

Corpus of git commit related data for OpenStack Nova

Sources:

- Launchpad (<https://launchpad.net/nova> (<https://launchpad.net/nova>))
- Gerrit (<https://review.openstack.org/#/q/status:open,n,z> (<https://review.openstack.org/#/q/status:open,n,z>))
- Git (<https://github.com/openstack/nova> (<https://github.com/openstack/nova>))

Structure of Corpus:

- combined_commits.jsonz : Basic record for each commit, integrating data from Launchpad, Gerrit and Git
- all_blame.jsonz : Applies GIT BLAME to the parent commit of every bug fix related commit. Only includes those lines included within the window for diff. Proximity is distance of line within blame for a changed line in a bug fix commit, where proximity=1 for changed lines or lines adjacent to inserts

Note: All .jsonz files are gzipped json, with individual json entries for each record. Apply json.loads() to each individual line.

Examples:

- Sample entry from combined_commits.jsonz
- Sample entry from all_blame.jsonz
- Example of blame normalization

Code to load corpus

```
In [1]: import gzip
import json
import pprint as pp
```

```
In [2]: def jload_helper(f):
        #restore the object
        result = []
        dict_result = {}
        for line in f:
            v = json.loads(line)
            if 'json_key' in v:
                #individual entries are items in a dict:
                k = v['json_key']
                del v['json_key']
                dict_result[k] = v
                pass
            else:
                result.append(v)
        if dict_result:
            print 'returning dict'
            return dict_result
        elif len(result) > 1:
            print 'returning list'
            return result
        elif len(result) == 1:
            print 'returning singleton'
            return result[0]
        else:
            return False
```

```
In [3]: def jload(name):
        if name.endswith('.z'):
            print 'gzipped'
            with gzip.open(name, 'rb') as f:
                return jload_helper(f)
        else:
            with open(name, 'rb') as f:
                return jload_helper(f)
```

Sample entry from combined_commits.jsonz

```
In [4]: combined_commits = jload('combined_commits.jsonz')
```

```
gzipped  
returning dict
```

Basic commit entry (without corresponding Launchpad info)

```
In [5]: pp.pprint(combined_commits.items()[0])
```

```
(u'0b698186b56af6580633dedd7916df2897945f29',  
{u'author': u'<git.Actor "MORITA Kazutaka <morita.kazutaka@gmail.com>">',  
  u'cid': u'0b698186b56af6580633dedd7916df2897945f29',  
  u'committer': u'<git.Actor "MORITA Kazutaka <morita.kazutaka@gmail.com>">',  
  u'date': 1305808274,  
  u'files': [],  
  u'msg': u'Avoid wildcard import.',  
  u'parents': [u'5a3bd10a2ada4ab1287b115051578b7b27bba6bc']})
```

