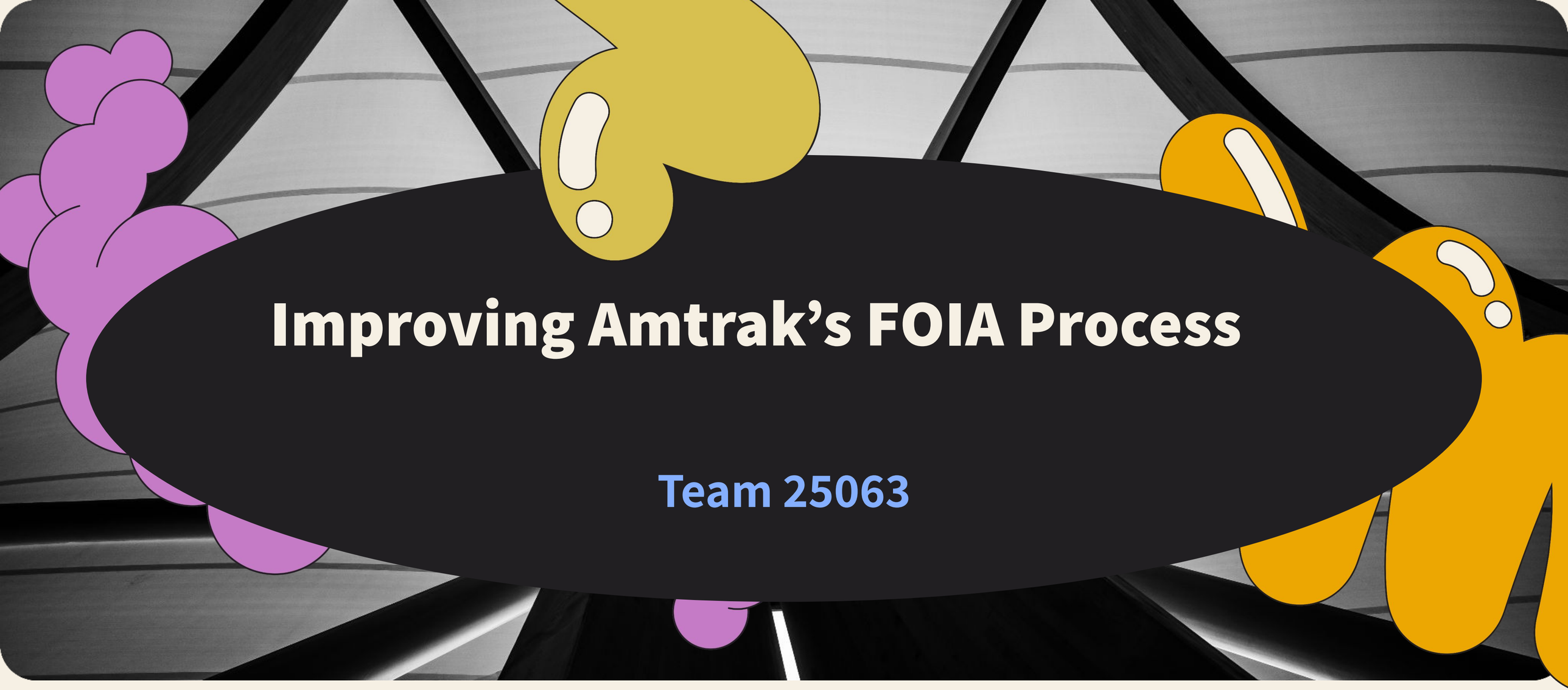


Improving Amtrak’s FOIA Process

Team 25063

Agenda

01 Introduction

02 Problem

03 Solution

04 Analysis of Results

05 Prototype Demo

Introduction

Luis D. Gascon
Towson University

Andreas Papacharalampous
Towson University

Rayquan Williams
Towson University

Use of AI

Generative AI was heavily used throughout this project

Platform	Google AI Studio. Gemini 2.0 Flash model
How it was used	Explain concepts concisely. How to generate a chart using Altair. Debug through errors. Determine where a keyword should go based on a department's name
Learning points	Filtering dataframes. Applying functions to each row of a dataframe. What stop words are in the context of keyword extraction. Libraries to use

Problem

The way Amtrak processes FOIA requests is inefficient and contains a single point of failure. This can introduce bottlenecks, and create confusing processes.

