

# Stack Overflow Customer Behavior Analysis

Text Mining

Qianchao Liu

# Outline

- Data Description
- Basic Customer Behavior Analysis
- Text Classification
- Sentiment Analysis

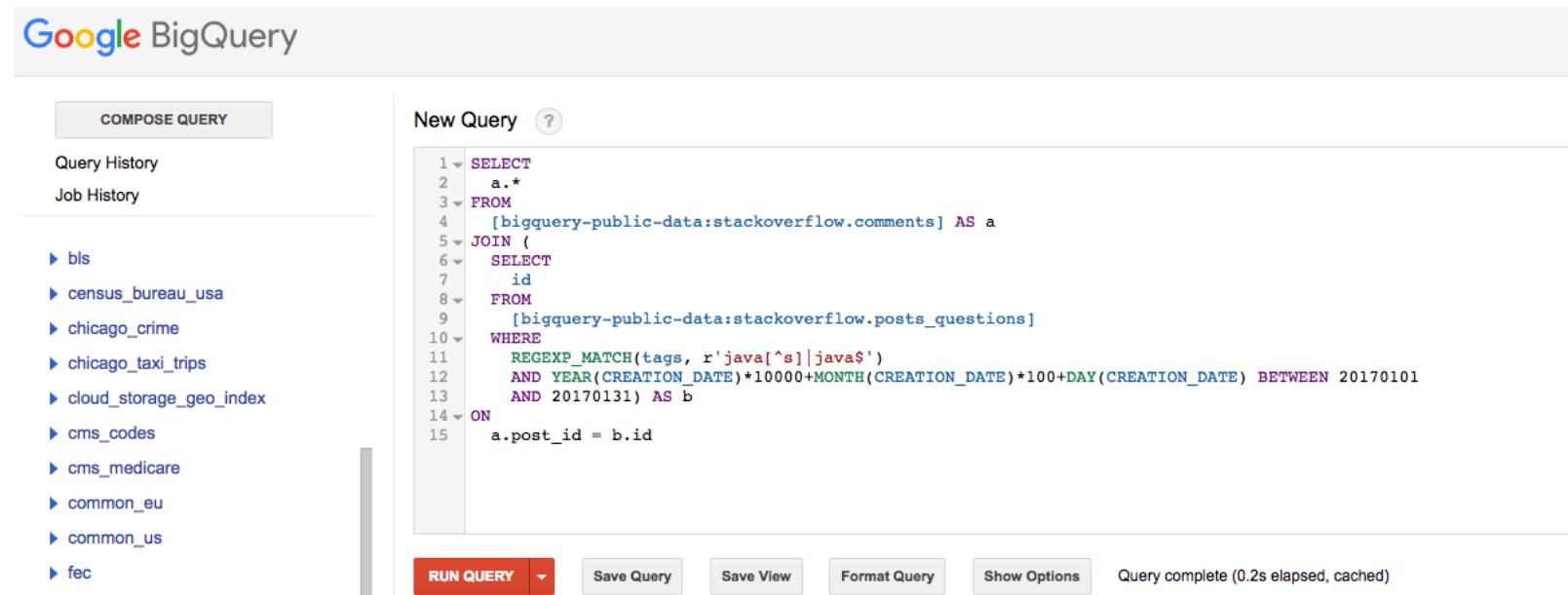
# Part 1 Data Description

# Data Description



- Questions, answers and comments from Stack Overflow
- Time period: January 2017
- Programming Languages: 7 (including Python, Ruby, Cpp ...)

## Data Source:

A screenshot of the Google BigQuery web interface. The top bar shows the "Google BigQuery" logo. On the left, there's a sidebar with a "COMPOSE QUERY" button and a list of datasets including "bls", "census\_bureau\_usa", "chicago\_crime", "chicago\_taxi\_trips", "cloud\_storage\_geo\_index", "cms\_codes", "cms\_medicare", "common\_eu", "common\_us", and "fec". The main area is titled "New Query" and contains a SQL query. The query is a JOIN between "bigquery-public-data:stackoverflow.comments" (aliased as 'a') and "bigquery-public-data:stackoverflow.posts\_questions" (aliased as 'b'). It filters for comments where the tags contain 'java' and the creation date is in January 2017. The query is: 

```
1 SELECT
2   a.*
3 FROM
4   [bigquery-public-data:stackoverflow.comments] AS a
5 JOIN (
6   SELECT
7     id
8   FROM
9     [bigquery-public-data:stackoverflow.posts_questions]
10  WHERE
11    REGEXP_MATCH(tags, r'java[^s]|java$')
12    AND YEAR(CREATION_DATE)*10000+MONTH(CREATION_DATE)*100+DAY(CREATION_DATE) BETWEEN 20170101
13    AND 20170131) AS b
14 ON
15   a.post_id = b.id
```

 At the bottom, there are buttons for "RUN QUERY", "Save Query", "Save View", "Format Query", and "Show Options". A status message at the bottom right says "Query complete (0.2s elapsed, cached)".

# Data Description



▲

1939

▼

Best way to check if a list is empty

For example, if passed the following:

```
a = []
```

How do I check to see if `a` is empty?