Entry associated with bug fix. 'g:' prefix is data from gerrit, 'lp:' prefix is data from launchpad. All other fields from git commit

```
In [6]: bug_fix_commits = [k for k,v in combined_commits.items() if 'lp:id' in v]  
print 'commits associated with big fixes:', len(bug_fix_commits)  
print bug_fix_commits[0]  
print  
pp.pprint (combined_commits[bug_fix_commits[0]] )
```

```
commits associated with big fixes: 4456  
3a915546507e9a5617ada90546342a9239735ff8
```

```
{u'author': u'<git.Actor "Dan Prince <dprince@redhat.com>">',  
  u'bug': u'1307582',  
  u'change_id': u'I6855b03f00492fdf51e18c0dcb0a73590a71b374',  
  u'cid': u'3a915546507e9a5617ada90546342a9239735ff8',  
  u'committer': u'<git.Actor "Dan Prince <dprince@redhat.com>">',  
  u'date': 1397491076,  
  u'files': [],  
  u'g:_number': 87305,  
  u'g:branch': u'master',
```

```
u'g:change_id': u'I6855b03f00492fdf51e18c0dcb0a73590a71b374',
u'g:created': u'2014-04-14 15:58:36.000000000',
u'g:labels': {u'Code-Review': {u'all': [{u'_account_id': 6873,
                                         u'name': u'Matt Riedemann',
                                         u'value': 2},
                                         {u'_account_id': 4393,
                                          u'name': u'Dan Smith',
                                          u'value': 2}],
                                u'approved': {u'_account_id': 6873,
                                              u'name': u'Matt Riedemann'}}},
u'Verified': {u'all': [{u'_account_id': 3,
                        u'name': u'Jenkins',
                        u'value': 2},
                       {u'_account_id': 9578,
                        u'name': u'turbo-hipster',
                        u'value': 1}],
              u'approved': {u'_account_id': 3,
                            u'name': u'Jenkins'}}},
u'Workflow': {u'all': [{u'_account_id': 4393,
                        u'name': u'Dan Smith',
                        u'value': 1}],
              u'approved': {u'_account_id': 4393,
                            u'name': u'Dan Smith'}}},
u'g:messages': [{u'_revision_number': 1,
                  u'author': {u'_account_id': 3,
                              u'avatars': [],
                              u'name': u'Jenkins',
                              u'username': u'jenkins'},
                  u'date': u'2014-04-14 15:58:40.000000000',
                  u'id': u'AAAW3//bwM=',
                  u'message': u'Patch Set 1:\n\nStarting check jobs.\nhttp://status.openstack.org/zuul/'},
                 {u'_revision_number': 1,
                  u'author': {u'_account_id': 6873,
                              u'avatars': [],
                              u'email': u'mriedem@us.ibm.com',
                              u'name': u'Matt Riedemann',
                              u'username': u'mriedem'},
                  u'date': u'2014-04-14 16:02:15.000000000',
                  u'id': u'AAAW3//btY=',
                  u'message': u'Patch Set 1: Looks good to me (core reviewer)\n\n'},
                 {u'_revision_number': 1,
                  u'author': {u'_account_id': 4393,
```

```

        u'avatars': [],
        u'email': u'dms@danplanet.com',
        u'name': u'Dan Smith',
        u'username': u'danms'},
    u'date': u'2014-04-14 16:02:39.000000000',
    u'id': u'AAAW3//btE=',
    u'message': u'Patch Set 1: Looks good to me (core reviewer); Approved\n\n'},
    {u'_revision_number': 1,
     u'author': {u'_account_id': 9578,
                  u'avatars': [],
                  u'email': u'turbo-hipster@lists.rcbops.com',
                  u'name': u'turbo-hipster',
                  u'username': u'turbo-hipster'},
     u'date': u'2014-04-14 17:00:35.000000000',
     u'id': u'AAAW3//bN4=',
     u'message': u'Patch Set 1: Works for me\n\nDatabase migration testing successful.\n\n-
gate-real-db-upgrade_nova_mysql_devstack_131007 http://www.rcbops.com/turbo_hipster/results/87/87305/1/ch
eck/gate-real-db-upgrade_nova_mysql_devstack_131007/b27c818/131007_devstack_export.log : SUCCESS in 10m 4
8s\n- gate-real-db-upgrade_nova_mysql_user_001 http://www.rcbops.com/turbo_hipster/results/87/87305/1/che
ck/gate-real-db-upgrade_nova_mysql_user_001/cd3e4b6/user_001.log : SUCCESS in 18m 44s\n- gate-real-db-upg
rade_nova_percona_devstack_131007 http://www.rcbops.com/turbo_hipster/results/87/87305/1/check/gate-real-
db-upgrade_nova_percona_devstack_131007/994969a/131007_devstack_export.log : SUCCESS in 10m 12s\n- gate-r
eal-db-upgrade_nova_percona_user_001 http://www.rcbops.com/turbo_hipster/results/87/87305/1/check/gate-re
al-db-upgrade_nova_percona_user_001/2bc8d1f/user_001.log : SUCCESS in 21m 58s\n- gate-real-db-upgrade_nov
a_mysql_user_002 http://www.rcbops.com/turbo_hipster/results/87/87305/1/check/gate-real-db-upgrade_nova_m
ysql_user_002/9d54f01/user_002.log : SUCCESS in 13m 58s\n- gate-real-db-upgrade_nova_percona_user_002 htt
p://www.rcbops.com/turbo_hipster/results/87/87305/1/check/gate-real-db-upgrade_nova_percona_user_002/77cd
fe9/user_002.log : SUCCESS in 13m 48s\n'},
    {u'_revision_number': 1,
     u'author': {u'_account_id': 5170,
                  u'avatars': [],
                  u'email': u'hyper-v_ci@microsoft.com',
                  u'name': u'Hyper-V CI',
                  u'username': u'hyper-v-ci'},
     u'date': u'2014-04-14 17:09:17.000000000',
     u'id': u'AAAW3//bIE=',
     u'message': u'Patch Set 1:\n\nBuild succeeded.\n\n- http://64.119.130.115/87305/1/ : Lo
oks good to me in 1h 10m 21s\n'},
    {u'_revision_number': 1,
     u'author': {u'_account_id': 3,
                  u'avatars': [],
                  u'name': u'Jenkins',

