Integro - Search it all!

Amey Ambade - aa4075
Alaap Sivaram - as5446
Maitri Shroff - ms5524
Asha Jacob - aj2782

Introduction

Often times we see users having multiple tabs open to search for a particular thing and the search recommendations that are made to them by search engines, are not very specific in nature. We present a new service that integrates search results and relevant posts, and curates a personalized feed for each user.

Motivation

Several informational sources already exist for multiple services online. We envisioned having a single service that could integrate multiple informational and online social services, and refine them based on the user's likes.

Furthermore, as of now searches on one service does not affect recommendations on other services.

Abstract

- Stream relevant data from multiple APIs like MediaWiki (WikiAPI), Twitter (Tweepy), Reddit (praw), Youtube for a user
- Recommend content to user based on various algorithms

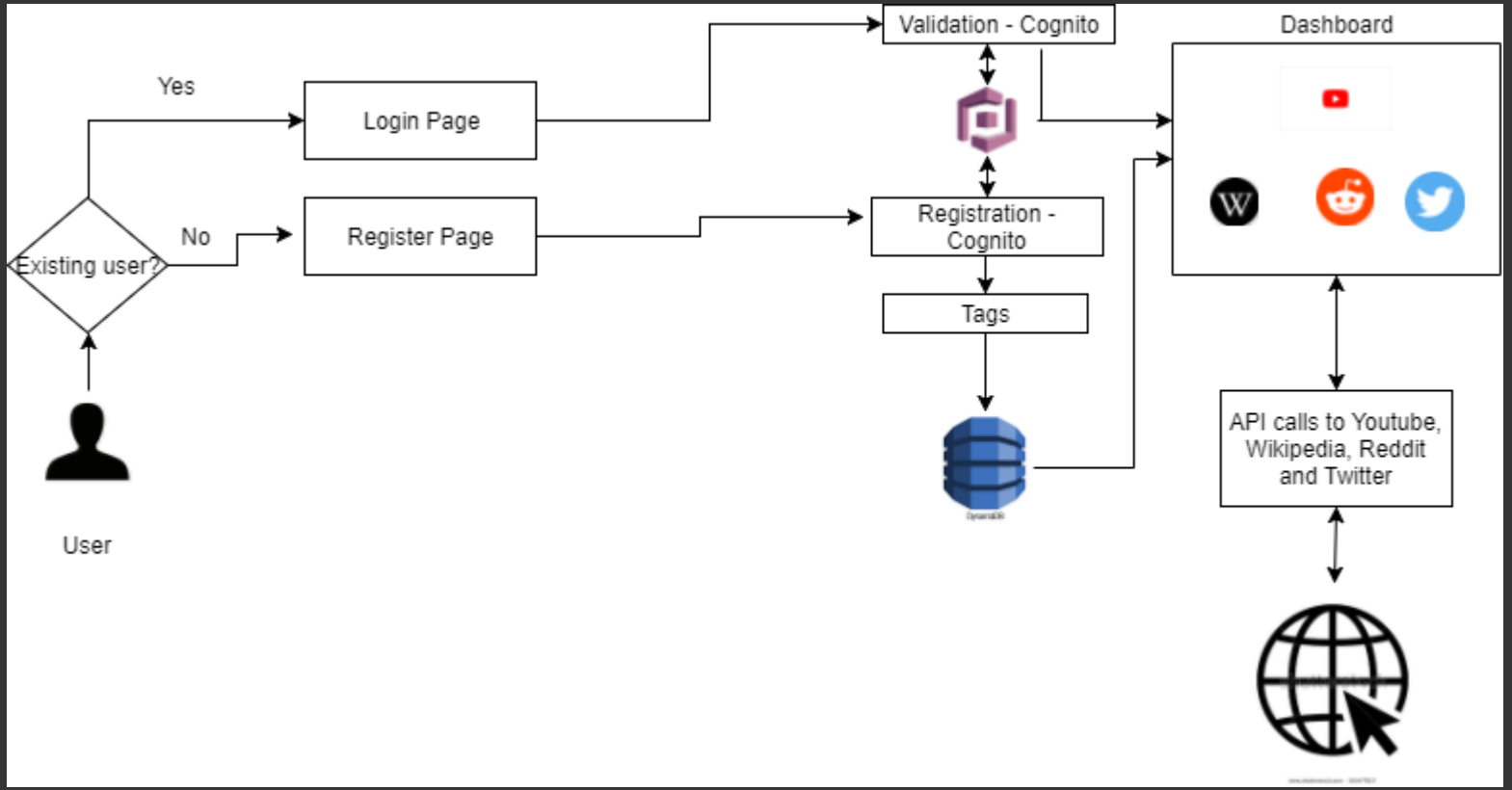
Recommendation Algorithms:

- The first algorithm is based on collaborative filtering.
  - Recommend search tag based on maximum user similarity score.
  - Recommendations are personalised based on the age and gender of the user(demographics).

- The second algorithm is content based.
  - Tags recommended are based on user search history and the frequency of his searches.
  - More weightage is given to recent searches with higher frequency than older searches with lower frequency.