Question

★ 323

python list

share edit

edited Apr 18 '14 at 9:42

Martijn Pieters ♦ 550k ● 85 ● 1599 ● 1673

asked Sep 10 '08 at 6:20

Ray Vega 58.7k ● 83 ● 177 ● 181

- 4
- Checking for emptiness might be useful if you are consuming your list in random order. Something like `item = get_item_from_somewhere(); if item in list: {consume_item(item); list.remove(item); if not list: {print('You are done')}}` – Vladimir Cravero Oct 21 '15 at 9:24
- 2
- @VladimirCravero - Of course. However, if the OP doesn't know enough about Python to know about the 'truthiness' of a list, it is not unthinkable to imagine they may not understand the benefit of just processing a list regardless of its contents. – MrWonderful Jun 21 '16 at 17:27

Comment

29 Answers

active oldest votes

▲

2791

▼

```
if not a:
    print("List is empty")
```

Using the implicit booleanness of the empty list is quite pythonic.

share edit

edited Apr 27 at 2:52

answered Sep 10 '08 at 6:28

Patrick 40.8k ● 10 ● 41 ● 60

Answer

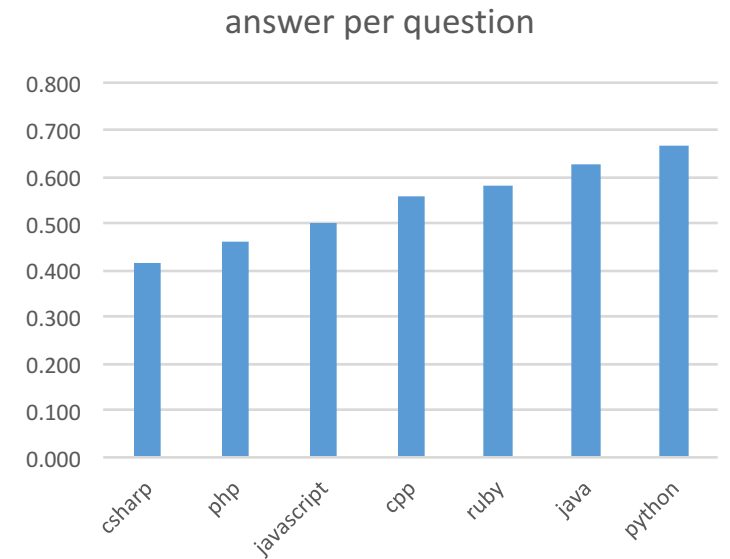
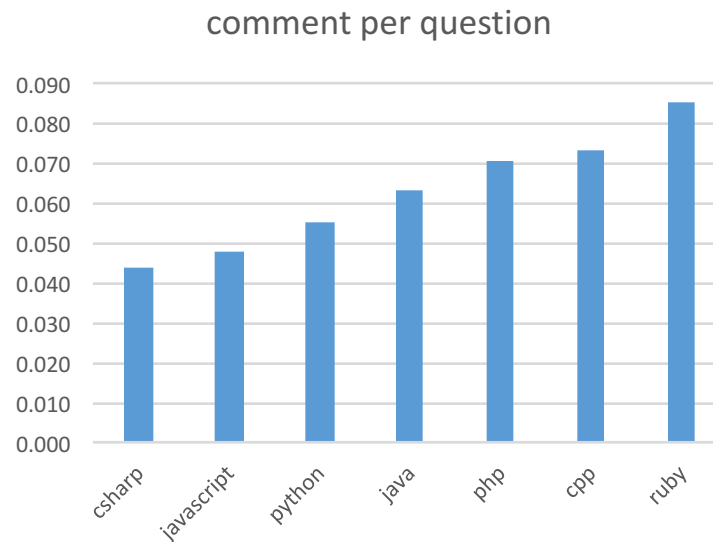
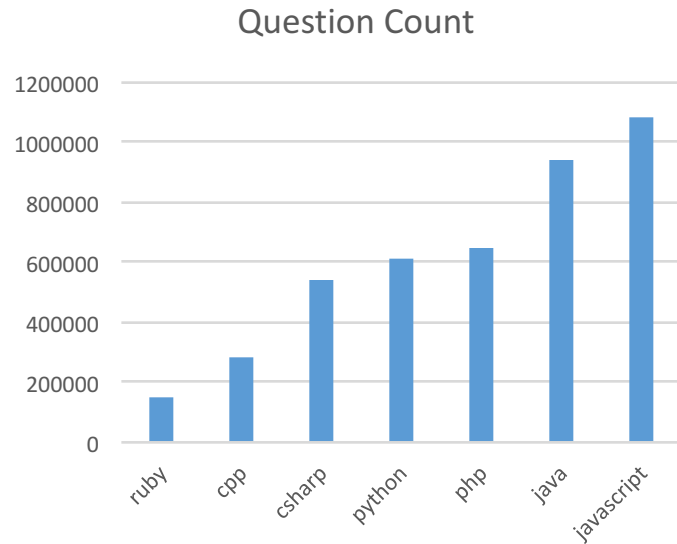
560

Playing devil's advocate. I don't understand why this idiom is considered pythonic. 'Explicit is better than implicit', correct? This check doesn't seem very explicit about what is is checking. – James McMahon Nov 22 '11 at 6:14