```

```
        u'username': u'jenkins'},
        u'date': u'2014-04-14 17:11:26.000000000',
        u'id': u'AAAAW3//bHQ=',
        u'message': u'Patch Set 1: Works for me\n\nBuild succeeded.\n\n- gate-nova-pep8 http://
logs.openstack.org/05/87305/1/check/gate-nova-pep8/f9330cd : SUCCESS in 9m 06s\n- gate-nova-docs http://d
ocs-draft.openstack.org/05/87305/1/check/gate-nova-docs/dc91f91/doc/build/html/ : SUCCESS in 8m 34s\n- ga
te-nova-python26 http://logs.openstack.org/05/87305/1/check/gate-nova-python26/9b65d13 : SUCCESS in 29m 0
2s\n- gate-nova-python27 http://logs.openstack.org/05/87305/1/check/gate-nova-python27/b656a8f : SUCCESS
in 22m 12s\n- check-tempest-dsvm-full http://logs.openstack.org/05/87305/1/check/check-tempest-dsvm-full/
49fclae : SUCCESS in 1h 02m 42s\n- check-tempest-dsvm-neutron http://logs.openstack.org/05/87305/1/check/
check-tempest-dsvm-neutron/51342f6 : SUCCESS in 49m 07s\n- check-tempest-dsvm-neutron-heat-slow http://lo
gs.openstack.org/05/87305/1/check/check-tempest-dsvm-neutron-heat-slow/425b1f1 : SUCCESS in 22m 53s\n- ch
eck-grenade-dsvm http://logs.openstack.org/05/87305/1/check/check-grenade-dsvm/3c85e98 : SUCCESS in 43m 5
9s\n- check-grenade-dsvm-neutron http://logs.openstack.org/05/87305/1/check/check-grenade-dsvm-neutron/aa
de7e1 : FAILURE in 24m 07s (non-voting)\n- check-devstack-dsvm-cells http://logs.openstack.org/05/87305/1
/check/check-devstack-dsvm-cells/399655d : SUCCESS in 12m 43s\n- check-tempest-dsvm-postgres-full http://
logs.openstack.org/05/87305/1/check/check-tempest-dsvm-postgres-full/ec7cca3 : SUCCESS in 45m 14s\n- gate
-tempest-dsvm-large-ops http://logs.openstack.org/05/87305/1/check/gate-tempest-dsvm-large-ops/clc822a :
SUCCESS in 14m 31s\n- gate-tempest-dsvm-neutron-large-ops http://logs.openstack.org/05/87305/1/check/gate
-tempest-dsvm-neutron-large-ops/9fdb525 : SUCCESS in 17m 27s\n- check-tempest-dsvm-cells-full http://logs
.openstack.org/05/87305/1/check/check-tempest-dsvm-cells-full/21dcdeb : FAILURE in 25m 07s (non-voting)\n
- gate-nova-pylint http://logs.openstack.org/05/87305/1/check/gate-nova-pylint/d438009 : SUCCESS in 17m 2
6s (non-voting)\n- check-grenade-dsvm-partial-ncpu http://logs.openstack.org/05/87305/1/check/check-grena
de-dsvm-partial-ncpu/914ca6d : SUCCESS in 40m 32s\n'},
        {u'_revision_number': 1,
          u'author': {u'_account_id': 3,
                      u'avatars': [],
                      u'name': u'Jenkins',
                      u'username': u'jenkins'},
          u'date': u'2014-04-14 17:11:29.000000000',
          u'id': u'AAAAW3//bHM=',
          u'message': u'Patch Set 1: No score\n\nStarting gate jobs.\nhttp://status.openstack.org
/zuul/'},
        {u'_revision_number': 1,
          u'author': {u'_account_id': 3,
                      u'avatars': [],
                      u'name': u'Jenkins',
                      u'username': u'jenkins'},
          u'date': u'2014-04-14 18:22:01.000000000',
          u'id': u'AAAAW3//ajo=',
          u'message': u'Patch Set 1: Verified\n\nBuild succeeded.\n\n- gate-nova-docs http://docs
-draft.openstack.org/05/87305/1/gate/gate-nova-docs/ccf6488/doc/build/html/ : SUCCESS in 9m 28s\n- gate-n
ova-pep8 http://logs.openstack.org/05/87305/1/gate/gate-nova-pep8/6469dc3 : SUCCESS in 9m 56s\n- gate-nov
```

