



All Data Big and Small

<http://redmonk.com/public/lucene.pdf>

**In the beginning, there was
the **database**...**

1979

ORACLE®

1983



1989

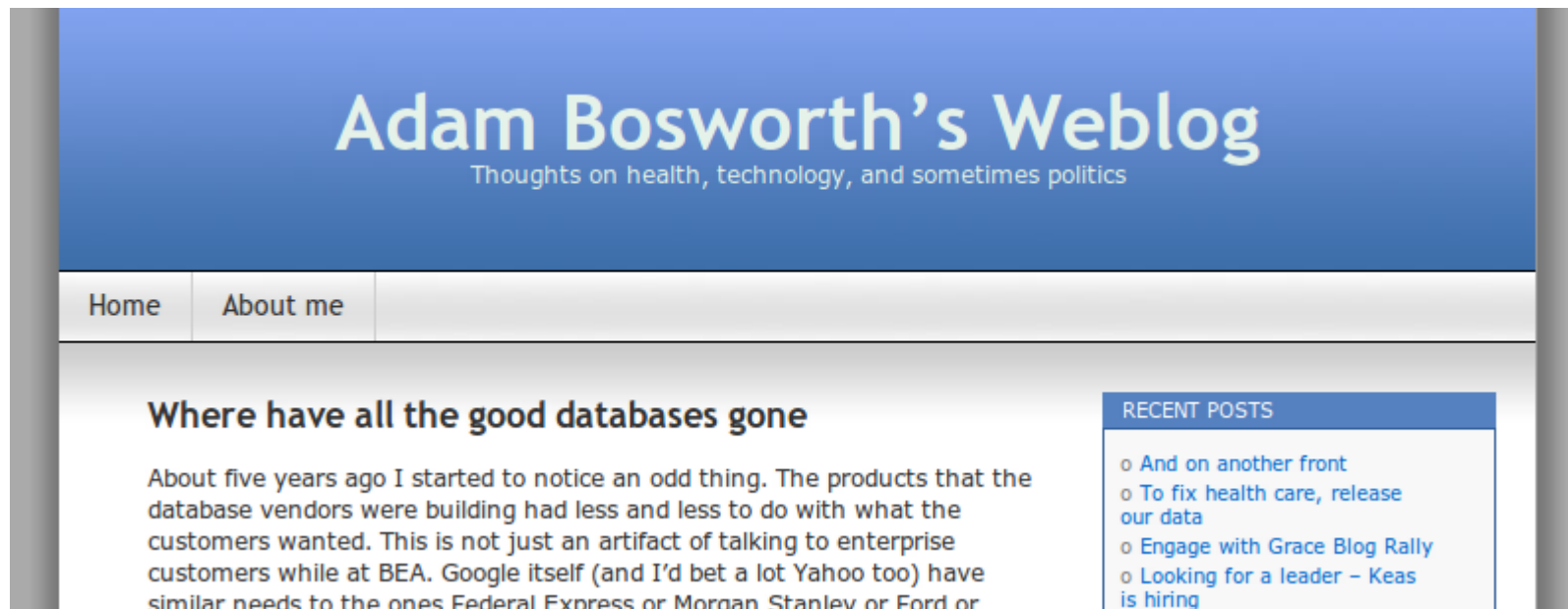
Microsoft®
SQL Server™

**When you have a hammer
and so on**



Source: <http://www.flickr.com/photos/pagedooley/2234031789/>

December 29, 2004



The screenshot shows the homepage of Adam Bosworth's Weblog. The header is a blue gradient bar with the title "Adam Bosworth's Weblog" in white, and a subtitle "Thoughts on health, technology, and sometimes politics" in a smaller white font. Below the header is a navigation bar with two links: "Home" and "About me". The main content area has a light gray background. On the left, there is a post titled "Where have all the good databases gone" with a paragraph of text. On the right, there is a sidebar titled "RECENT POSTS" with a list of four links.

