FOOGLE

Search Engine

Allen Lau & Syed Faquaruddin Quadri

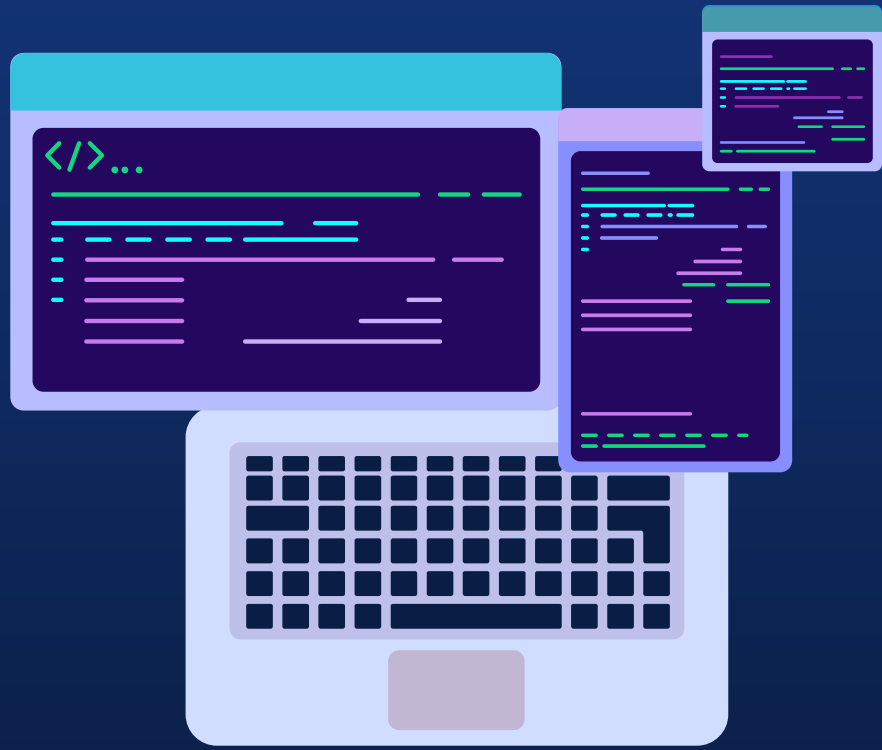


TABLE OF CONTENTS

01
Project Overview
What is Foogle?

02
Contributions
Who did what?

03
Search Engine Architecture
How do we go from query to results?

04
Demo
What does Foogle look like?

06
Database
How do we store the data?

06
Backend
What is the logic for populating and querying?

01

Project Overview

What is Foogle?

Foogle is a search engine aggregator, which compiles the results from other widely used search engines, stores the results into a mySQL database, and then displays the most relevant results by keyword to the consumer.

Database	MySQL Database
Backend	Python, Flask
Frontend	HTML, CSS

THEM VS US

THEM

- Search Engine companies are ultimately for-profit organizations
- Ranking algorithms are often business driven
- Ads are given priority

Foogle

- Utilize word frequency instead of ranking algorithms
- Ads are filtered
- Refined secondary searches from resultant first set of searches