- The third algorithm is based on Apache Spark's MLlib library.
  - We use the well tested ALS (Alternating Least Squares) algorithm to present collaborative filtering recommendations for a user.
  - The model is trained on data we generated based on the 2500 most commonly searched nouns (wordlib) and their relative frequency of occurrence on the web for 17000 distinct users. Total number of entries are ~ 150,000

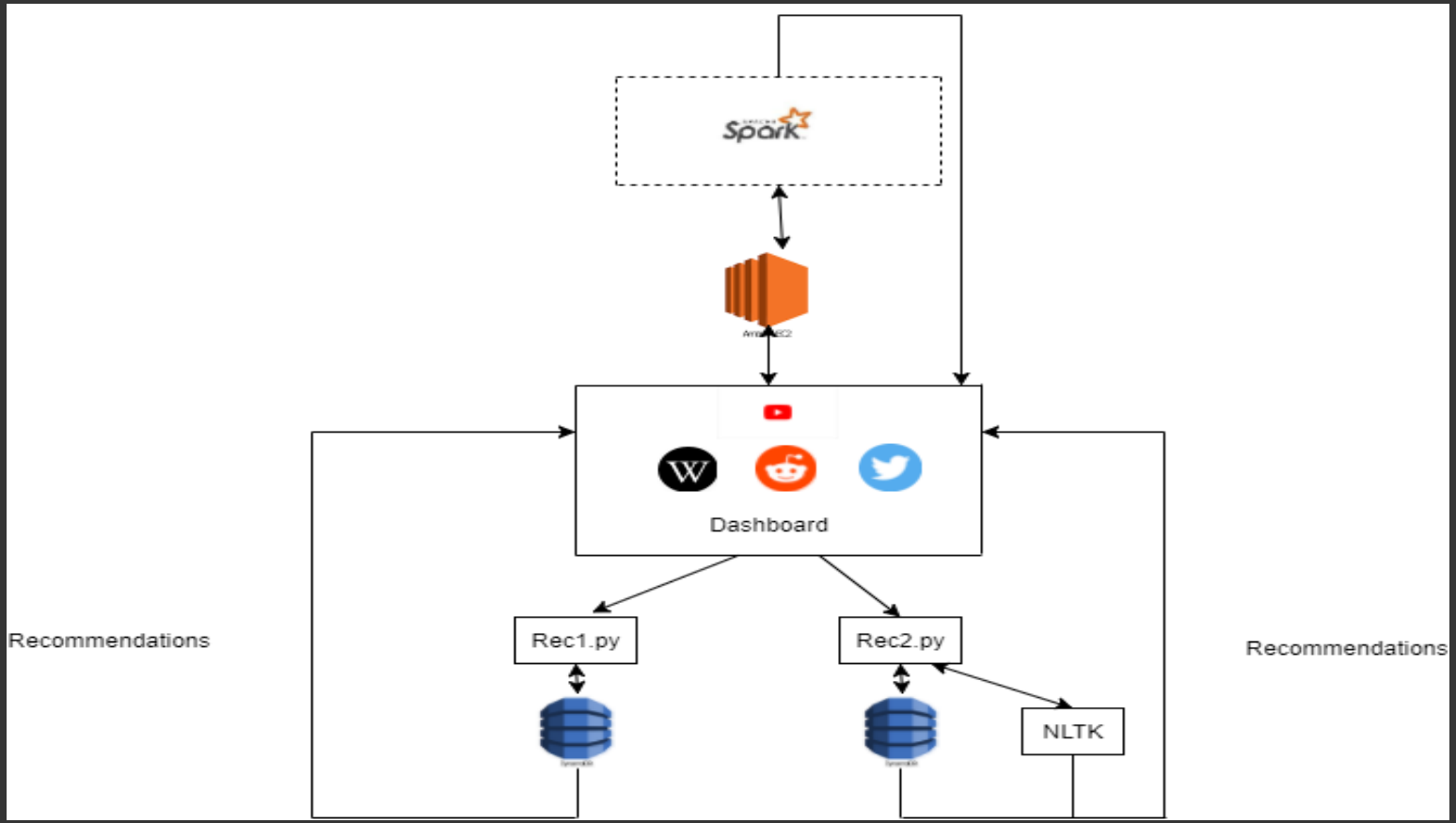
Architecture - I:



Working:

- Amazon Cognito is used for sign-up and log in operations
- The user’s age and gender are captured to assign demographics
- After successfully logging in or signing up, the user is redirected to a dashboard which has recommended feeds
- The user also has an option to select the services that he/she can view
- Moreover, the user can give feedback for the search that was made to improve the recommendation

Architecture - II:



Future Scope:

- Integrating other APIs like News, Yelp, Pinterest, Instagram etc.
- Cache search result data to ElasticSearch for presenting results faster
- Learning to train Spark based recommendation algorithm for every user that signs up
- Developing a hybrid model that captures the best of both, collaborative and content-based recommendation

THANK YOU