## Part 2 Basic Customer Behavior Analysis

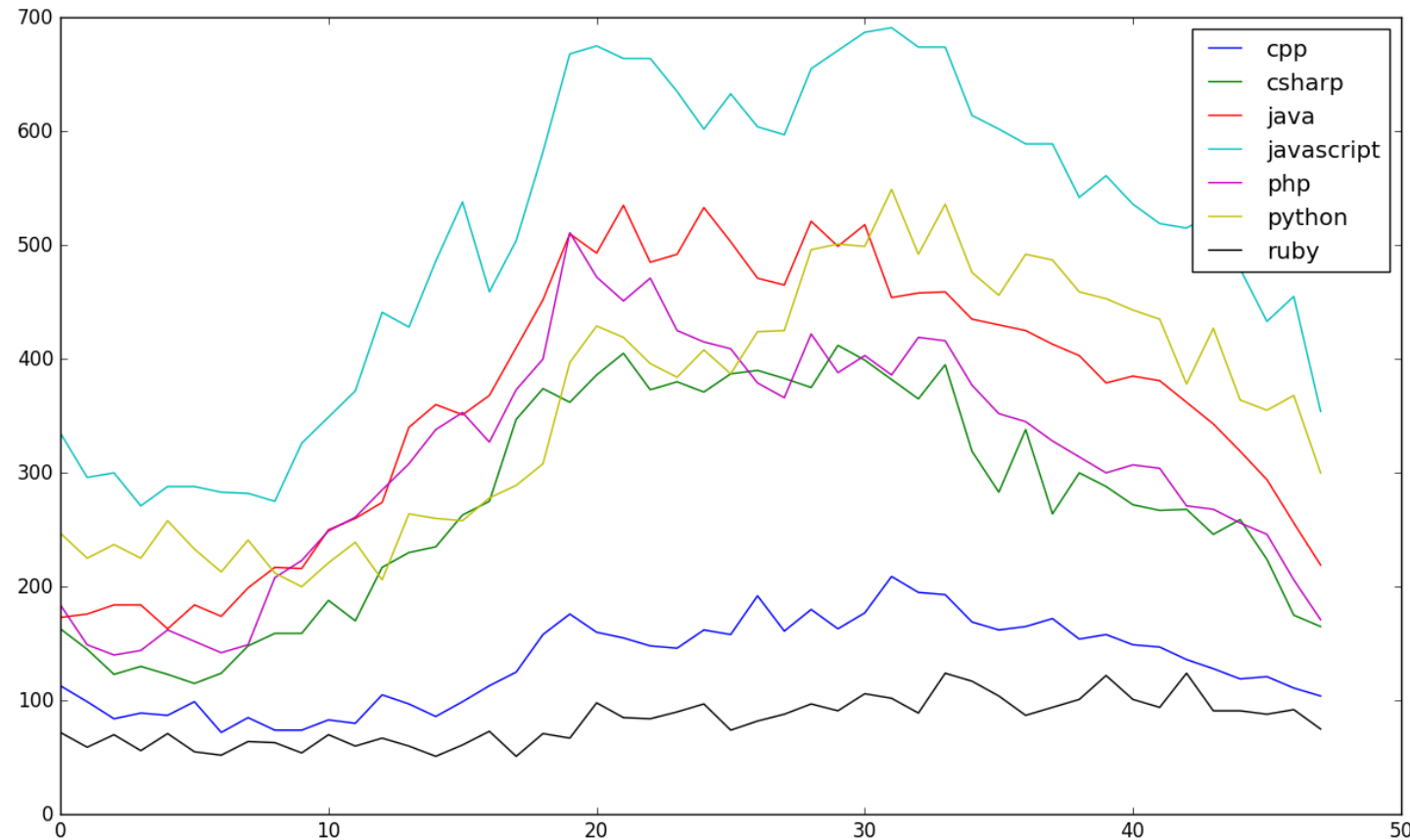
# Basic Customer Behavior Analysis

- Question, answer and comment number over programming languages



# Basic Customer Behavior Analysis

- Questions post time over Programming Language