02

Contributions

Contributions

Allen Lau

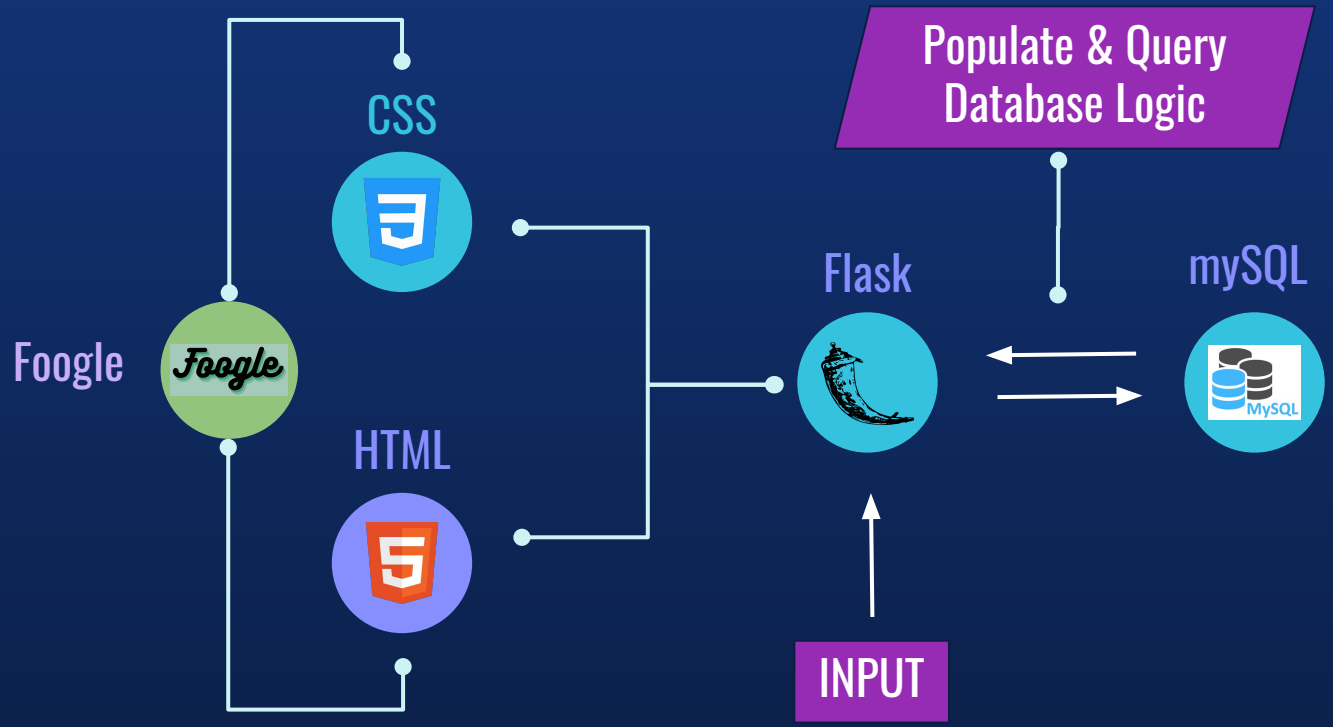
- MySQL Database
- Populate & Query Database Development