Proposed Solution

Use of cosine similarity to compare texts, and good ol’ algorithms!

Data Cleanup

There are FOIAs that contain the following keywords:

- “Duplicate”
- “Not a proper FOIA request”
- “unclear”

These FOIAs will only pollute our keyword analysis and slow down the mapping process

Classifying FOIAs

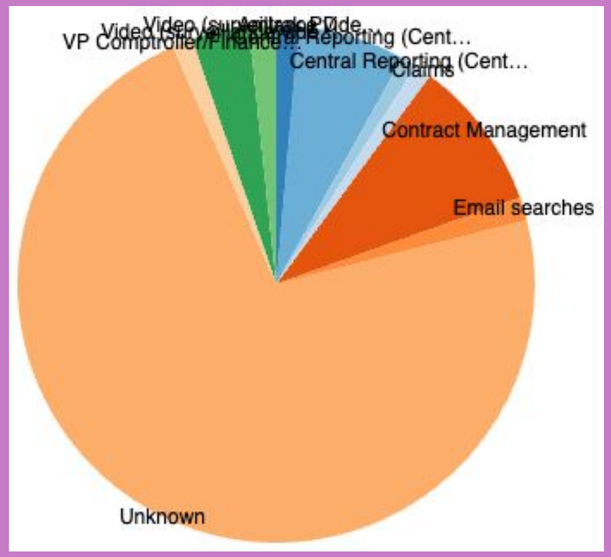
- Our initial solution was to create an algorithm that will look at each FOIA in the database, and categorize them based on *exact* keywords that are matched to a department.

- Can slow down with a larger data set
  - ... takes 2 seconds with ~4,000 requests!

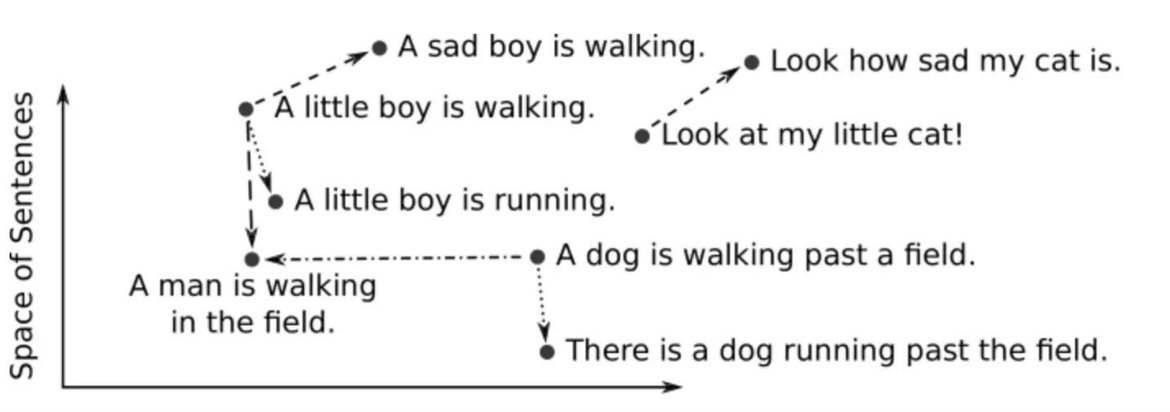
```
a ~/D/C/I/2/C/IC25-2025 main λ time python3 src/classification.py

------------------------------------------------------
Executed in    2.35 secs    fish           external
   usr time    2.25 secs    0.08 millis    2.25 secs
   sys time    0.07 secs    1.26 millis    0.06 secs
```

Timing stats, takes ~2s to run.