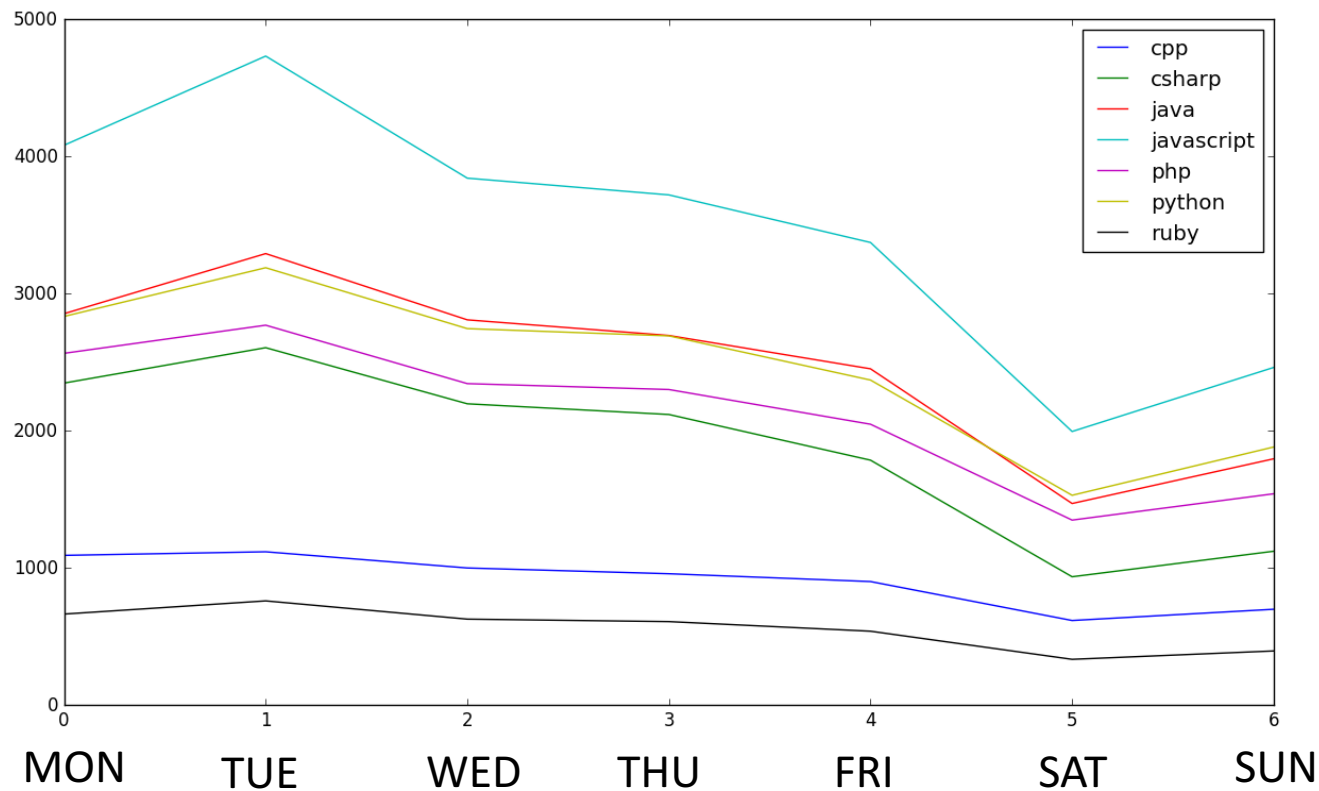


- C# may mainly be used during working time
- Python is more likely be used during the latter half of a day



# Basic Customer Behavior Analysis

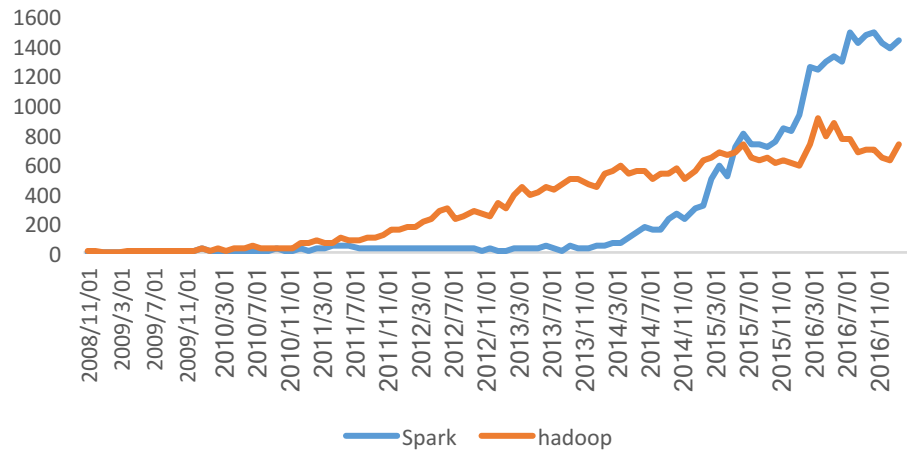
- Questions post time over Programming Language



