Hops
A Beer Recommender

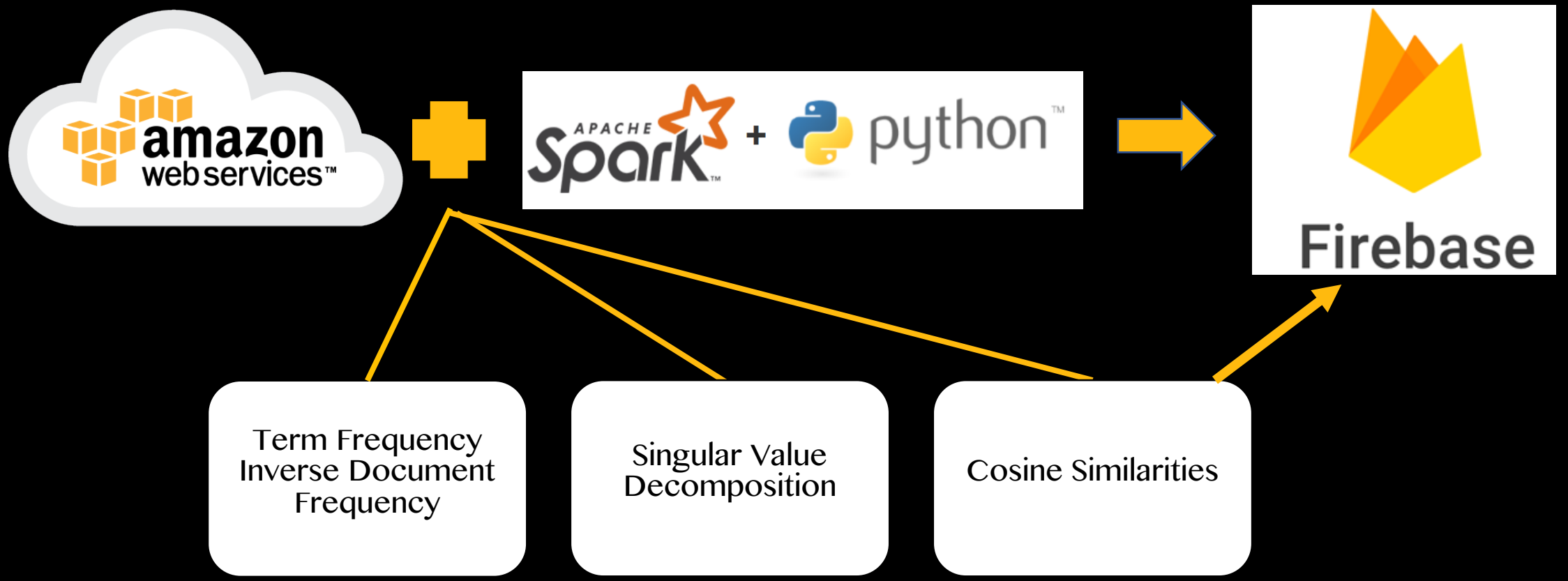
Presented by Kelly Ochikubo

Motivations

Interest in Natural Language Processing & Beer

Pipeline:
Data Collection to Web Application

Pipeline:
Data Collection to Web Application



Latent Semantic Analysis

Document 1

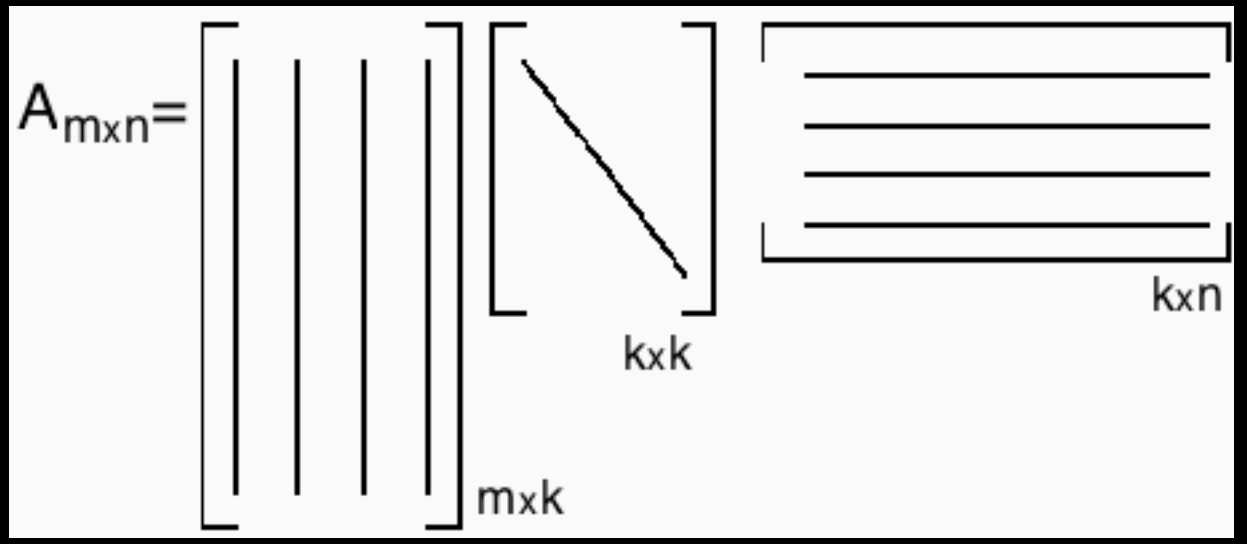
Term	Term Count
this	1
is	1
a	2
sample	1

