

# RESTful Interface to Healthcare Twitter Analysis MongoDB database

[Introduction](#)[Interface](#)[Choropleth](#)[Instructions](#)

## Healthcare Twitter Analysis

The objective of the Healthcare Twitter Analysis (HTA) project is to find ways to add value to all sectors of the worldwide medical community by utilizing data science to analyze very large volumes of Twitter messages. Starting with a collection of csv files containing health-related tweets from Jan-Jun 2014, the project recruited data-science expertise via crowd-sourcing to help push the analysis forward and at one point had several hundred participants.

This website and all the technology & data that powers it was created by George Fisher\*, one of the participants.

## Technology

This website is constructed as a single-page web app which means that after the initial download there are no more page loads; all requests to the server for data are done in the background with none of the usual flickering and stuttering of a Web 2.0 website.

The "stack" for this website is the following:

- Front End
  - HTML5 & CSS3
  - JavaScript, jQuery & jQuery UI
  - D3.js
- HTTP Server & REST interface
  - Bottle
  - Pymongo
  - Python 2.7
- Database
  - MongoDB

The [status report](#) on my [repo for this project](#) contains a detailed record of the work that went into creating the ~5 million documents in the current database as well as the code that was used to create those records; plus all the code for this website.

\* [GitHub](#), [LinkedIn](#)