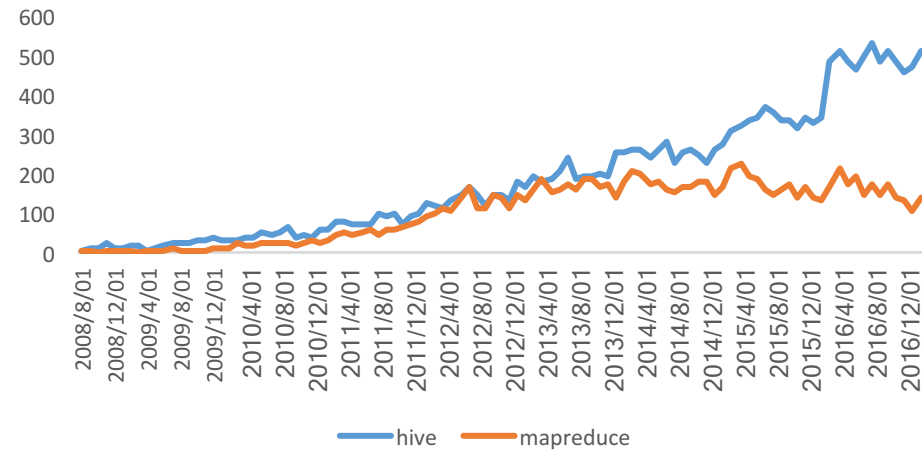
- Tuesday
- Python

# Basic Customer Behavior Analysis

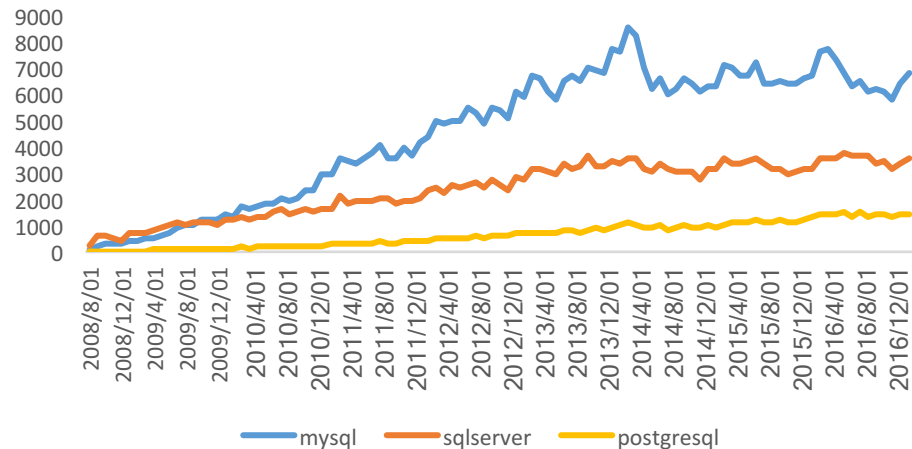
## Apache Tools



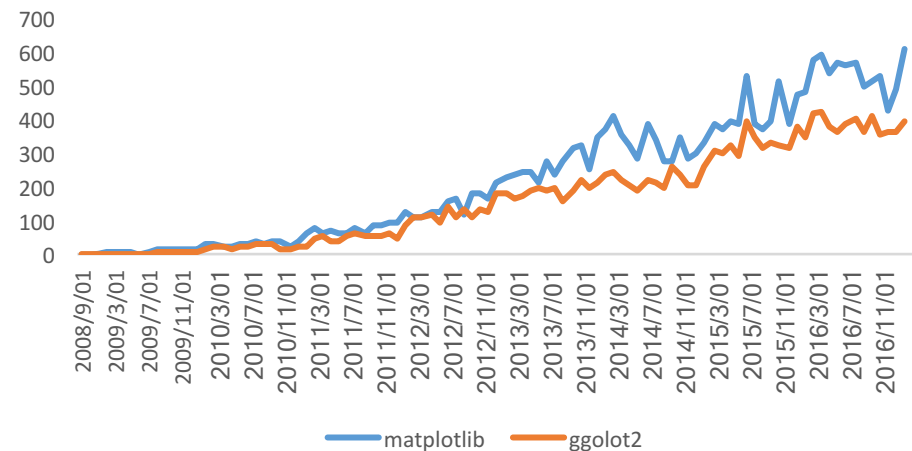
## Processing data tools



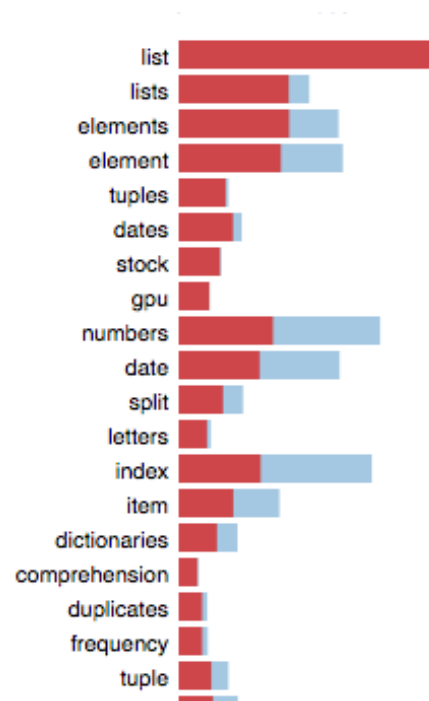
## Database



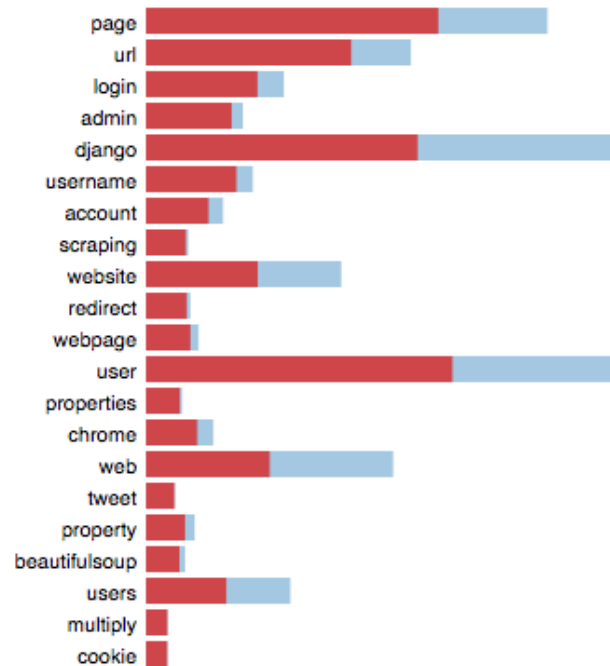
## Plot Tools for python and R



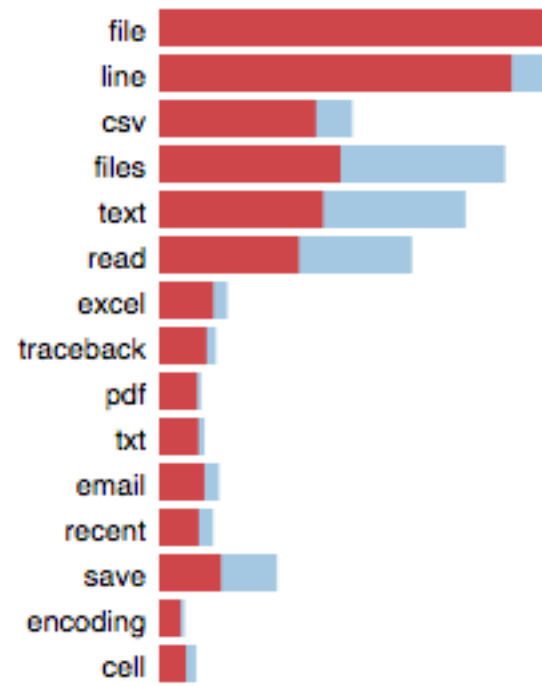
# Topics for python questions(LDA)