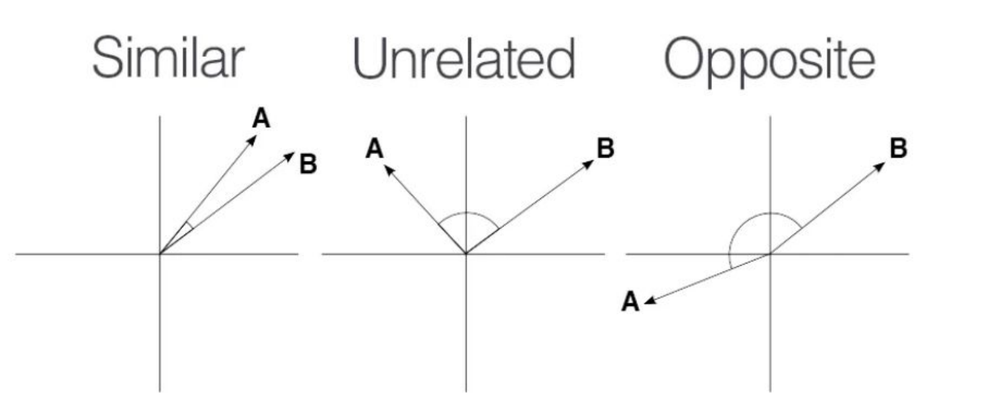


Final results. Too many unknowns :(



Encode department name plus keywords into a vector

Encode FOIA description into a vector



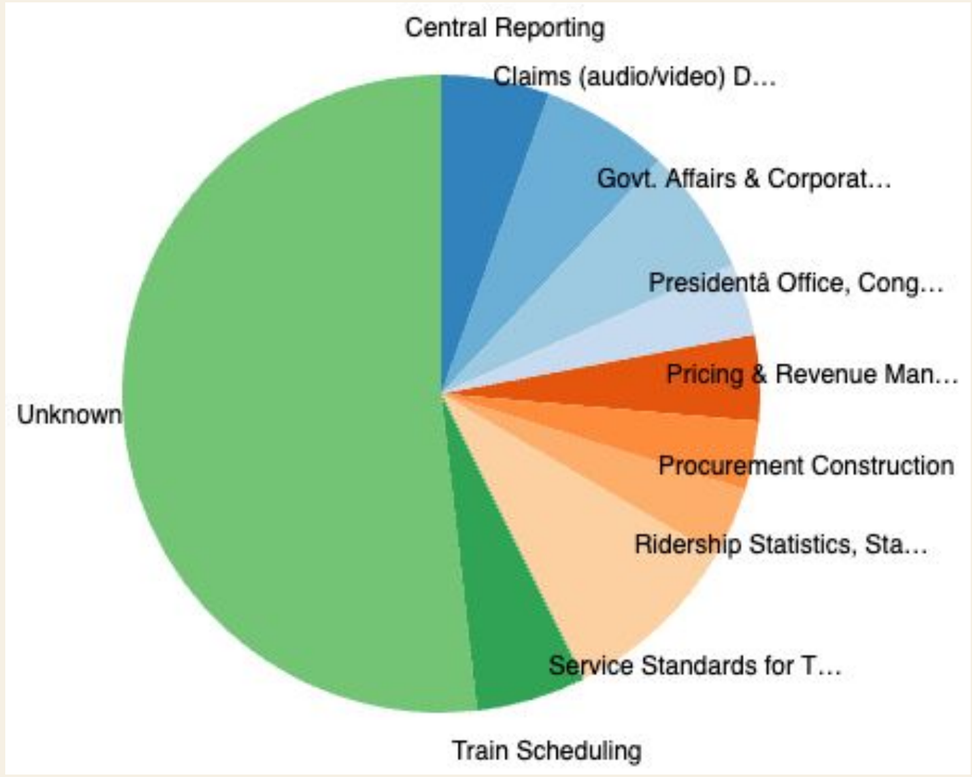
Map FOIAs by the department whose similarity score is above .50, otherwise, set as unknown

Trick to Remember Trigonometric Table

cuemath
THE MATH EXPERT

	0°	30°	45°	60°	90°
SIN	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
COS	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
TAN	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	Not Defined

- Only 85 out of 4,119 FOIAs have a similarity score above 60%!
- Lots of unknowns because of missing keyword matches. Not good! :(

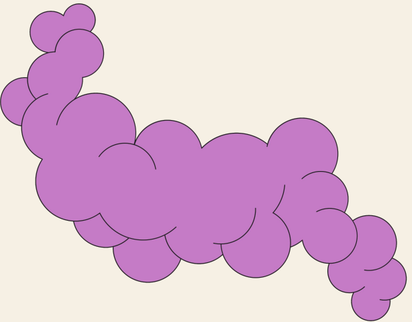


Little over ~1,770 unknown FOIAs w/ cosine similarity!