# RESTful Interface to Healthcare Twitter Analysis MongoDB database

Introduction Interface Choropleth Instructions

MongoDB Command-Line Query

{ \$text: { \$search: '#Sepsis' } }

Enter the maximum to return

5

Click to send the query

Display the first

Display all

The query: { '\$text': { '\$search': '#Sepsis' } }

Limited to: 5

Number returned: 5

```
{
  "id_list": [
    {
      "_id": {
        "$oid": "5448e5bcc7b27e52998fb32d"
      }
    },
    {
      "_id": {
        "$oid": "5448e5bcc7b27e52998fb328"
      }
    },
    {
      "_id": {
        "$oid": "5448e5bcc7b27e52998fb31e"
      }
    },
    {
      "_id": {
        "$oid": "5448e5bcc7b27e52998fb326"
      }
    },
    {
      "_id": {
        "$oid": "5448e5bcc7b27e52998fb32b"
      }
    }
  ],
  "num": 5,
  "example": {
    "contributors": null,
    "truncated": false,
    "text": "RT @SepsisSurvivors: @JoyCluff And that is the very best attitude to have. #contentment #mrssa #sepsis",
    "in_reply_to_status_id": null,
    "id": 434879289935138800,
    "favorite_count": 0,
    "topsy": {
      "file_counter": 9,
      "firstpost_date": "02/16/14",
      "url": "http://twitter.com/joycluff/status/434879289935138819",
      "timestamp": 1392508800,
      "score": 7.232268,
      "trackback_author_nick": "joycluff",
      "trackback_author_url": "http://twitter.com/joycluff",
      "trackback_permalink": "http://twitter.com/JoyCluff/status/434879289935138819",
      "short_file_name": "/Jan to May/Blood/Tweets_Sepsis.csv"
    },
    "source": "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter for iPhone</a>",
    "retweeted": false,
    "coordinates": null,
    "entities": {
      "symbols": [],
      "user_mentions": [
        {
          "indices": [
            3,
            19
          ],
          "id": 1122623600,
          "id_str": "1122623600",
          "name": "SurvivorsOfSepsis",
          "screen_name": "SepsisSurvivors"
        }
      ],
      "indices": [
        21,
        30
      ],
      "id": 593274264,
      "id_str": "593274264",
      "name": "MOM vs. MRSA",
      "screen_name": "JoyCluff"
    }
  },
  "hashtags": [
    {
      "indices": [
        75,
        87
      ],
      "text": "contentment"
    },
    {
      "indices": [
        88,
        93
      ],
      "text": "mrssa"
    },
    {
      "indices": [
        94,
        101
      ],
      "text": "sepsis"
    }
  ],
  "urls": []
},
{
  "in_reply_to_screen_name": null,
  "id_str": "434879289935138819",
  "retweet_count": 1,
  "in_reply_to_user_id": null,
  "favorited": false,
  "retweeted_status": {
    "contributors": null,
    "truncated": false,
    "text": "@JoyCluff And that is the very best attitude to have. #contentment #mrssa #sepsis",
    "in_reply_to_status_id": 43487753536546300,
    "id": 434879193357099000,
    "favorite_count": 0,
    "source": "<a href='\"http://twitter.com/download/android\"' rel='\"nofollow\"'>Twitter for Android</a>",
    "retweeted": false,
    "coordinates": null,
    "entities": {
      "symbols": [],
      "user_mentions": [
        {
          "indices": [
            0,
            9
          ],
          "id": 593274264,
          "id_str": "593274264",
          "name": "MOM vs. MRSA",
          "screen_name": "JoyCluff"
        }
      ],
      "hashtags": [
        {
          "indices": [
            54,
            66
          ],
          "text": "contentment"
        },
        {
          "indices": [
            67,
            72
          ],
          "text": "mrssa"
        },
        {
          "indices": [
            73,
            80
          ],
          "text": "sepsis"
        }
      ],
      "urls": []
    }
  },
  "in_reply_to_screen_name": "JoyCluff",
  "id_str": "434879193357099008",
  "retweet_count": 1,
  "in_reply_to_user_id": 593274264,
  "favorited": false,
  "user": {
    "follow_request_sent": false,
    "profile_use_background_image": false,
    "id": 1122623600,
    "verified": false,
    "profile_text_color": "333333",
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/486717377115996160/U6kzTo7u_normal.jpeg",
    "profile_sidebar_fill_color": "F6F6F6",
    "is_translator": false,
    "entities": {
      "url": {
        "urls": [
          {
            "url": "http://t.co/hL9oMDWxRP",
            "indices": [
              0,
              22
            ],
            "expanded_url": "http://survivorsofsepsis.blogspot.ca",