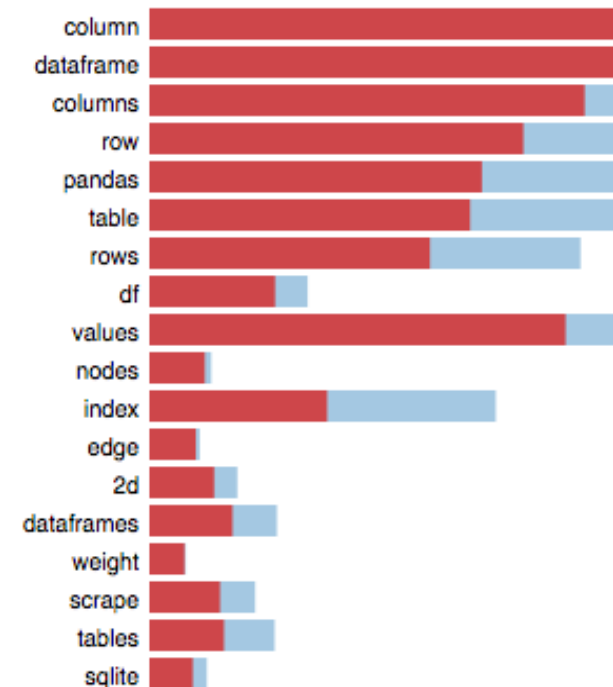
Python Syntax



Web

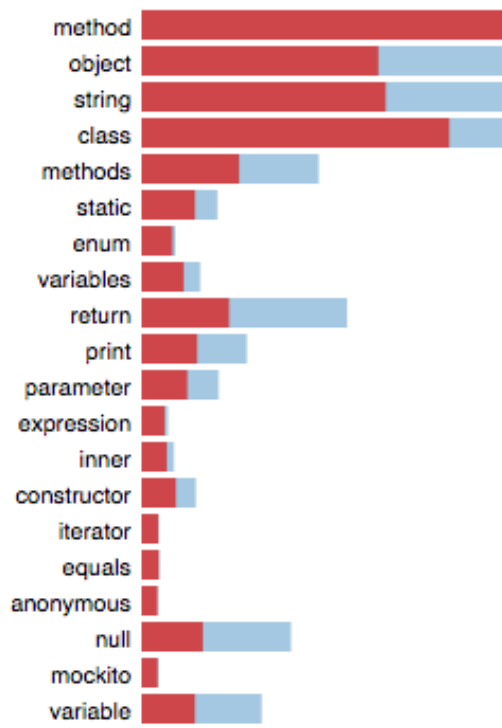


File

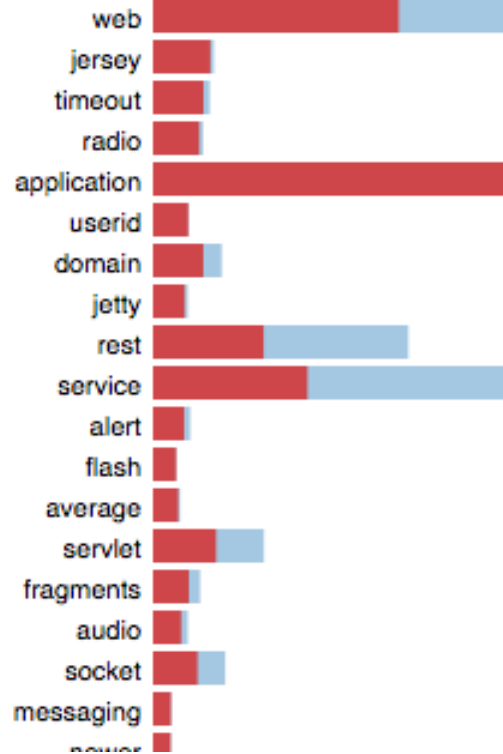


Pandas

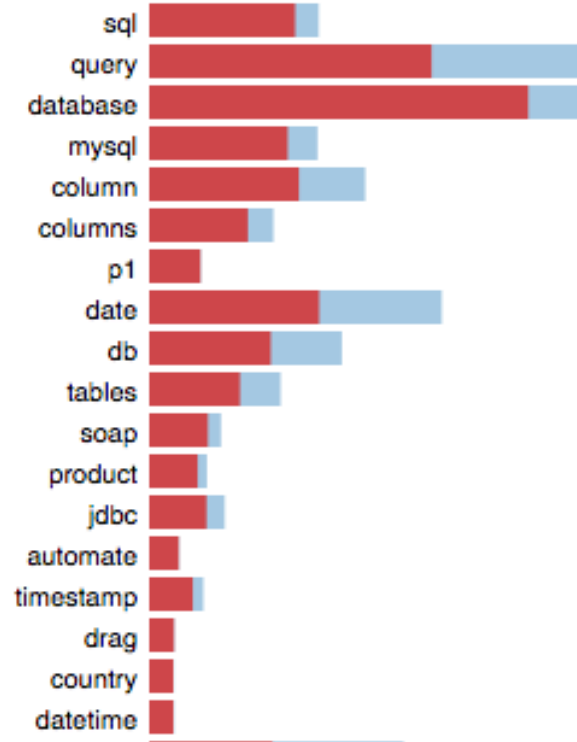
# Topics for Java questions(LDA)



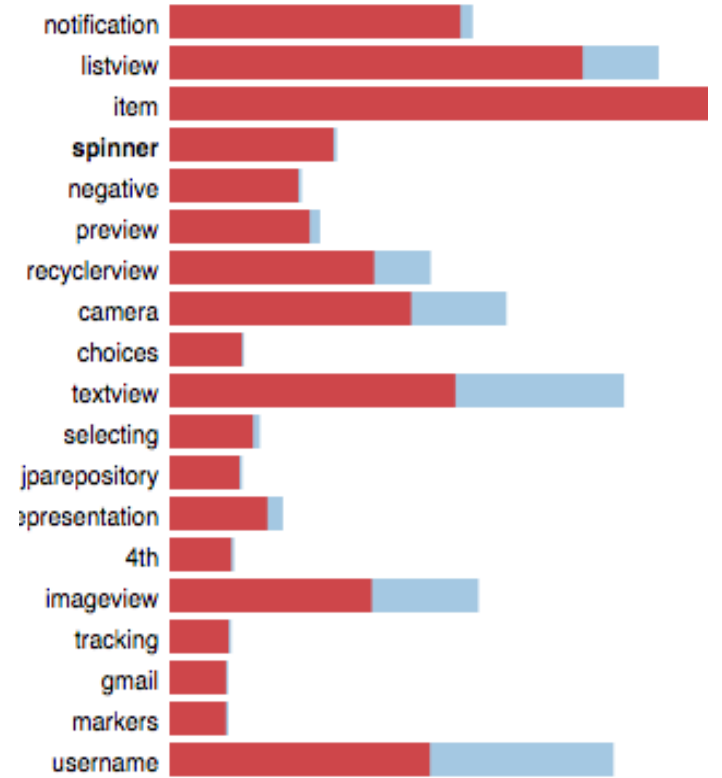
Java Syntax



Web



Database



Android

## Part 3 Text Classification for Better Tags

# Classification

- Only 20% of problems' description contain the programming language name.

Title

**B** *I*

Links Images Styling/Headers Lists Blockquotes Code HTML [advanced help >](#)

how to use google sentiment analysis api?

**Tags**

☐ Send me new responses to my posts via email ([settings](#))

[Post Your Question](#) [discard](#)

Tags should be added by hand

# Classification

- Using TF-IDF Vector + Classification Model



## Part 4 Sentiment Analysis for Comments



# Sentiment Analysis Using Word2vec and LSTM

Text Source: Comments for questions