Adam Bosworth's Weblog

Thoughts on health, technology, and sometimes politics

[Home](#) [About me](#)

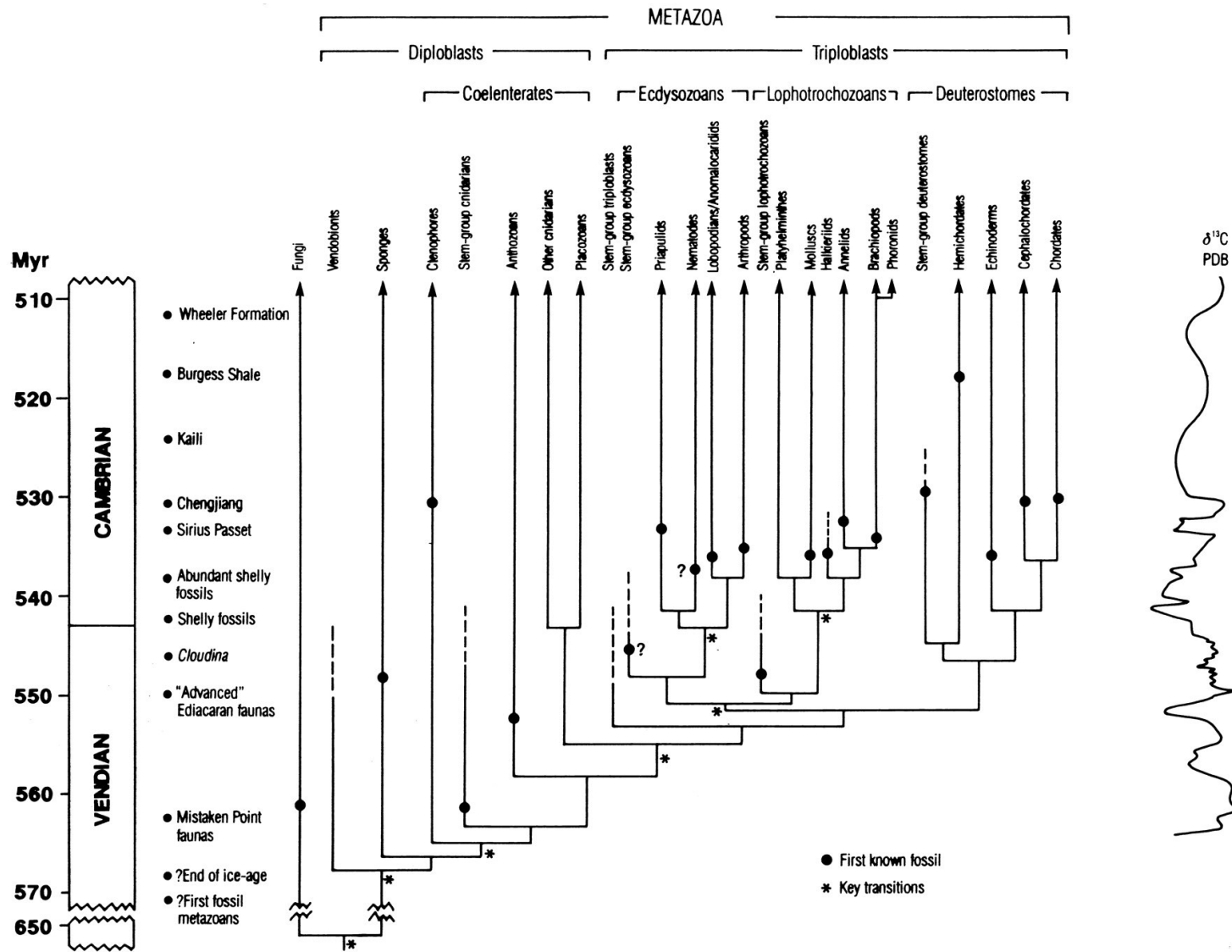
Where have all the good databases gone

About five years ago I started to notice an odd thing. The products that the database vendors were building had less and less to do with what the customers wanted. This is not just an artifact of talking to enterprise customers while at BEA. Google itself (and I'd bet a lot Yahoo too) have similar needs to the ones Federal Express or Morgan Stanley or Ford or

RECENT POSTS

- [And on another front](#)
- [To fix health care, release our data](#)
- [Engage with Grace Blog Rally](#)
- [Looking for a leader - Keas is hiring](#)