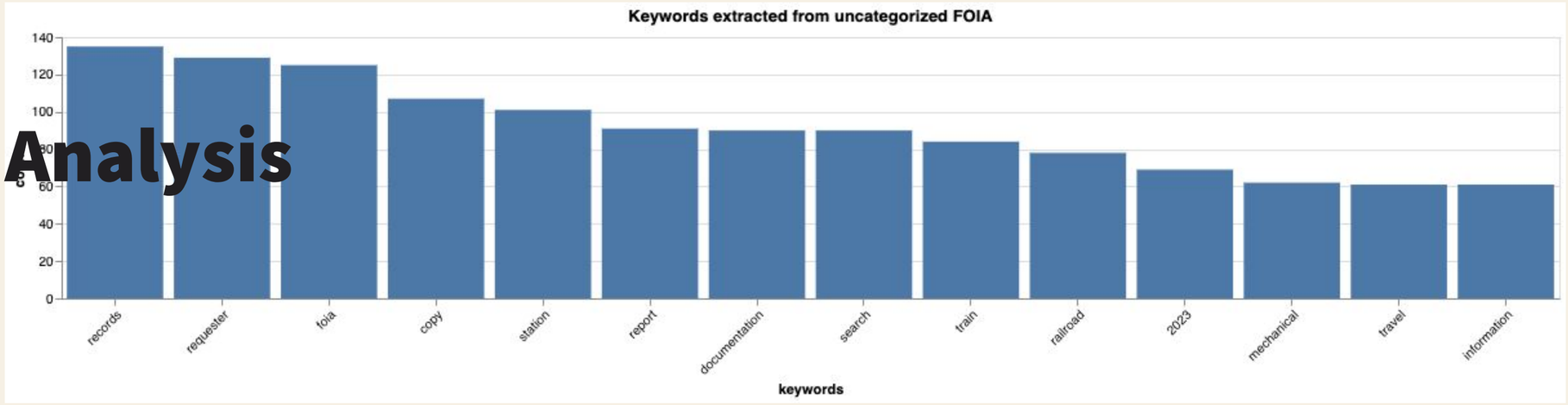
Analysis of Results

How can we improve this?

- Departmental data provided by Amtrak could include key contextual information to help us create matching keywords.
  - Performed our own research on common keywords such as “RFP (Request for Proposals)”, and “OIG (Office of the Inspector General)”
- Filter out irrelevant keywords from FOIA requests, such as common English words, or keywords that have already been matched.
  - Machine learning opportunity!