Syed Faquaruddin Quadri

- Flask Web Framework
- HTML/CSS Website Development

03

Search Engine Architecture

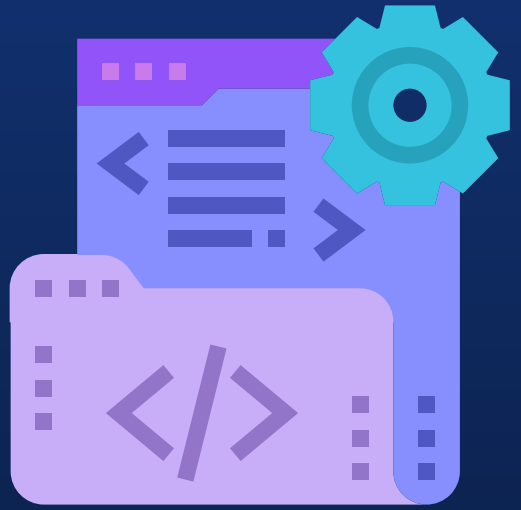

Populate & Query
Database Logic
CSS
Flask
mySQL
Foogle
Foogle
HTML
MySQL
INPUT

04

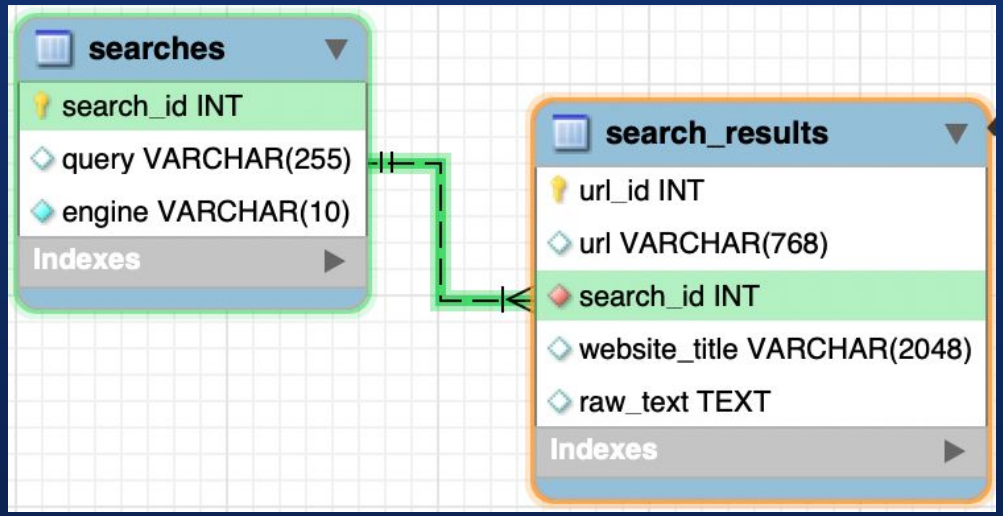
Demo

05

Database



mySQL Database ER Diagram



searches

search_id	query	engine
13	when was the start of the civil war	Google
14	when was the start of the civil war	Bing
15	when was the start of the civil war	Yahoo
16	when was the start of the civil war	DuckDuckGo
17	what is the tallest building in the word	Google
18	what is the tallest building in the word	Bing
19	what is the tallest building in the word	Yahoo
20	what is the tallest building in the word	DuckDuckGo

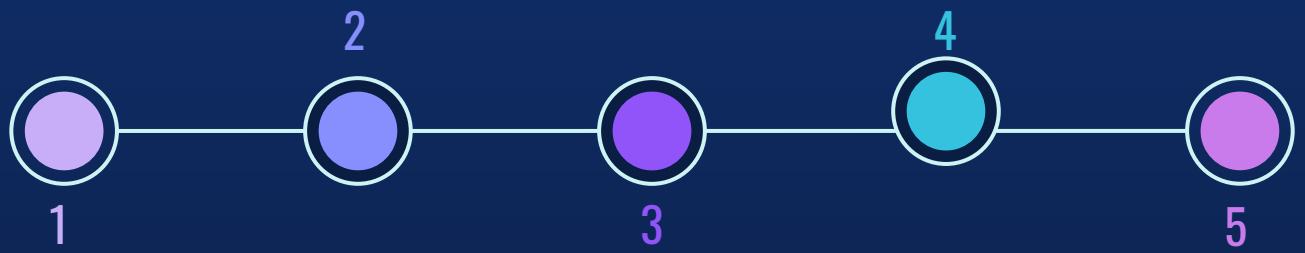
search_results

url_id	url	search_id	website_title	raw_text
164	https://www.history....	13	Civil War - Caus...	By: History.com Edit...
165	https://americanhis...	13	The Civil War B...	Today's Hours10:00...
166	http://www.abraha...	13	Beginning of the...	After Lincoln's electi...
167	https://www.nps.go...	13	Civil War Timeli...	Library of Congress...
169	https://www.though...	14	American Civil...	The Civil War Transf...
170	https://www.wester...	14	Post I The West...	The Western Theate...
172	https://www.history...	14	Where did the C...	The Civil War starte...
173	https://bing.com/ali...	14	Civil War Photo...	Still Pictures Engine...
174	https://www.studynl...	14	Causes of the A...	Causes of the Ameri...