The Cambrian Non-relational Explosion



Source: <http://www.pnas.org/content/97/9/4426/F1.expansion.html>



Cassandra

Apache

Solr



Tokyo Cabinet 8192PIB



COUCHBase

riak



Neo4j
the graph database



 **mongoDB**



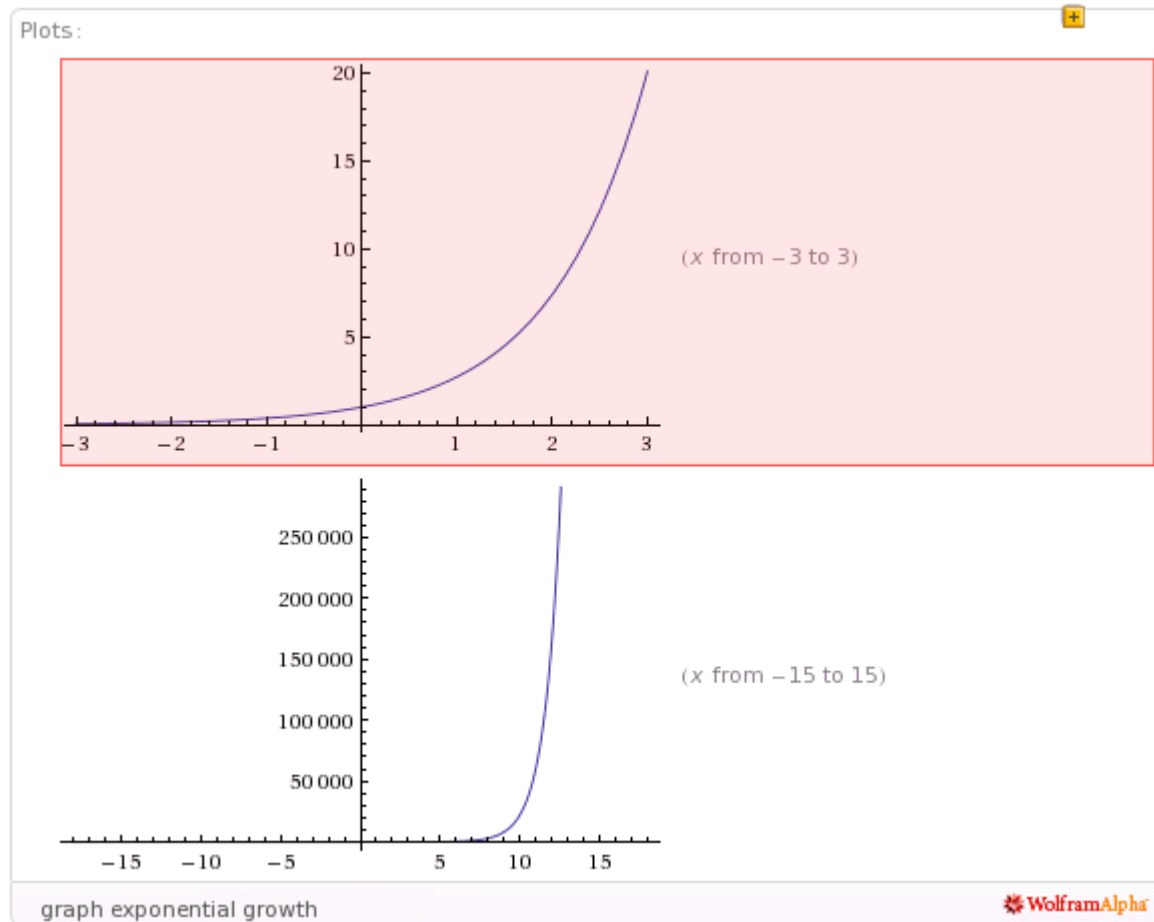
redis

Why?