Document 2

Term	Term Count
this	1
is	1
another	2
example	3

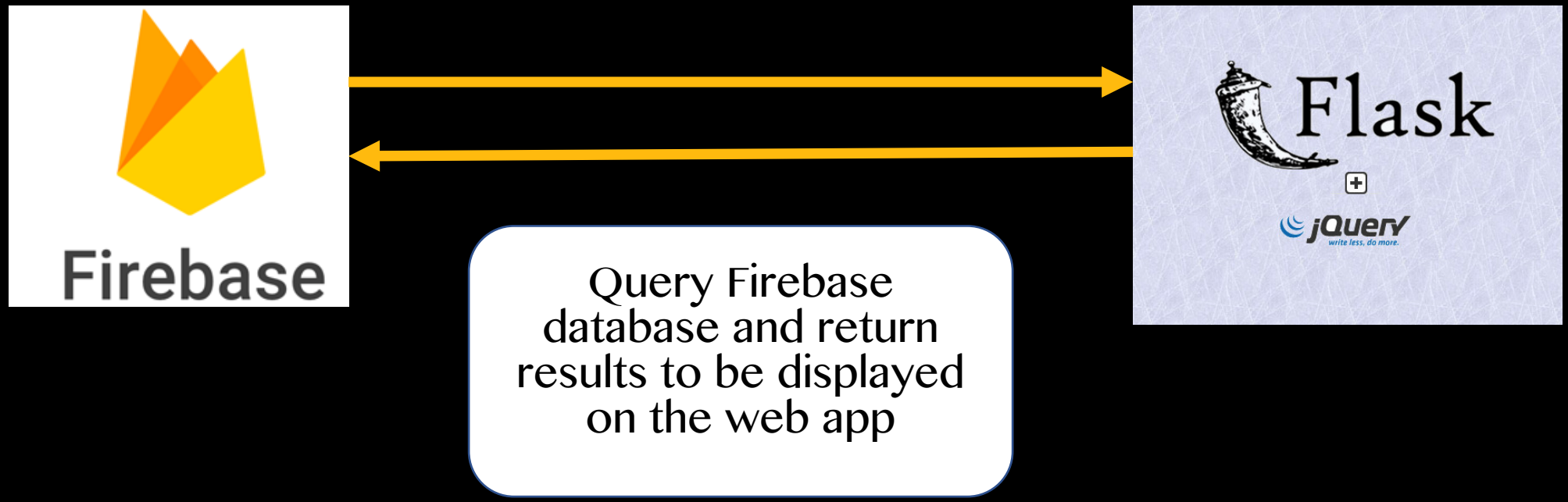
TF-IDF

+

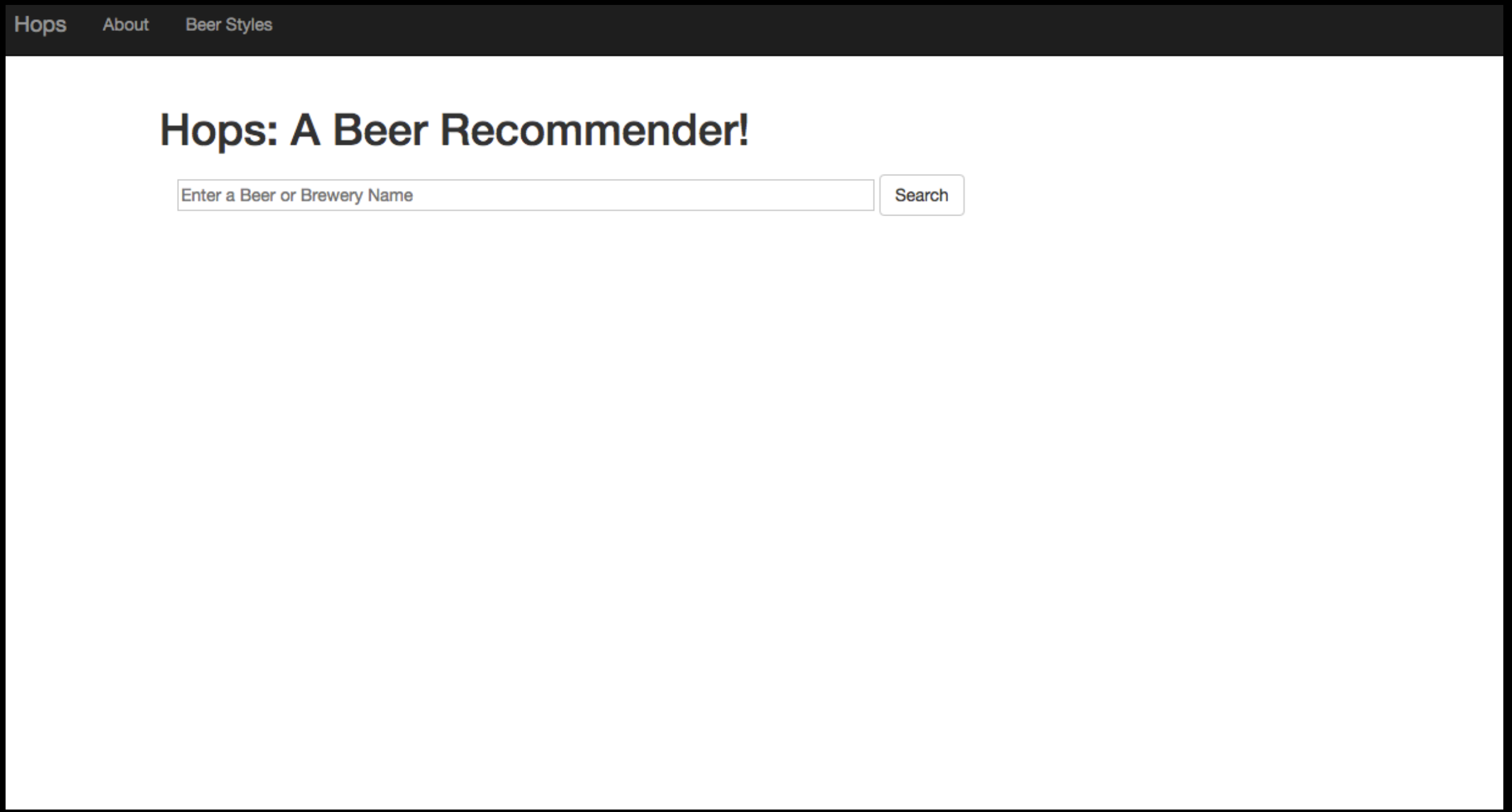


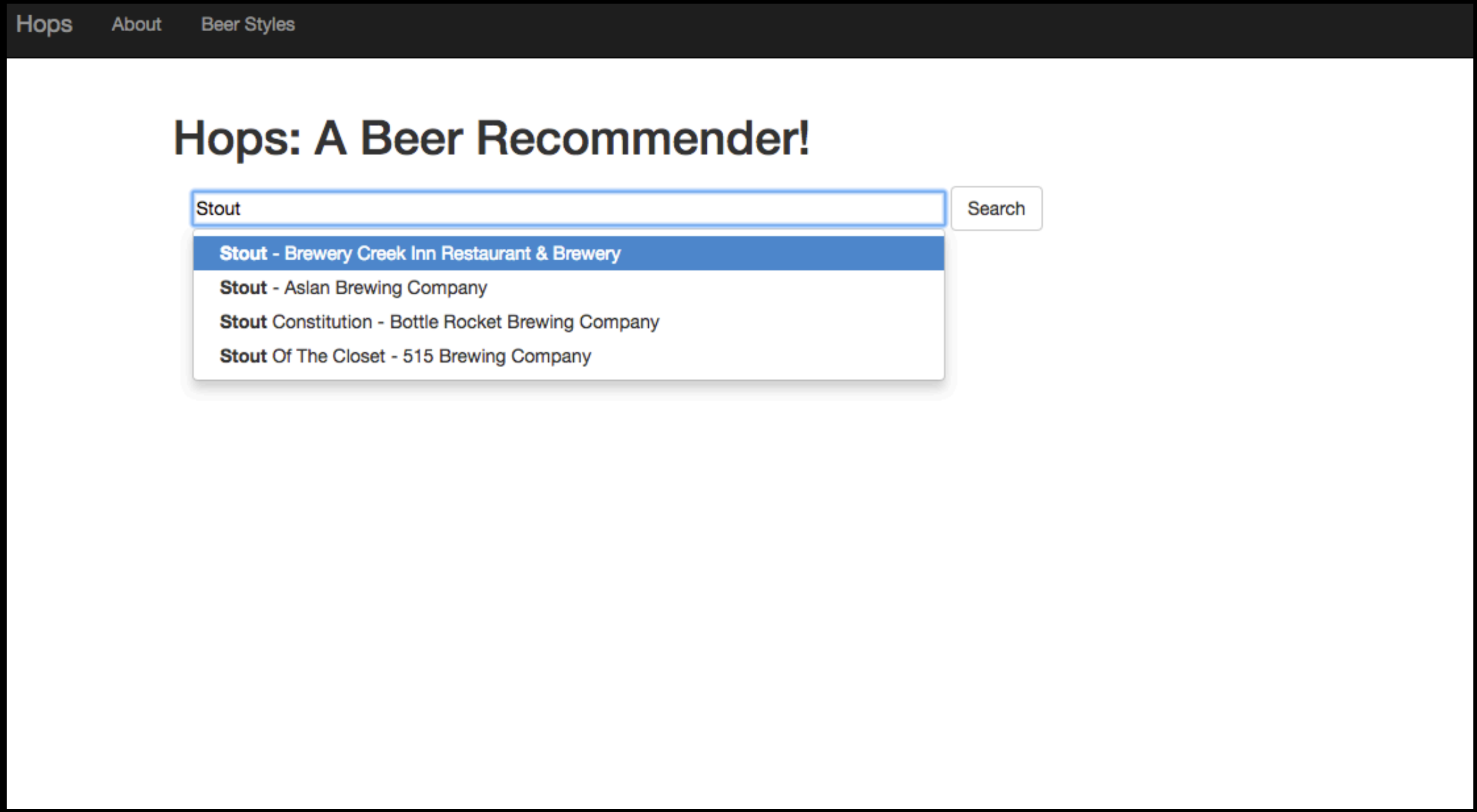
SVD

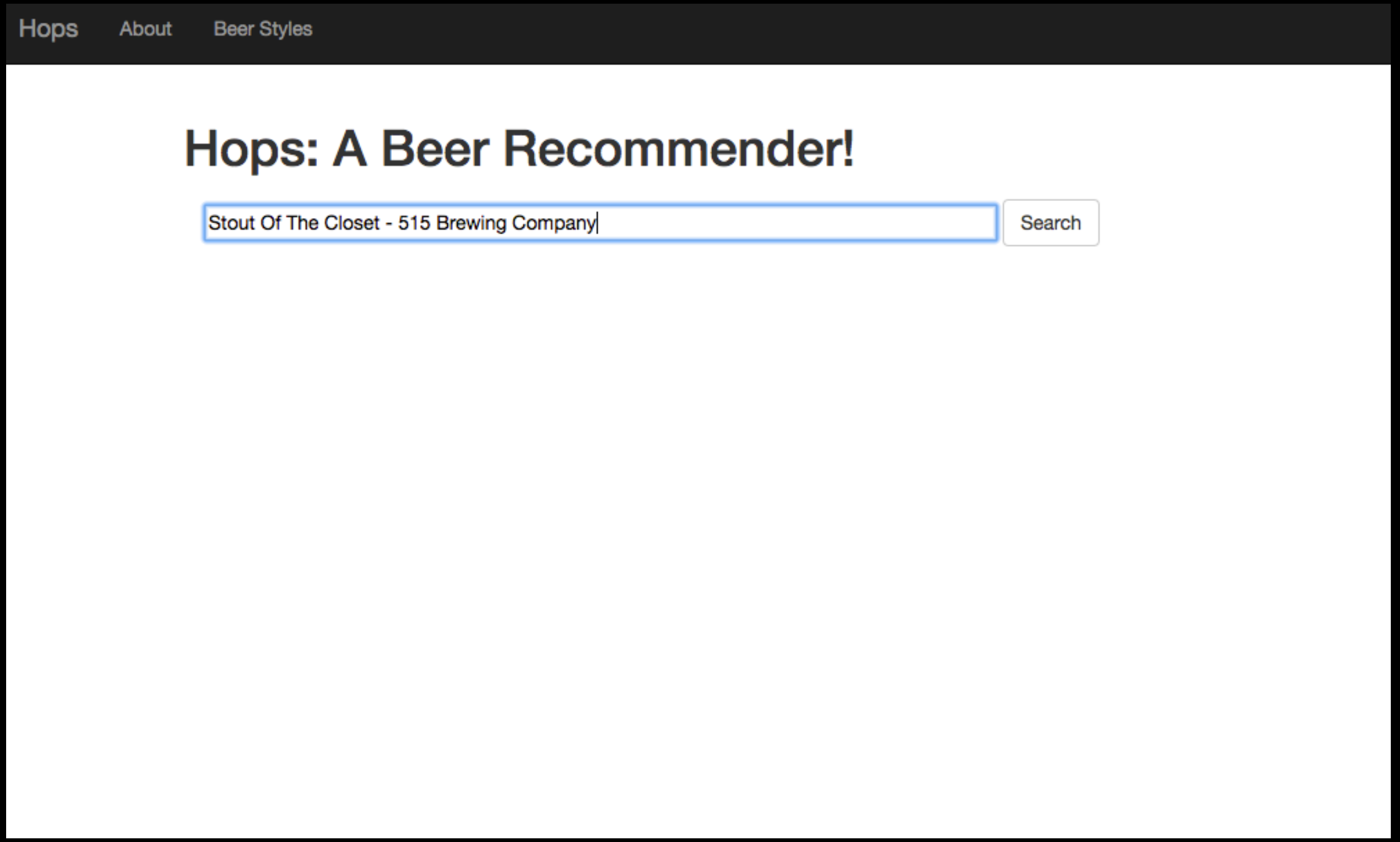
Pipeline:
Data Collection to Web Application



Hops Web Application


Hops
About
Beer Styles
Hops: A Beer Recommender!
Enter a Beer or Brewery Name
Search


Hops
About
Beer Styles
Hops: A Beer Recommender!
Stout
Search
Stout - Brewery Creek Inn Restaurant & Brewery
Stout - Aslan Brewing Company
Stout Constitution - Bottle Rocket Brewing Company