a-python26 http://logs.openstack.org/05/87305/1/gate/gate-nova-python26/78ef19f : SUCCESS in 16m 47s\n- gate-nova-python27 http://logs.openstack.org/05/87305/1/gate/gate-nova-python27/d24c47c : SUCCESS in 21m 54s\n- gate-tempest-dsvm-full http://logs.openstack.org/05/87305/1/gate/gate-tempest-dsvm-full/4fd4dc8 : SUCCESS in 54m 25s\n- gate-tempest-dsvm-neutron http://logs.openstack.org/05/87305/1/gate/gate-tempest-dsvm-neutron/9fbb8f9 : SUCCESS in 50m 00s\n- gate-tempest-dsvm-neutron-heat-slow http://logs.openstack.org/05/87305/1/gate/gate-tempest-dsvm-neutron-heat-slow/ad16dfd : SUCCESS in 22m 35s\n- gate-grenade-dsvm http://logs.openstack.org/05/87305/1/gate/gate-grenade-dsvm/3a40977 : SUCCESS in 39m 53s\n- gate-grenade-dsvm-neutron http://logs.openstack.org/05/87305/1/gate/gate-grenade-dsvm-neutron/b7d56ac : FAILURE in 24m 16s (non-voting)\n- gate-devstack-dsvm-cells http://logs.openstack.org/05/87305/1/gate/gate-devstack-dsvm-cells/d513a0f : SUCCESS in 11m 22s\n- gate-tempest-dsvm-postgres-full http://logs.openstack.org/05/87305/1/gate/gate-tempest-dsvm-postgres-full/767982f : SUCCESS in 1h 06m 51s\n- gate-tempest-dsvm-large-ops http://logs.openstack.org/05/87305/1/gate/gate-tempest-dsvm-large-ops/9b90688 : SUCCESS in 19m 01s\n- gate-tempest-dsvm-neutron-large-ops http://logs.openstack.org/05/87305/1/gate/gate-tempest-dsvm-neutron-large-ops/bcf9923 : SUCCESS in 16m 25s\n- gate-grenade-dsvm-partial-ncpu http://logs.openstack.org/05/87305/1/gate/gate-grenade-dsvm-partial-ncpu/3c03c7a : SUCCESS in 54m 19s\n'},

```
{u'_revision_number': 1,
 u'author': {u'_account_id': 3,
             u'avatars': [],
             u'name': u'Jenkins',
             u'username': u'jenkins'},
 u'date': u'2014-04-14 18:22:05.000000000',
 u'id': u'AAAW3//ajk=',
 u'message': u'Change has been successfully merged into the git repository.'},
```

```
{u'_revision_number': 1,
 u'author': {u'_account_id': 10118,
             u'avatars': [],
             u'email': u'kvmpower@linux.vnet.ibm.com',
             u'name': u'IBM PowerKVM Testing',
             u'username': u'powerkvm'},
 u'date': u'2014-04-14 18:34:53.000000000',
 u'id': u'AAAW3//afA=',
 u'message': u'Patch Set 1:\n\nBuild failed. For rechecking only on the IBM PowerKVM CI,
```