            "display_url": "survivorsofsepsis.blogspot.ca"
          }
        ]
      },
      "description": {
        "urls": []
      }
    },
    "followers_count": 1450,
    "protected": false,
    "location": "Alberta, Canada",
    "default_profile_image": false,
    "id_str": "1122623600",
    "lang": "en",
    "utc_offset": null,
    "statuses_count": 21971,
    "description": "Respect Earns Respect ~ None of us are as strong as all of us. I ReTweet all credible #sepsis information; this story is about everyone.",
    "friends_count": 1411,
    "profile_background_image_url_https": "https://pbs.twimg.com/profile_images/460477532173647872/vMfiJWQi.png",
    "profile_link_color": "000000",
    "profile_image_url": "http://pbs.twimg.com/profile_images/486717377115996160/U6kzTo7u_normal.jpeg",
    "notifications": false,
    "geo_enabled": false,
    "profile_background_color": "000000",
    "profile_banner_url": "https://pbs.twimg.com/profile_banners/1122623600/1398309883",
    "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/460477532173647872/vMfiJWQi.png",
    "name": "SurvivorsOfSepsis",
    "is_translation_enabled": false,
    "profile_background_tile": false,
    "favourites_count": 5900,
    "screen_name": "SepsisSurvivors",
    "url": "http://t.co/hL9oMDWxRP",
    "created_at": "Sat Jan 26 17:17:19 +0000 2013",
    "contributors_enabled": false,
    "time_zone": null,
    "profile_sidebar_border_color": "FFFFFF",
    "default_profile": false,
    "following": false,
    "listed_count": 26
  },
  "geo": null,
  "in_reply_to_user_id_str": "593274264",
  "lang": "en",
  "created_at": "Sun Feb 16 02:37:34 +0000 2014",
  "in_reply_to_status_id_str": "43487753536546304",
  "place": null
},
{
  "timestamp": 1392518277,
  "user": {
    "follow_request_sent": false,
    "profile_use_background_image": true,
    "id": 593274264,
    "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme1/bg.png",
    "verified": false,
    "profile_text_color": "333333",
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/490697074027945986/2boM38C8_normal.jpeg",
    "profile_sidebar_fill_color": "D0E0F6",
    "is_translator": false,
    "entities": {
      "url": {
        "urls": [
          {
            "url": "http://t.co/dZPVuT6AVD",
            "indices": [
              0,
              22
            ],
            "expanded_url": "http://www.mrsasurvivors.org",
            "display_url": "mrsasurvivors.org"
          }
        ]
      },
      "description": {
        "urls": []
      }
    },
    "followers_count": 381,
    "protected": false,
    "location": "Saint Joseph, MI",
    "default_profile_image": false,
    "id_str": "593274264",
    "lang": "en",
    "utc_offset": -10800,
    "statuses_count": 1979,
    "description": "son survived #MRSA, I now advocate #ptsafety prevention,education, awareness & host events. Get to know #MRSA & #HAI #Superbugs!",
    "friends_count": 1120,
    "location_geoinfo": {},
    "profile_link_color": "008484",
    "profile_image_url": "http://pbs.twimg.com/profile_images/490697074027945986/2boM38C8_normal.jpeg",
    "notifications": false,
    "geo_enabled": true,
    "profile_background_color": "000000",
    "profile_banner_url": "https://pbs.twimg.com/profile_banners/593274264/1394944323",
    "profile_background_image_url": "http://abs.twimg.com/images/themes/theme1/bg.png",
    "name": "MOM vs. MRSA",
    "is_translation_enabled": false,
    "profile_background_tile": false,
    "favourites_count": 346,
    "screen_name": "JoyCluff",
    "url": "http://t.co/dZPVuT6AVD",
    "created_at": "Tue May 29 00:21:53 +0000 2012",
    "contributors_enabled": false,
    "time_zone": "Atlantic Time (Canada)",
    "profile_sidebar_border_color": "000000",
    "default_profile": true,
    "following": false,
    "listed_count": 3
  },
  "geo": {
    "type": "Point",
    "coordinates": [
      42.109581,
      -86.479954
    ]
  },
  "in_reply_to_user_id_str": null,
  "lang": "en",
  "created_at": "Sun Feb 16 02:37:57 +0000 2014",
  "in_reply_to_status_id_str": null,
  "geo_reverse": {
    "city": "Shorewood-Tower Hills-Harbert",
    "area_code": "",
    "country": "United States",
    "land_sq_mi": "",
    "zipcode": "",
    "county": "",
    "state": "",
    "fips": "",
    "state_abbr": "",
    "country_code": "US",
    "Pop_2010": "",
    "type": ""
  },
  "place": null,
  "_id": {
    "$oid": "5448e5bcc7b27e52998fb32d"
  }
}
```