Stout Of The Closet - 515 Brewing Company


Hops
About
Beer Styles
Hops: A Beer Recommender!
Stout Of The Closet - 515 Brewing Company
Search

Top 5 Beers Similar to Stout Of The Closet by 515 Brewing Company

Montpelier Mocha
American Stout

Bent Hill Brewery

Vlad The Inhaler
Russian Imperial Stout

Burley Oak Brewery

Hot Chocolate Porter
American Porter

Freedom's Edge Brewing Co.

Cascadian Dark Ale
American Black Ale

7 Seas Brewery and Taproom

Bourbon Oak Aged Breakfast Stout
American Double / Imperial Stout

Cloud 9 Brewery

Challenges & Takeaways

- Firebase
  - Limited ways to query database, limit to space on free tier
- Apache Spark
  - No PySpark implementation for SVD (only exists for Scala/Java)
    - Time to learn Scala!
  - To take maximum advantage of Spark's capabilities, one must think in "Map Reduce"
  - Troubleshooting can be difficult especially if using PySpark because the issue could lie either in your high level code or the source code
    - Documentation is your BFF (obviously!)

Future Work

- Model Improvements
  - Only consider beers with a set minimum # of reviews
  - Take into consideration other features like location and beer rating

- Extensions
  - Apply same LSA concept to beer styles and see what can be learned about the similarities and differences between them
  - Add visualizations to the web application to show user these similarities and differences

Application

- Text Mining/Natural Language Processing can be difficult in terms of the curse of dimensionality
- As the dimensionality increases, the volume of space increases so fast that the data becomes sparse
- Matrix Factorization techniques such as Singular Value Decomposition can be used to reduce dimensionality while maintaining as much “power” as possible

THANK YOU

KELLY.OCHIKUBO@GMAIL.COM