Different tools for different **jobs**

Or, rather, different data

A **lot** of different data



Most of the attention goes to **Big Data**

**In spite of the fact that
comparatively **few** have it**

Less heralded is **unstructured data**


```
GNU nano 1.3.10      File: access.log

79.171.81.215 - - [22/May/2011:06:26:01 -0400] "GET /robots.txt HTTP/1.1" 200 297 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
79.171.81.215 - - [22/May/2011:06:26:02 -0400] "GET /cote/topic/podcasts/austintech/ HTTP/1.1" 200 50927 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
79.171.81.218 - - [22/May/2011:06:26:04 -0400] "GET /cote/topic/cloud/ HTTP/1.1" 200 53011 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
207.46.12.241 - - [22/May/2011:06:26:04 -0400] "GET /jgovernor/tag/hadoop/ HTTP/1.1" 200 23361 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)"
222.232.53.204 - - [22/May/2011:06:26:05 -0400] "GET /jgovernor/2011/04/19/you-should-watch-this-video-redmonks-developer-philosophy-wrapped-in-humour/#comments HTTP/1.0" 404 20$
69.55.231.75 - - [22/May/2011:06:26:04 -0400] "POST /jgovernor/wp/wp-cron.php?doing_wp_cron HTTP/1.0" 200 - "-" "WordPress/3.1.2; http://www.redmonk.com/jgovernor"
63.119.10.10 - - [22/May/2011:06:26:06 -0400] "GET /sogrady/feed/ HTTP/1.1" 200 97201 "-" "Apple-PubSub/65.21"
220.181.94.233 - - [22/May/2011:06:26:06 -0400] "GET /energy-efficiency-in-the-enterprise-chris-oconnors-pulse-keynote/ HTTP/1.1" 200 86648 "-" "Sogou web spider/4.0(+http://www$
67.195.112.154 - - [22/May/2011:06:26:07 -0400] "GET /tag/president-elect/ HTTP/1.0" 200 56331 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/$
67.195.111.184 - - [22/May/2011:06:26:12 -0400] "GET /cote/2007/02/01/ HTTP/1.0" 200 39323 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slur$
193.200.241.126 - - [22/May/2011:06:26:14 -0400] "GET /feed HTTP/1.1" 304 - "-" "Java/1.6.0_16"
94.13.105.130 - - [22/May/2011:06:26:13 -0400] "GET /sogrady/2008/10/22/drizzle_changes/ HTTP/1.1" 200 69887 "-" "Chilkat/1.0.0 (+http://www.chilkatsoft.com/ChilkatHttpUA.asp)"
65.52.28.252 - - [22/May/2011:06:26:18 -0400] "GET /sogrady/2009/12/15/jim-corgel-interview/trackback/ HTTP/1.1" 302 - "-" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0)"
79.171.81.216 - - [22/May/2011:06:26:17 -0400] "GET /cote/topic/enterprise-software/ HTTP/1.1" 200 54638 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
76.13.114.106 - - [22/May/2011:06:26:20 -0400] "GET /jgovernor/atom.xml HTTP/1.1" 302 353 "-" "BuzzTracker/1.02 +http://www.buzztracker.com"
65.52.28.252 - - [22/May/2011:06:26:19 -0400] "GET /sogrady/2009/12/15/jim-corgel-interview/ HTTP/1.1" 200 78721 "-" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.0)"
79.171.81.217 - - [22/May/2011:06:26:20 -0400] "GET /cote/topic/systems-management/ HTTP/1.1" 200 52507 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
79.171.81.217 - - [22/May/2011:06:26:24 -0400] "GET /cote/topic/rss/ HTTP/1.1" 200 54351 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
209.85.238.209 - - [22/May/2011:06:26:25 -0400] "GET /jgovernor/2011/05/09/on-changing-horses-open-source-licenses-foundations-and-the-software-freedom-conservancy/feed/ HTTP/1.$
79.171.81.216 - - [22/May/2011:06:26:26 -0400] "GET /jgovernor/topic/cloud-computing/ HTTP/1.1" 200 29320 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
50.19.14.128 - - [22/May/2011:06:26:29 -0400] "GET /cote/2008/03/26/osgi-and-prosyst-at-eclipsecon-2008/ HTTP/1.1" 200 47173 "-" "Mozilla/5.0 (compatible; ptd-crawler; +http://b$
66.249.68.40 - - [22/May/2011:06:26:30 -0400] "GET /cote/page/43/?lang=pt_b HTTP/1.1" 200 143126 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
88.15.176.167 - - [22/May/2011:06:26:30 -0400] "POST /wp/wp-admin/admin-ajax.php HTTP/1.1" 200 217 "http://greenmonk.net/wp/wp-admin/post.php?post=3406&action=edit&message=6" "M$