add a review comment with recheck-pkvm\n\n- gate-ibm-tempest-dsvm-full http://dal05.objectstorage.softlayer.net/v1/AUTH_3d8e6ecb-f597-448c-8ec2-164e9f710dd6/pkvmci/87305/1/gate-ibm-tempest-dsvm-full/34dd253 : SUCCESS in 31m 12s\n- gate-ibm-tempest-dsvm-postgres-full http://dal05.objectstorage.softlayer.net/v1/AUTH_3d8e6ecb-f597-448c-8ec2-164e9f710dd6/pkvmci/87305/1/gate-ibm-tempest-dsvm-postgres-full/2dbf010 : FAILURE in 32m 45s\n'},

```
{u'_revision_number': 1,
 u'author': {u'_account_id': 10385,
             u'avatars': [],
             u'email': u'openstack@citrix.com',
             u'name': u'XenServer CI',
```

```

        u'username': u'citrix_xenserver_ci'},
    u'date': u'2014-04-15 11:00:34.000000000',
    u'id': u'AAAAW3//U80=',
    u'message': u'Patch Set 1:\n\nPassed using XenAPI driver with XenServer 6.2: Logs at ht
tp://dd6b71949550285df7dc-dda4e480e005aaa13ec303551d2d8155.r49.cf1.rackcdn.com/05/87305/1/results.html\n\
nStandard recheck supported; use "recheck xenserver" to trigger only xenserver re-check. XenServer CI co
ntact: openstack@citrix.com.'}],
    u'g:owner': {u'_account_id': 360, u'name': u'Dan Prince'},
    u'g:project': u'openstack/nova',
    u'g:status': u'MERGED',
    u'g:subject': u'Revert "Fix network-api direct database hits in metadata server"',
    u'g:topic': u'fix_metadata_neutron',
    u'g:updated': u'2014-04-15 11:00:34.000000000',
    u'lp:activity': [],
    u'lp:attachments': [],
    u'lp:change_id': u'I6855b03f00492fdf51e18c0dcb0a73590a71b374',
    u'lp:cves': [],
    u'lp:date_created': u'2014-04-14 15:54:46.662334+00:00',
    u'lp:description': u"We recently landed a patch to make the metadata API's base.py use the FixedIP objec
t for address lookups.\n\nThis causes problems when using the metadata API with neutron because the Nova
FixedIP object class does not yet appear to support neutron.\n\nWhen using the metadata server today (wit
h Neutron) you'll see the following:\n\nApr 14 15:36:40 localhost nova-api[3373]: 2014-04-14 15:36:40.035
3505 ERROR nova.api.metadata.handler [-] Failed to get metadata for ip: 172.19.0.5",
    u'lp:duplicate_of': u'None',
    u'lp:heat': u'6',
    u'lp:id': u'1307582',
    u'lp:information_type': u'Public',
    u'lp:message_count': u'3',
    u'lp:messages': [{u'bug_attachments': [],
        u'content': u"We recently landed a patch to make the metadata API's base.py use the Fi
xedIP object for address lookups.\n\nThis causes problems when using the metadata API with neutron becaus
e the Nova FixedIP object class does not yet appear to support neutron.\n\nWhen using the metadata server
today (with Neutron) you'll see the following:\n\nApr 14 15:36:40 localhost nova-api[3373]: 2014-04-14 1
5:36:40.035 3505 ERROR nova.api.metadata.handler [-] Failed to get metadata for ip: 172.19.0.5",
        u'date_created': u'2014-04-14 15:54:46.662334+00:00',
        u'resource_type': u'#message',
        u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/1307582/comments/0',
        u'subject': u'FixedIp object does not support Neutron'},
    {u'bug_attachments': [],
        u'content': u'Fix proposed to branch: master\nReview: https://review.openstack.org/873
05',
        u'date_created': u'2014-04-14 15:58:46.587410+00:00',
        u'resource_type': u'#message',

