# Jarvis for Piazza

By Pranav, Isaac and Ke Yu

# Problems we Tackle

1. Duplicate questions asked by students
2. Too many posts for students to go through
3. Average search facilities

**Scale ~600 students**

# How we solve it

**Duplicate Detection**

- ➔ Compute Similarity
- ➔ Features
  - ◆ Subject
  - ◆ Body
  - ◆ Folder
  - ◆ Date
- ➔ Top ‘n’ documents with high confidence

**Weekly FAQ**

- ➔ Determine the top posts/questions of the week
- ➔ Features
  - ◆ Good question endorsement
  - ◆ Good answer endorsement
  - ◆ Upvotes from students
  - ◆ Unique Views
  - ◆ # of students involved in followups

**Bad search feature**

- ➔ Relevance based
- ➔ N-gram based look up
- ➔ Post importance
  - ◆ Recency also a factor
  - ◆ Key words
- ➔ Stats based boosting

Question History:

question

# edits to post

173 views

Actions

## Final_Practice Complexity Classes

In the best/worst/average complexity table question, the average complexity of both hash table searches is listed as O(n/m). It feels as though in most cases m will equal n, making this an O(1) complexity. It also just feels strange using a separate variable when using big-O complexity. When is it appropriate or necessary to use additional n-related variables for the complexity of an algorithm?

final_exam

~ An instructor (Nathan Fenner (nrfenner)) thinks this is a good question ~

edit | good question | 1

Updated 4 months ago by [redacted]

**the students' answer,** *where students collectively construct a single answer*

M is the number of buckets. If the number of buckets is equal to n, the number of items in the hash table, then an average case of O(1) should be expected. If there are a lot more elements than buckets, you would expect a worse complexity.

You should use a number other than n when the complexity is dependent on more than just one variable. Usually, input size is all that really matters, bu there are cases where complexity can vary dramatically dependent on a different variable. If the number of buckets is 1, wouldn't you expect a vastly different complexity that if the number was infinite?

~ An instructor (Andrew Marino (aymarino)) endorsed this answer ~

edit | good answer | 2

Updated 4 months ago by [redacted]

**the instructors' answer,** *where instructors collectively construct a single answer*

Another quick example of where complexity depends on multiple variables: searching for an element in a 2-D matrix of size $n$ by $m$ is $O(nm)$. If $n$ where on the same order as $m$, we could simplify it to $O(n^2)$ but we don't always know this.

edit | good answer | 0

Updated 4 months ago by Andrew Marino (aymarino)

# What tools did we use?

- Built using Python
- Whoosh library for Indexing documents
  - Full indexing at start
  - Incremental indexing thereafter
- Piazza Unofficial API (Lot of reverse engineering)
  - For reading from and posting to Piazza with Jarvis bot
- Sklearn for similarity (cosine)

# Demo

# Evaluation (Pending)

**Duplicate Detection**

- ➔ If an instructor marked it as a duplicate with '@65', could we find it too?
- ➔ Did we find duplicates that instructors couldn't?

**Weekly FAQ**

- ➔ Top weekly post is highly subjective, but looking at the result, the FAQ post definitely brings out common questions asked on piazza.

**Search**

- ➔ Compare with piazza's search results. MAP, precision @ k

# Thanks!

**Source Code:**

https://github.com/pranavr93/piazza_bot