65.52.28.252 - - [22/May/2011:06:26:31 -0400] "GET /sogrady/2009/12/15/jim-corgel-interview/trackback/ HTTP/1.1" 302 - "-" "-"
50.19.14.128 - - [22/May/2011:06:26:30 -0400] "GET /cote/2009/09/17/root-cause-analysis-upcoming-webinar/ HTTP/1.1" 200 47149 "-" "Mozilla/5.0 (compatible; ptd-crawler; +http://$
65.52.28.252 - - [22/May/2011:06:26:33 -0400] "GET /sogrady/2009/12/15/jim-corgel-interview/ HTTP/1.1" 200 78741 "-" "-"
69.191.249.202 - - [22/May/2011:06:26:39 -0400] "GET / HTTP/1.1" 200 147004 "-" "BLP_bbot/0.1"
157.82.156.130 - - [22/May/2011:06:26:40 -0400] "GET /sogrady/2006/07/ HTTP/1.1" 200 55173 "http://redmonk.com/sogrady/2008/04/09/clouds-rolling-in-the-google-app-engine-qa/" "M$
125.163.228.8 - - [22/May/2011:06:26:39 -0400] "GET /greenmonk-news-roundup-11042008/ HTTP/1.1" 200 65743 "-" "Chilkat/1.0.0 (+http://www.chilkatsoft.com/ChilkatHttpUA.asp)"
208.80.194.27 - - [22/May/2011:06:26:46 -0400] "GET /sogrady/2010/01/12/links-for-2010-01-12/tvt\t\t\t\t\t HTTP/1.0" 200 68558 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows N$
69.55.231.75 - - [22/May/2011:06:26:48 -0400] "GET /feed/ HTTP/1.0" 302 - "-" "WordPress/3.1.2; http://redmonk.com/master"
217.113.33.215 - - [22/May/2011:06:26:47 -0400] "GET /cote/ HTTP/1.0" 200 134294 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5 (.N$
217.113.33.215 - - [22/May/2011:06:26:47 -0400] "GET /jgovernor/ HTTP/1.0" 200 98595 "-" "Mozilla/4.0"
69.55.231.75 - - [22/May/2011:06:26:49 -0400] "GET /feed/ HTTP/1.0" 200 163828 "-" "WordPress/3.1.2; http://redmonk.com/master"
209.85.238.246 - - [22/May/2011:06:26:45 -0400] "GET /master/feed HTTP/1.1" 301 - "-" "Feedfetcher-Google; (+http://www.google.com/feedfetcher.html; 1 subscribers; feed-id=84924$
209.85.238.246 - - [22/May/2011:06:26:50 -0400] "GET /master/feed/ HTTP/1.1" 304 - "-" "Feedfetcher-Google; (+http://www.google.com/feedfetcher.html; 1 subscribers; feed-id=8492$
24.128.190.130 - - [22/May/2011:06:26:55 -0400] "GET /favicon.ico HTTP/1.0" 200 1406 "-" "Safari/6533.21.1 CFNetwork/454.11.12 Darwin/10.7.0 (i386) (MacBook5%2C1)"
198.185.25.120 - - [22/May/2011:06:26:58 -0400] "GET /wp/wp-content/plugins/wordpress-popular-posts/style/wpp.css HTTP/1.1" 200 548 "http://greenmonk.net/more-oversight-needed-f$
216.39.58.18 - - [22/May/2011:06:27:09 -0400] "GET /robots.txt HTTP/1.1" 200 297 "-" "Yahoo Pipes 2.0"
216.39.58.18 - - [22/May/2011:06:27:09 -0400] "GET /sogrady/2011/03/17/framework-traction-on-hacker-news/ HTTP/1.1" 200 66258 "-" "Mozilla/5.0 (compatible; Yahoo Pipes 2.0; +htt$
222.232.53.204 - - [22/May/2011:06:27:15 -0400] "GET /jgovernor/2011/04/18/on-purpose-meaning-and-poetry-at-ibm-impact/ HTTP/1.0" 200 32900 "-" "larbin_2.6.3 larbin2.6.3@unspeci$
66.249.68.40 - - [22/May/2011:06:27:15 -0400] "GET /cote/2007/09/11/links-for-2007-09-11/ HTTP/1.1" 200 45877 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com$
38.101.148.126 - - [22/May/2011:06:27:16 -0400] "GET /jgovernor/2009/12/10/our-friend-don/feed/ HTTP/1.1" 200 1572 "-" "Mozilla/5.0 (compatible; discobot/1.1; +http://discoverye$
79.171.81.218 - - [22/May/2011:06:27:19 -0400] "GET /cote/topic/reviews-2/ HTTP/1.1" 200 48670 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
209.85.238.246 - - [22/May/2011:06:27:20 -0400] "GET /feed/ HTTP/1.1" 200 6409 "-" "Feedfetcher-Google; (+http://www.google.com/feedfetcher.html; 36 subscribers; feed-id=7899190$
67.195.111.184 - - [22/May/2011:06:27:26 -0400] "GET /jgovernor/2005/12/22/on-micromuse-and-ibm/ HTTP/1.0" 200 29474 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yaho$
119.63.196.117 - - [22/May/2011:06:27:27 -0400] "GET /jgovernor/?p=83 HTTP/1.1" 301 - "-" "Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)"

AG Get Help          AO WriteOut      AR Read File        AY Prev Page
AX Exit              AJ Justify      AW Where Is         AV Next Page
                     AL Append      AU UnCut Text
                     AT To Spell
```