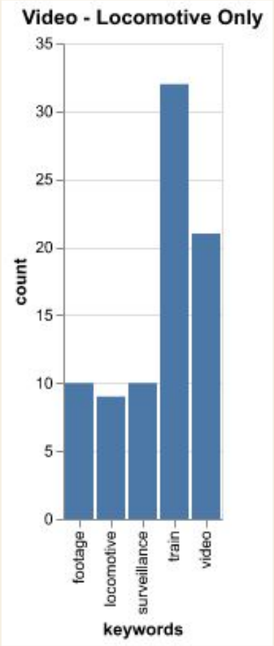


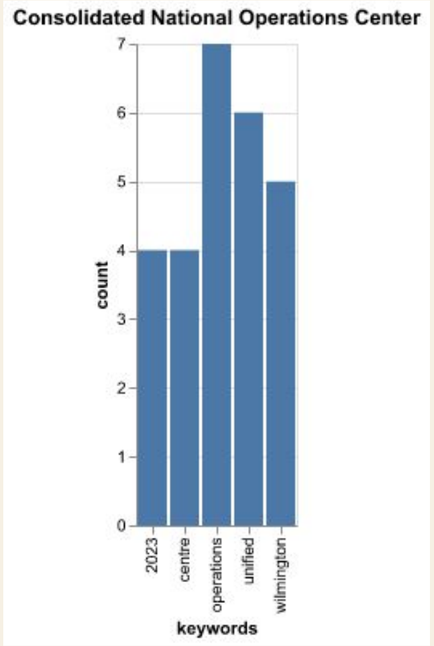
Keyword Analysis

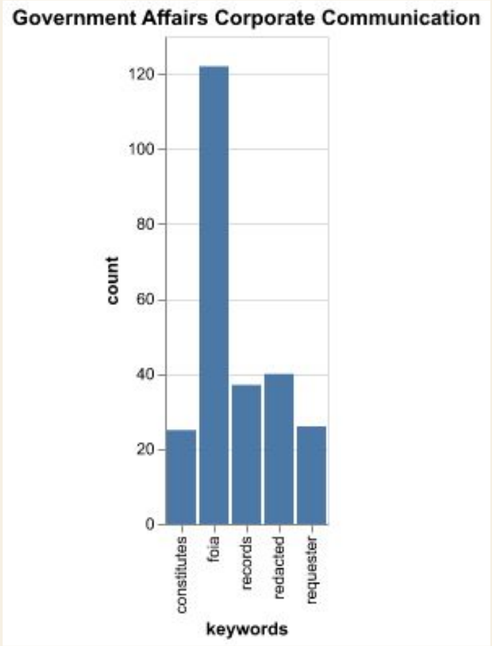


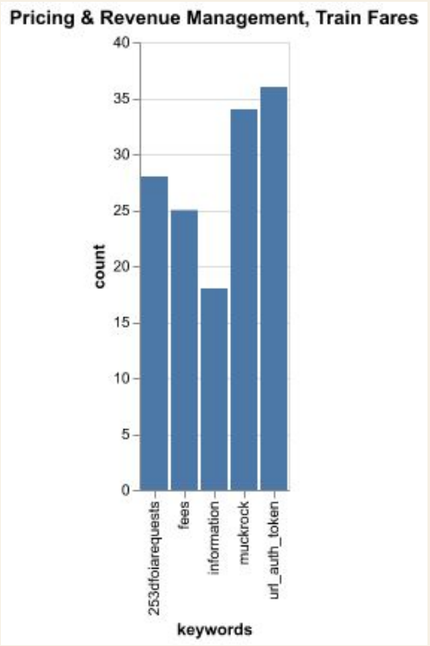
KeyBERT

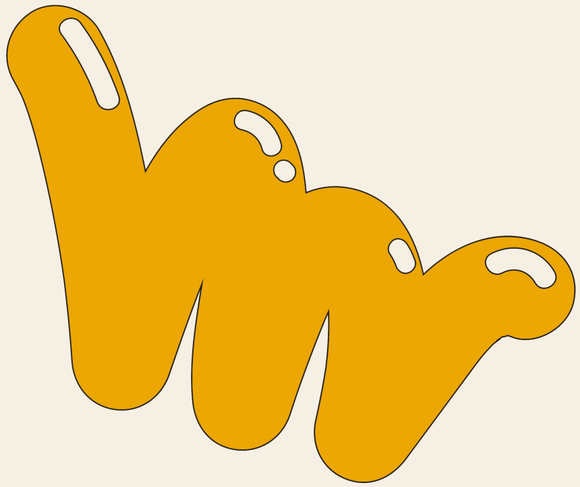
Extracted keywords to understand why cosine similarity scores are low...











Conclusion

- Although the results of cosine similarity look poor, the accuracy will only improve as more relevant keywords are added to a department.

- We believe this solution is scalable as we'll show in our prototype demo

- Introducing a more sophisticated traditional algorithm could also be beneficial in the long term.
  - ... or a combination of both solutions!

Prototype Demo

THANK YOU!

Any