06

Backend

Populate Database Process



Retrieve query from end-user

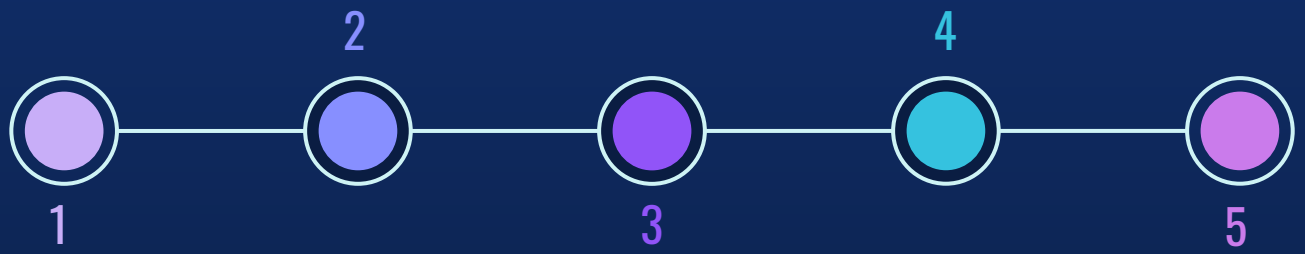
Website → Flask → Python

Populate Database Process

Send HTTP requests from Google, Bing, Yahoo, and DuckDuckGo to obtain URLs

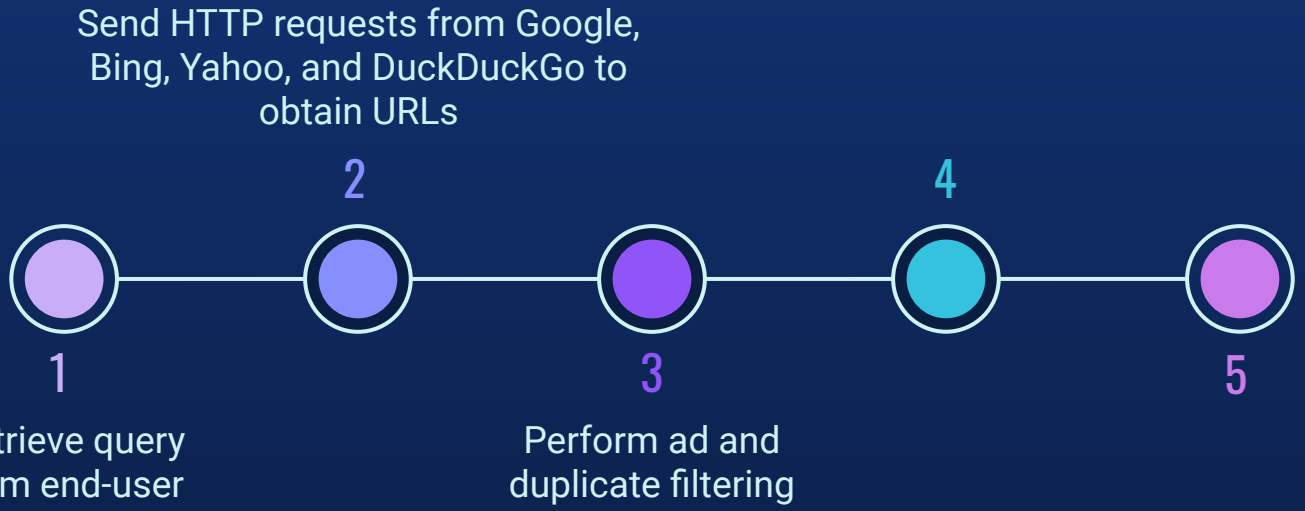
Python Requests to retrieve URLs from search engines.

Python Asyncio & Aiohttp to asynchronously retrieve HTML text from URLS.