**Between the size and (un)structure,
it's amazing anything **gets found****



Source: <http://www.flickr.com/photos/28705377@N04/4142872268/>

It's hard to ask the **right question**

**To make matters worse,
you may only get **one chance****

**The most important answer
is the next question**

Some questions

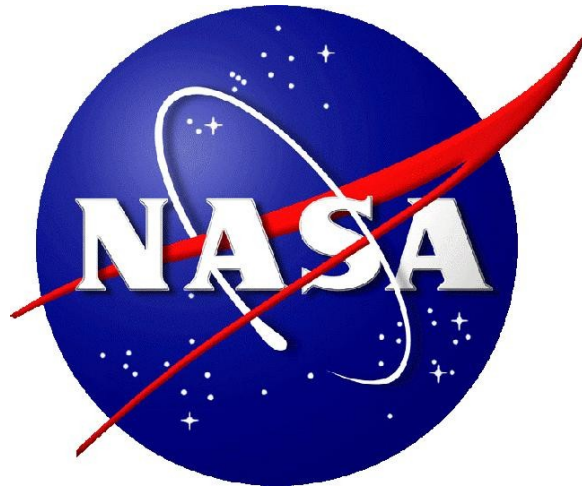


at&t



J. Craig Venter™

I N S T I T U T E



```
sog@apone: ~
GNU nano 2.2.6 File: ...apikey=api_test-W1cipwpcdu9Cbd9pmm8D4Cjc469 Modified
{
  "infochimps.com/soc/net/tw/influence.json?screen_name=infochimps&apikey=api_test-W1cipwpcdu9Cbd9pmm8D4Cjc469"
  {
    "enthusiasm":0.08884297,
    "fo_trstrank":0.9268817,
    "created_at":"20080806103437",
    "trstrank" "interesting":10.167786,
    "resting" "outflux":0.014605695,
    "ex" "influx":533.0094,
    "lux" "followers":"2433.0",
    "rstrank" "feedness":0.7231405,
    "ness" "at_trstrank":0.17347828,
    "id" "user_id":15748351,
    "tiness" "sway":2.9026845,
    " " "chattiness":0.28512397,
    "ow_rate" "follow_rate":0.08248787,
    "on_name" "screen_name":"infochimps",
    "w_churn" "follow_churn":0.8888889
  }
}

^G Get Help ^O WriteOut ^R Read File ^Y Prev Page ^K Cut Text ^C Cur Pos
^X Exit ^J Justify ^W Where Is ^V Next Page ^U UnCut Text ^T To Spell
```

OTHER QUESTIONS