```



```

        u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/1307582/comments/1',
        u'subject': u'Fix proposed to nova (master)'},
        {u'bug_attachments': [],
         u'content': u'Reviewed: https://review.openstack.org/87305\nCommitted: https://git.op
enstack.org/cgiit/openstack/nova/commit/?id=3a915546507e9a5617ada90546342a9239735ff8\nSubmitter: Jenkins\n
Branch:      master\n\nncommit 3a915546507e9a5617ada90546342a9239735ff8\nAuthor: Dan Prince <email address h
idden>\nDate:   Mon Apr 14 11:56:17 2014 -0400\n\n      Revert "Fix network-api direct database hits in met
adata server"\n      \n      This reverts commit 79ab96e34ba5b8dd3e4e542dd3a7f65624b13367.\n      \n      This co
mmit broke the Nova metadata API when using neutron.\n      \n      Change-Id: I6855b03f00492fdf51e18c0dcb0a7
3590a71b374\n      Closes-bug: #1307582\n',
         u'date_created': u'2014-04-14 18:22:13.267313+00:00',
         u'resource_type': u'#message',
         u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/1307582/comments/2',
         u'subject': u'Fix merged to nova (master)'}}],
    u'lp:name': u'None',
    u'lp:resource_type': u'#bug',
    u'lp:review': u'87305',
    u'lp:security_related': u'False',
    u'lp:self_link': u'https://api.launchpad.net/1.0/bugs/1307582',
    u'lp:tags': [],
    u'lp:title': u'FixedIp object does not support Neutron',
    u'msg': u'Revert "Fix network-api direct database hits in metadata server"\n\nThis reverts commit 79ab96
e34ba5b8dd3e4e542dd3a7f65624b13367.\n\nThis commit broke the Nova metadata API when using neutron.\n\nCha
nge-Id: I6855b03f00492fdf51e18c0dcb0a73590a71b374\nCloses-bug: #1307582\n',
    u'parents': [u'7a1b8ff31db82c42c6d163b728d7fd6fbe5ae0cd']}

```

Sample entry from all_blame.jsonz

```
In [7]: all_blame = jload('all_blame.jsonz')
```

```

gzipped
returning list

```

```
In [8]: pp.pprint(all_blame[0])
```

```

{u'blame': {u'nova/api/metadata/base.py': [{u'commit': u'3222d8bb021effb3156ed86e263cdc0ac64fd44a',
      u'lineno': 29,
      u'proximity': 4},
      {u'commit': u'91db10a0fd43c30ed088082583b7d43667339ac5',