Retrieve query from end-user

Populate Database Process



Python logic to remove duplicate URLs and URLs marked as ads

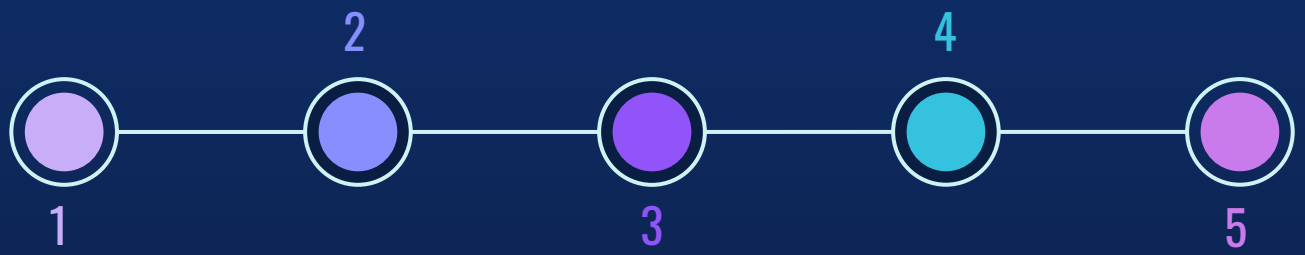
Ad Filtering Specifics

Filter on Ad Domains

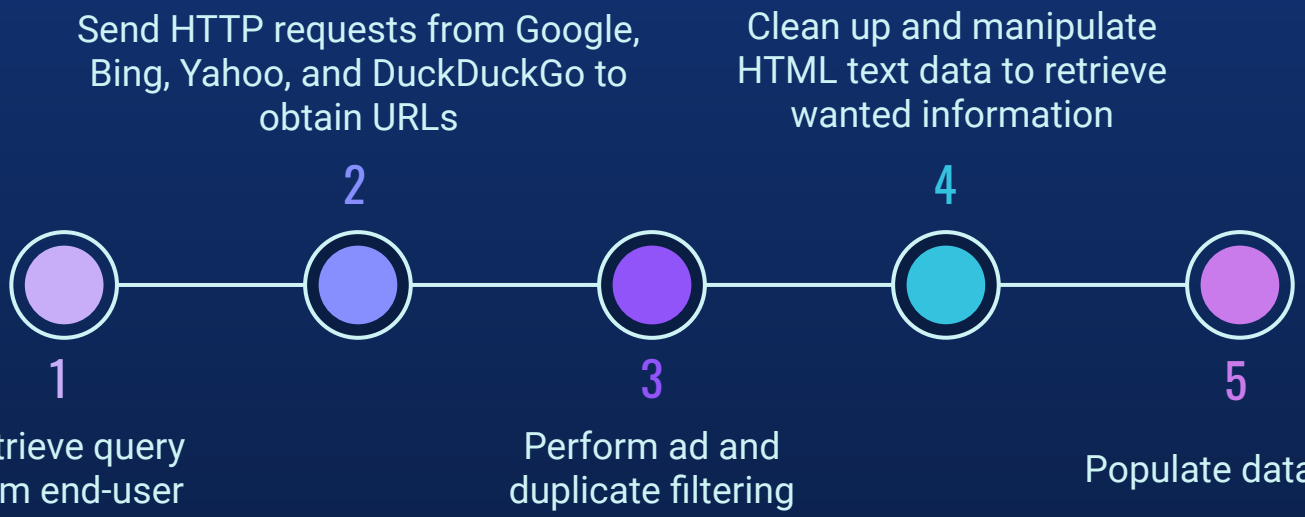
- http://www.google.com/acl
- help.ads.microsoft
- https://help.ads.microsoft.co
- https://advertising.yahoo.com
- https://help.duckduckgo.com/duckduckgo-help-pages/company/ads-by-microsoft-on-duckduckgo-private-search
- https://duckduckgo.com/y.js?ad_domain'

Populate Database Process

Python BeautifulSoup

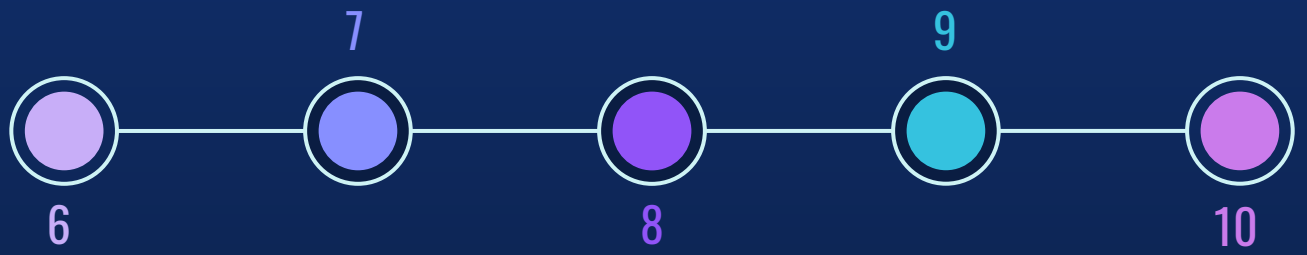


Populate Database Process



Python mySQL Connector

Query Database Process



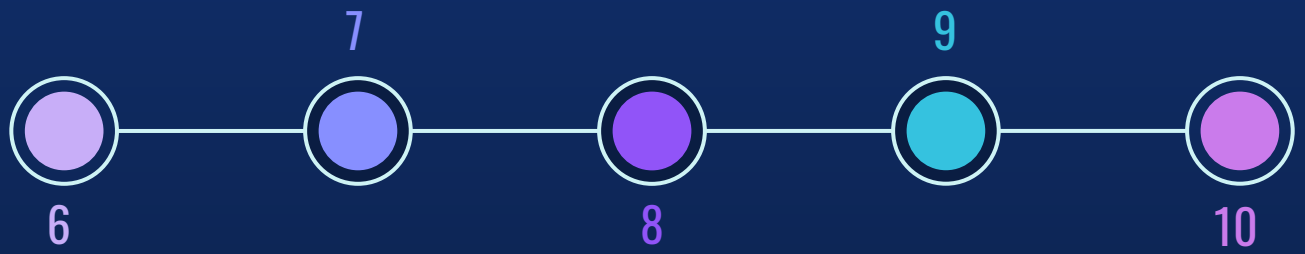
Retrieve websites from database based on the end-user query relevance with the text from the URL