# RESTful Interface to Healthcare Twitter Analysis MongoDB database

[Introduction](#) [Interface](#) [Choropleth](#) [Instructions](#)

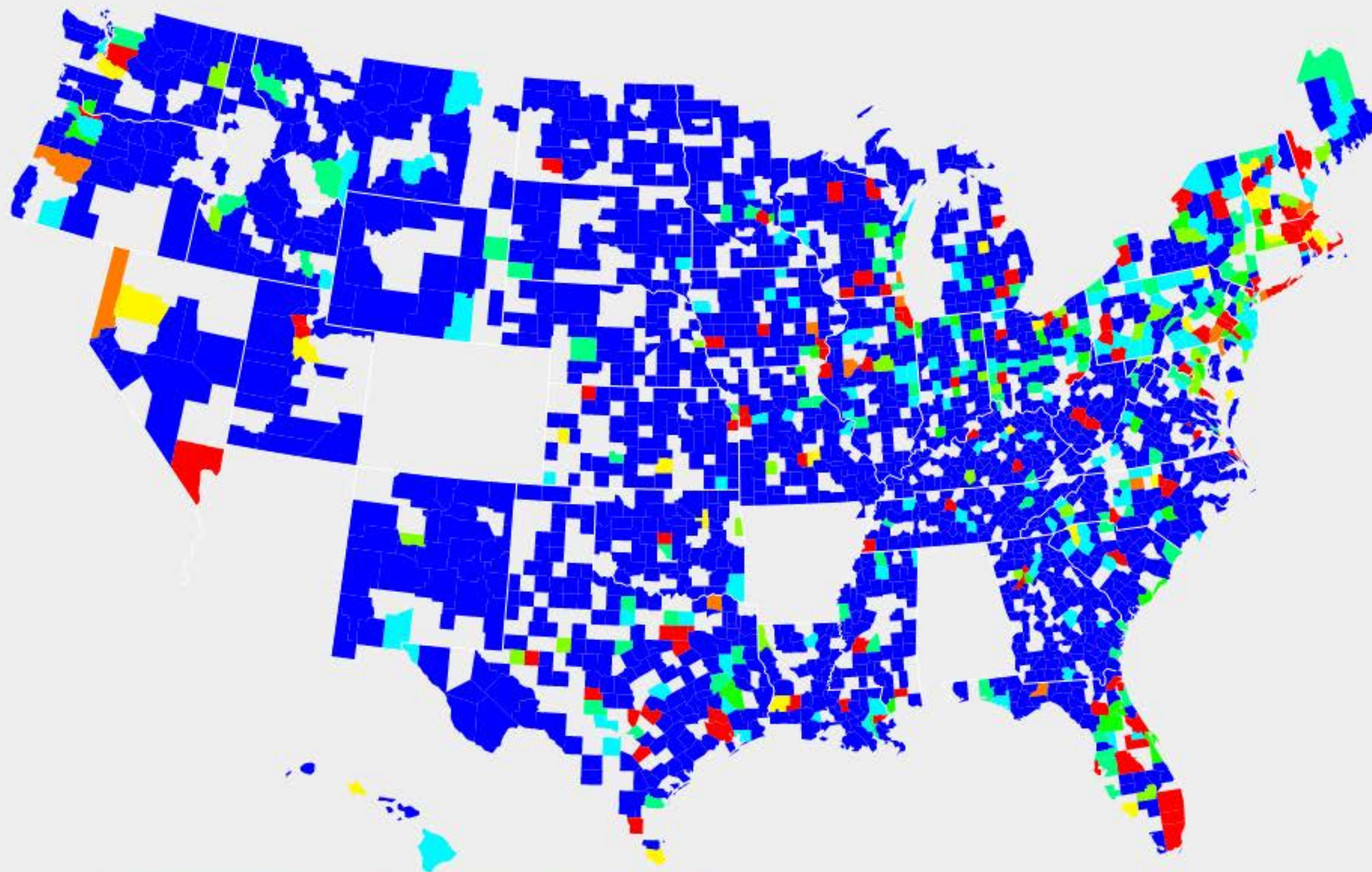
Enter a term to search for

[Click to send the query](#)

Progressbar



Frequency of the term cancer



LOW

frequency

HIGH



# RESTful Interface to Healthcare Twitter Analysis MongoDB database

Introduction Interface **Choropleth** Instructions

These instructions assume the following:

- you have created a MongoDB database as follows  
`mongoimport -d HTA -c grf --file HTA_noduplicates.json`  
on an external hard drive, location `E:\HTA`  
(the status report on the repo ... [HTA repo](#) ... has detailed instructions including suggested indexes to add for performance).
- you have the Bottle Server, `bottle_server.py`  
on the same external hard drive in location `E:\HTA\RESTful Interface`
- you have all the supporting folders and files in the structure shown in the repo.
- you are using the latest version of the Chrome browser; others may work, but only by luck.

The [status report](#) on my [repo for this project](#) will answer most questions.

These instructions apply to a Windows machine; Linux and Mac folks will have to adapt a bit.

1. Start the MongoDB Server using the database on the external drive
  1. Start a command window as Administrator
  2. `net stop MongoDB`
  3. `mongod --dbpath "E:\HTA"`
2. Start the Bottle HTTP Server
  1. Start another command window
  2. `e:`
  3. `cd E:\HTA\RESTful Interface`
  4. `python bottle_server.py`
3. Using Chrome, load this single-page web app  
`http://localhost:8082/`

## Interface Tab

The interface is very simple: there are only two things you have to enter ...

- A MongoDB query in exactly the format used on the mongo command line  
(a placeholder is provided which shows how to perform a search for the hashtag '#sepsis' ... enter nothing and this query will be sent for you).
- A limit on the number of documents to return; set to zero (0) and the '\_id's of all the matching documents will be returned

At the moment you only see displayed the result of the query:

- a list of the '\_id's that match, up to the limit ("id\_list")
- the number in that list ("num")
- the first matching document ("example")

My expectation is to enhance this system so that you can scroll through all the matching documents; the plan is to make ajax requests behind the scenes as you scroll through so that an entire mass of data isn't sent down all at once, storing the documents received so far in Chrome's IndexedDB local storage facility for quick random-access retrieval.

## Choropleth Tab

Simply enter a term such as "cancer" and the REST interface implemented by Bottle will perform a search of the entire database and return the number of instances in each county in the United States, presented as a color gradient on a D3.js SVG map.

Performance is currently an issue ... the MongoDB Aggregation Framework query can be terribly slow the first time it's issued. I have to look into this and perhaps revert to doing the GROUP BY function in Bottle using Python. I've run a number of timing tests and MongoDB beats Python by as much as 30% once it has buffered whatever it's buffering, but getting to that point can take many minutes during which time you can watch the memory usage steadily growing.