## Machine translation for multilingual sentiment analysis

▲  
0  
▼  
★

I am trying to do sentiment analysis for non english languages like japanese, chinese, german etc. I want to know if any Machine translator available for translating documents in these languages to english. I am working on JAVA so I should be able to call the API or tool. I have used google translator API so please suggest anything apart from it.

java multilingual sentiment-analysis machine-translation

share improve this question

asked Dec 19 '10 at 11:43



Jagdeep

1

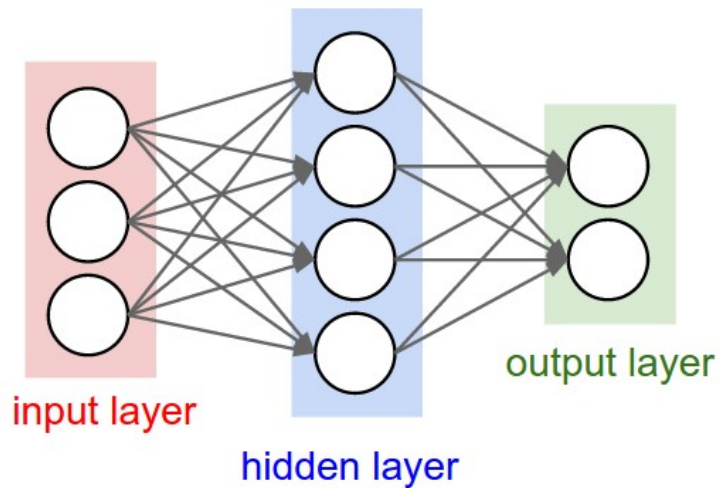
I sincerely doubt you'll be able to perform anything close to [sentiment analysis](#) on Japanese text using only an English translator. If you already have something implemented for English, chances are that it will need to be at least partially reimplemented in order to work with most other cultures. – [Frédéric Hamidi](#) Dec 19 '10 at 11:49