mySQL Natural Language Full-Text Search

Query Database Process

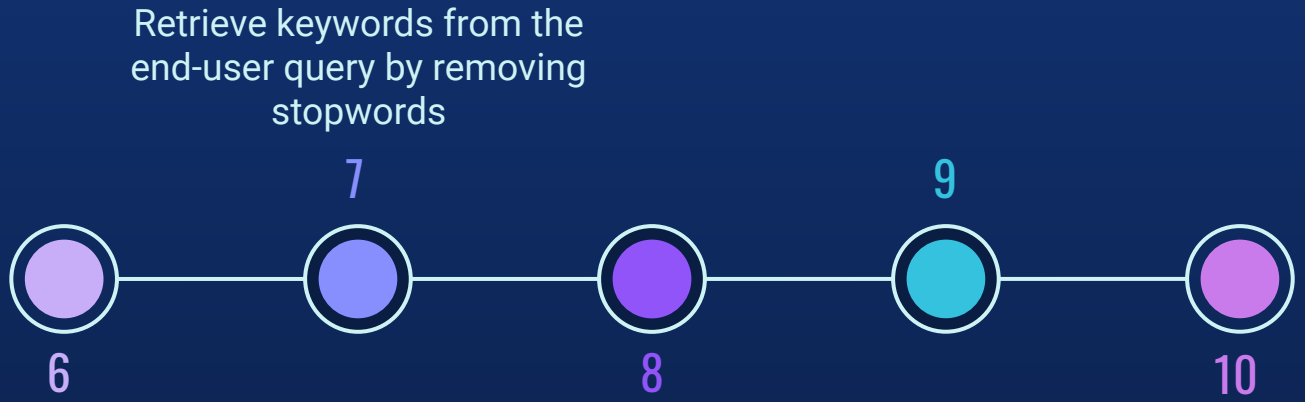
Python NLTK

Retrieve keywords from the end-user query by removing stopwords

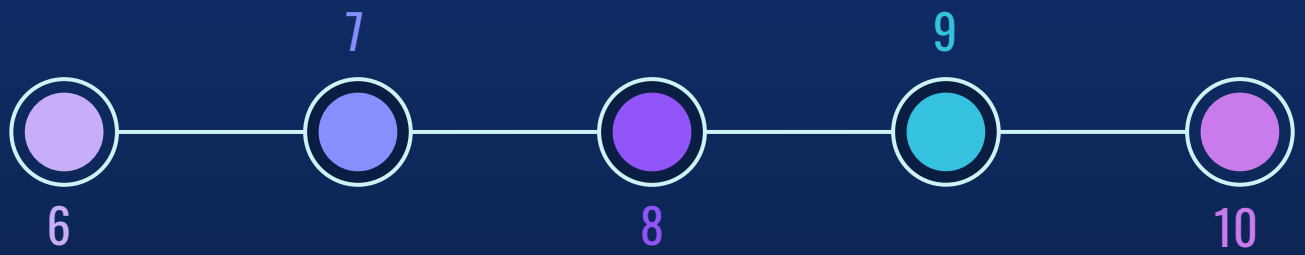


Retrieve websites from database based on the end-user query relevance with the text from the URL

Query Database Process



Query Database Process



Retrieve websites from database based on the end-user query relevance with the text from the URL

Retrieve keywords from the end-user query by removing stopwords

Count the keywords within the text of the retrieved websites

Rank the retrieved websites based on keyword count

Python logic for sorting by keywords

Keyword Ranking Algorithm

- Remove stop words from input query to get a list of keywords
- Use Python’s Collection Library to count the number of keywords in each website resulted from mySQL Natural Language Full-Text Search
- Sum the total number of keyword counts for each resultant website
- Display the website results based on keyword count, in descending order

Query Database Process

Flask → Website

Github Repository

https://github.com/alau002/search_engine.git

Syed Faquaruddin Quadri & Allen Lau

Thank You