```

```
u'lineno': 30,  
u'proximity': 3},  
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',  
u'lineno': 31,  
u'proximity': 2},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 32,  
u'proximity': 1},  
{u'commit': u'61e6f5b1e1a0a43b837e7c428a9624af06bbb4cf',  
u'lineno': 33,  
u'proximity': 2},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 34,  
u'proximity': 1},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 35,  
u'proximity': 1},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 36,  
u'proximity': 2},  
{u'commit': u'bd02a9189a7416097c274660d510eb85b8cd27e5',  
u'lineno': 37,  
u'proximity': 3},  
{u'commit': u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a',  
u'lineno': 118,  
u'proximity': 3},  
{u'commit': u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a',  
u'lineno': 119,  
u'proximity': 2},  
{u'commit': u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a',  
u'lineno': 120,  
u'proximity': 1},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 121,  
u'proximity': 1},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 122,  
u'proximity': 2},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 123,  
u'proximity': 1},  
{u'commit': u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a',  
u'lineno': 124,
```

```
u'proximity': 2},
{u'commit': u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a',
 u'lineno': 125,
 u'proximity': 3},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 126,
 u'proximity': 4},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 162,
 u'proximity': 4},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 163,
 u'proximity': 3},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 164,
 u'proximity': 2},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 165,
 u'proximity': 1},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 166,
 u'proximity': 1},
{u'commit': u'b0dalab2310316f55735827aa68ef0167274ce27',
 u'lineno': 167,
 u'proximity': 2},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 168,
 u'proximity': 3},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 169,
 u'proximity': 4},
{u'commit': u'bd02a9189a7416097c274660d510eb85b8cd27e5',
 u'lineno': 507,
 u'proximity': 4},
{u'commit': u'bd02a9189a7416097c274660d510eb85b8cd27e5',
 u'lineno': 508,
 u'proximity': 3},
{u'commit': u'bd02a9189a7416097c274660d510eb85b8cd27e5',
 u'lineno': 509,
 u'proximity': 2},
{u'commit': u'91db10a0fd43c30ed088082583b7d43667339ac5',
 u'lineno': 510,
 u'proximity': 1},
```

```
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',
 u'lineno': 511,
 u'proximity': 2},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 512,
 u'proximity': 3},
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',
 u'lineno': 513,
 u'proximity': 4}],
u'nova/tests/test_metadata.py': [{u'commit': u'2da5ea6e5685ae5c349fc1d175594aecbe0a0e23',
 u'lineno': 27,
 u'proximity': 4},
{u'commit': u'2da5ea6e5685ae5c349fc1d175594aecbe0a0e23',
 u'lineno': 28,
 u'proximity': 3},
{u'commit': u'2da5ea6e5685ae5c349fc1d175594aecbe0a0e23',
 u'lineno': 29,
 u'proximity': 2},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 30,
 u'proximity': 1},
{u'commit': u'706a1370056ffccc2c8811fc1ac0679944564ece',
 u'lineno': 31,
 u'proximity': 2},
{u'commit': u'd1b904d5b7d4a277adc156d8ab576b37b7e190fc',
 u'lineno': 32,
 u'proximity': 3},
{u'commit': u'd1b904d5b7d4a277adc156d8ab576b37b7e190fc',
 u'lineno': 33,
 u'proximity': 4},
{u'commit': u'61e6f5b1e1a0a43b837e7c428a9624af06bbb4cf',
 u'lineno': 47,
 u'proximity': 3},
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',
 u'lineno': 48,
 u'proximity': 2},
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',
 u'lineno': 49,
 u'proximity': 1},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 50,
 u'proximity': 1},
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',
```

```
u'lineno': 51,  
u'proximity': 2},  
{u'commit': u'b0da1ab2310316f55735827aa68ef0167274ce27',  
u'lineno': 52,  
u'proximity': 3},  
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',  
u'lineno': 71,  
u'proximity': 4},  
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',  
u'lineno': 72,  
u'proximity': 3},  
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',  
u'lineno': 73,  
u'proximity': 2},  
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',  
u'lineno': 74,  
u'proximity': 1},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 75,  
u'proximity': 2},  
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',  
u'lineno': 76,  
u'proximity': 3},  
{u'commit': u'b947ee558d168705d0080f55b4aac1ba1352df17',  
u'lineno': 77,  
u'proximity': 4},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 82,  
u'proximity': 4},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 83,  
u'proximity': 3},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 84,  
u'proximity': 2},  
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',  
u'lineno': 85,  
u'proximity': 1},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 86,  
u'proximity': 1},  
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',  
u'lineno': 87,
```

```
u'proximity': 2},
{u'commit': u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97',
 u'lineno': 88,
 u'proximity': 3},
{u'commit': u'265431666ac3c4b65bf74e40c170d0d62eafe242',
 u'lineno': 307,
 u'proximity': 4},
{u'commit': u'64fe85d526c7a8f88193b41edd4067a6d44c0dc0',
 u'lineno': 308,
 u'proximity': 3},
{u'commit': u'64fe85d526c7a8f88193b41edd4067a6d44c0dc0',
 u'lineno': 309,
 u'proximity': 2},
{u'commit': u'64fe85d526c7a8f88193b41edd4067a6d44c0dc0',
 u'lineno': 310,
 u'proximity': 1},
{u'commit': u'64fe85d526c7a8f88193b41edd4067a6d44c0dc0',
 u'lineno': 311,
 u'proximity': 1},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 312,
 u'proximity': 1},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 313,
 u'proximity': 1},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 314,
 u'proximity': 1},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 315,
 u'proximity': 1},
{u'commit': u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367',
 u'lineno': 316,
 u'proximity': 1},
{u'commit': u'b0da1ab2310316f55735827aa68ef0167274ce27',
 u'lineno': 317,
 u'proximity': 1},
{u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
 u'lineno': 318,
 u'proximity': 1},
{u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
 u'lineno': 319,
 u'proximity': 1},
```

```

        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 320,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 321,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 322,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 323,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 324,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 325,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 326,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 327,
         u'proximity': 1},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 328,
         u'proximity': 2},
        {u'commit': u'903df1fff10bfd9f37aa994981bddd53de7fefb',
         u'lineno': 329,
         u'proximity': 3}}],
    u'cid': u'3a915546507e9a5617ada90546342a9239735ff8'}