Knowing some Japanese; you can understand every word said but not understand what is meant. They imply much more in their language than we do in English which is much more direct. e.g. Japanese don't like to say no, but have many ways of sounding positive but really meaning no. It a bit like in English when you say thank you in a strained overly polite way, you are being sarcastic and actually mean the opposite of what you say. Japanese has many more subtleties like this than English. – [Peter Lawrey](#) Dec 19 '10 at 12:01

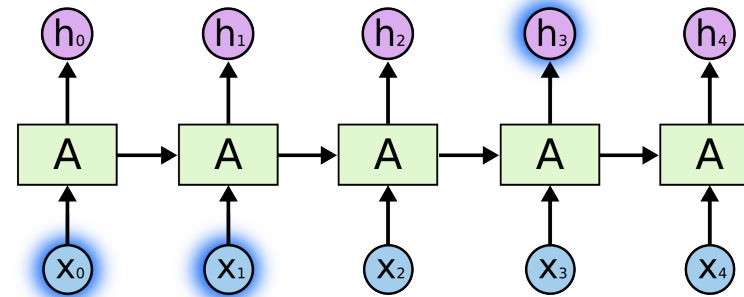
An example; a friend studied masters in Japanese literature, lived in Japan had a Japanese boyfriend for five years. Yet when he proposed marriage, she had no idea he had done so. :P There is a traditional poem about nature and the cycle of life which implies he wants her to spend the rest of their live together, she accepts by saying the second half of the poem. – [Peter Lawrey](#) Dec 19 '10 at 12:05

# Sentiment Analysis Using Word2vec and LSTM

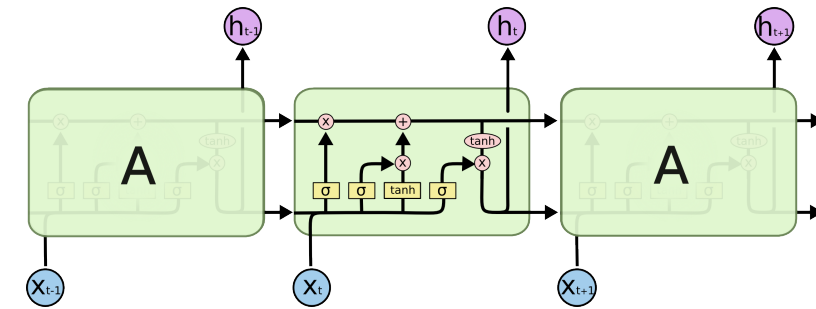
How Google sentiment works?



Neural Network

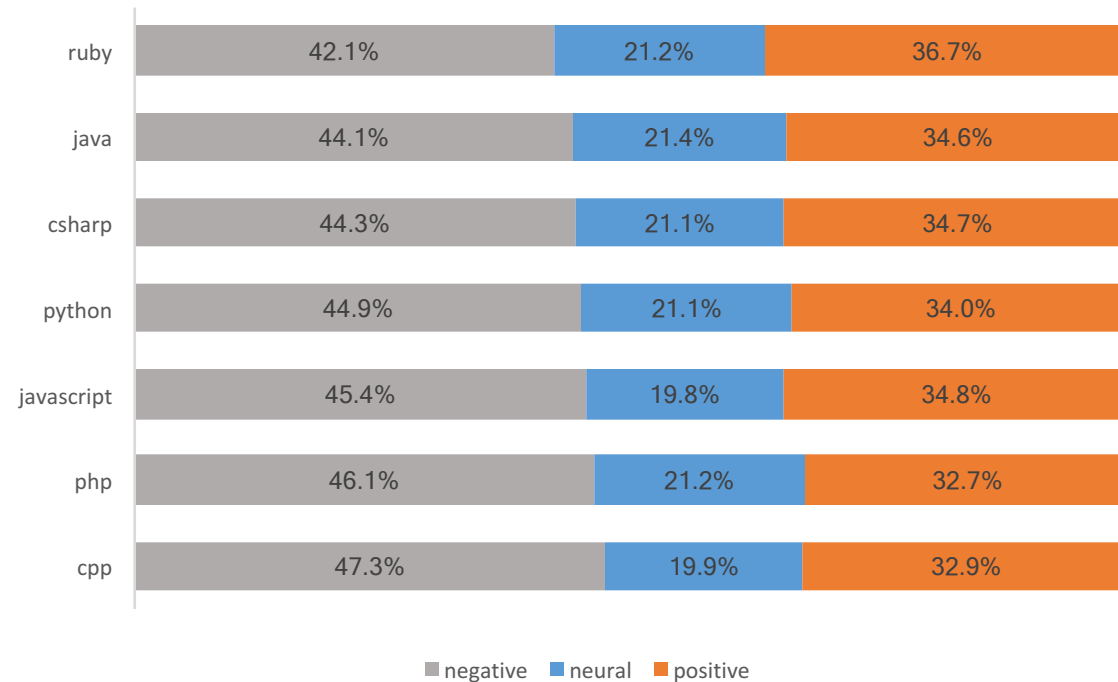


Recurrent Neural Network



Long Short Term Memory

# Sentiment Analysis Using Word2vec and LSTM



Sentiment Score