SQL

Table: Messages Filter in any column

	id	guid	sent_at	sender_id	subject_id	headers	body
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	<b7eaf76e-3366-4ae3-...	2005-12-09 05:34:30	1	1	BLOB	BLOB
2	2	<bb7ce631-ccf0-40a3-...	2005-12-09 05:58:01	2	2	BLOB	BLOB
3	3	<16703710.1134144397398.javamail....	2005-12-09 16:01:49	3	3	BLOB	BLOB
4	4	<9622225.1134139564303.javamail.t...	2005-12-09 14:43:12	4	4	BLOB	BLOB
5	5	<22585576.1134135343809.javamail....	2005-12-09 13:32:29	5	4	BLOB	BLOB
6	6	<29494755.1134157064976.javamail....	2005-12-09 19:32:31	6	5	BLOB	BLOB
7	7	<20623058.1134135893412.javamail....	2005-12-09 13:42:24	5	4	BLOB	BLOB
8	8	<14721465.1134163692331.javamail....	2005-12-09 21:23:09	6	5	BLOB	BLOB
9	9	<31872472.1134150954615.javamail....	2005-12-09 17:54:17	7	6	BLOB	BLOB
10	10	<1304695.1134163143767.javamail.t...	2005-12-09 21:16:39	7	5	BLOB	BLOB
11	11	<3598962.1134188359277.javamail.t...	2005-12-10 04:16:40	8	7	BLOB	BLOB
12	12	<590808.1134277104504.javamail.to...	2005-12-11 04:55:54	9	8	BLOB	BLOB
13	13	<27775220.1134286694278.javamail....	2005-12-11 07:34:43	10	9	BLOB	BLOB
14	14	<18541911.1134241825684.javamail....	2005-12-10 19:07:29	11	10	BLOB	BLOB
15	15	<20021180.1134321068392.javamail....	2005-12-11 17:08:15	12	9	BLOB	BLOB
16	16	<28266638.1134399879087.javamail....	2005-12-12 15:02:36	13	11	BLOB	BLOB
17	17	<1755944.1134395859482.javamail.t...	2005-12-12 13:55:09	14	12	BLOB	BLOB
18	18	<10733551.1134381175177.javamail....	2005-12-12 09:49:16	15	13	BLOB	BLOB
19	19	<13486694.1134400555513.javamail....	2005-12-12 15:13:38	14	11	BLOB	BLOB
20	20	<25165005.1134413878170.javamail....	2005-12-12 18:14:55	16	14	BLOB	BLOB
21	21	<d3dc4255-5936-4582-9aa1-05b2983...	2005-12-12 16:33:29	2	15	BLOB	BLOB
22	22	<6.0.1.1.2.20051212130514.0277fce...	2005-12-12 19:16:12	17	16	BLOB	BLOB
23	23	<6.0.1.1.2.20051212140958.027c5e1...	2005-12-12 20:20:41	17	17	BLOB	BLOB
24	24	<2931210.1134397540283.javamail.t...	2005-12-12 14:23:12	18	18	BLOB	BLOB
25	25	<23677089.1134422832868.javamail....	2005-12-12 21:24:48	19	19	BLOB	BLOB

1 - 26 of 59720

Go to: 1

3. gbasic.py program to compute basic histogram data

```
Command Prompt
59251 2015-04-22T10:08:08+02:00 bgarcia@entornosdeformacion.com
59501 2015-05-13T10:50:56+00:00 adam.marshall@ox.ac.uk

C:\Users\mirey\OneDrive\Desktop>gbasic.py
How many to dump? 20
Loaded messages= 59720 subjects= 29472 senders= 1800

Top 20 Email list participants
steve.swinsborg@swinsborg.com 3337
azeckoski@unicon.net 1907
ian@cam.ac.uk 1591
csev@umich.edu 1490
david.horwitz@uct.ac.za 1221
matthew@longsight.com 1209
neal.caidin@apereo.org 1035
stephen.marquard@uct.ac.za 931
adam.marshall@ox.ac.uk 929
arwhyte@umich.edu 868
adrian.r.fish@gmail.com 764
slt@columbia.edu 732
jimeng@umich.edu 709
hedrick@rutgers.edu 672
clay.fenlason@gatech.edu 670
markjnorton@earthlink.net 669
ottenhoff@longsight.com 653
chmaurer@indiana.edu 619
bkirsch@umich.edu 599
swgithen@mtu.edu 585

Top 20 Email list organizations
umich.edu 6805
gmail.com 5683
swinsborg.com 3337
cam.ac.uk 2626
uct.ac.za 2582
indiana.edu 2333
longsight.com 2315
unicon.net 2310
ox.ac.uk 1544
berkeley.edu 1388
stanford.edu 1355
ucdavis.edu 1184
rsmart.com 1109
apereo.org 1038
rutgers.edu 908
etudes.org 895
gatech.edu 865
virginia.edu 765
columbia.edu 746
earthlink.net 670

C:\Users\mirey\OneDrive\Desktop>
```