```

Normalize Blame Entry

Weight commits based on proximity to change. Normalize to that overall weight is 0. Normalization can either be reciprocal of distance or exponential

```
In [9]: from collections import defaultdict
```

```
In [10]: def normalize_blame(blameset, exp_weighting=True):
        """returns list of commits with weighted blame"""
        result = defaultdict(float)
        for per_file in blameset['blame'].values():
            for per_line in per_file:
                #print per_line
                if exp_weighting:
                    result[per_line['commit']] += 1.0/(2.0**((per_line['proximity']-1))
                else:
                    result[per_line['commit']] += 1.0/float(per_line['proximity'])

        total = sum(result.values())
        return dict([[k, v/total] for k, v in result.items()]])
```

```
In [11]: normalize_blame(all_blame[0], exp_weighting=True)
```

```
Out[11]: {u'265431666ac3c4b65bf74e40c170d0d62eafe242': 0.11224489795918367,
u'2da5ea6e5685ae5c349fc1d175594aecbe0a0e23': 0.017857142857142856,
u'3222d8bb021effb3156ed86e263cdc0ac64fd44a': 0.002551020408163265,
u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97': 0.0663265306122449,
u'61e6f5b1e1a0a43b837e7c428a9624af06bbb4cf': 0.015306122448979591,
u'64fe85d526c7a8f88193b41edd4067a6d44c0dc0': 0.05612244897959184,
u'706a1370056ffccc2c8811fc1ac0679944564ece': 0.01020408163265306,
u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367': 0.26785714285714285,
u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a': 0.05102040816326531,
u'903df1fff10bfd9f37aa994981bddd53de7fefb': 0.2193877551020408,
u'91db10a0fd43c30ed088082583b7d43667339ac5': 0.025510204081632654,
u'b0dalab2310316f55735827aa68ef0167274ce27': 0.03571428571428571,
u'b947ee558d168705d0080f55b4aac1ba1352df17': 0.08928571428571429,
u'bd02a9189a7416097c274660d510eb85b8cd27e5': 0.02295918367346939,
u'd1b904d5b7d4a277adc156d8ab576b37b7e190fc': 0.007653061224489796}
```



```
In [12]: normalize_blame(all_blame[0], exp_weighting=False)
```

```
Out[12]: {u'265431666ac3c4b65bf74e40c170d0d62eafe242': 0.11217948717948716,  
u'2da5ea6e5685ae5c349fc1d175594aecbe0a0e23': 0.02083333333333333,  
u'3222d8bb021effb3156ed86e263cdc0ac64fd44a': 0.004807692307692307,  
u'55128d1c5d2ffda769f7b939dfc7dc6af3bf2e97': 0.0705128205128205,  
u'61e6f5b1e1a0a43b837e7c428a9624af06bbb4cf': 0.01602564102564102,  
u'64fe85d526c7a8f88193b41edd4067a6d44c0dc0': 0.054487179487179474,  
u'706a1370056ffccc2c8811fc1ac0679944564ece': 0.009615384615384614,  
u'79ab96e34ba5b8dd3e4e542dd3a7f65624b13367': 0.2580128205128205,  
u'8ad28f924c7ff4c932bd280032cf7c8ffe5aa20a': 0.05128205128205127,  
u'903df1fff10bfd9f37aa994981bddd53de7fefb': 0.20833333333333331,  
u'91db10a0fd43c30ed088082583b7d43667339ac5': 0.025641025641025637,  
u'b0da1ab2310316f55735827aa68ef0167274ce27': 0.03525641025641025,  
u'b947ee558d168705d0080f55b4aac1ba1352df17': 0.09455128205128202,  
u'bd02a9189a7416097c274660d510eb85b8cd27e5': 0.027243589743589737,  
u'd1b904d5b7d4a277adc156d8ab576b37b7e190fc': 0.011217948717948715}
```

```
In [12]:
```