

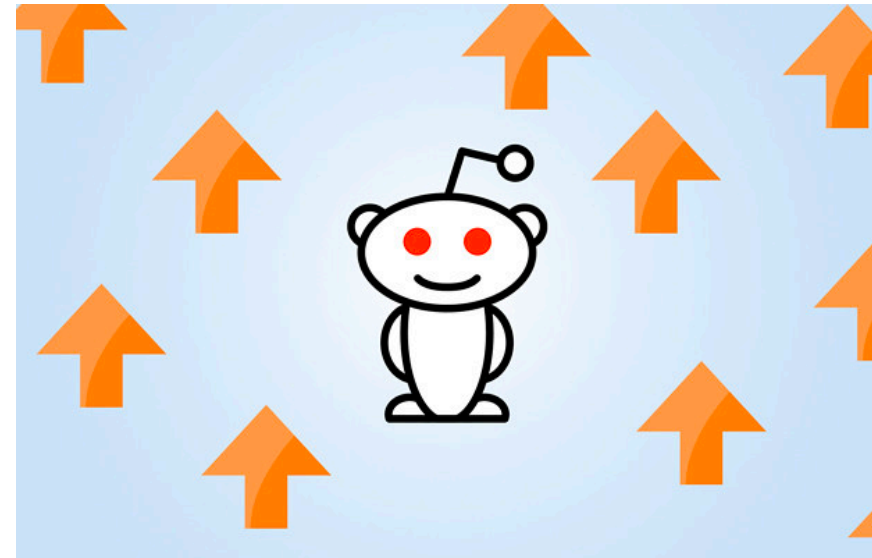


Reddit Recommendation System

James Dudley
Joseph Hamilton
Prasadi Patabadhige

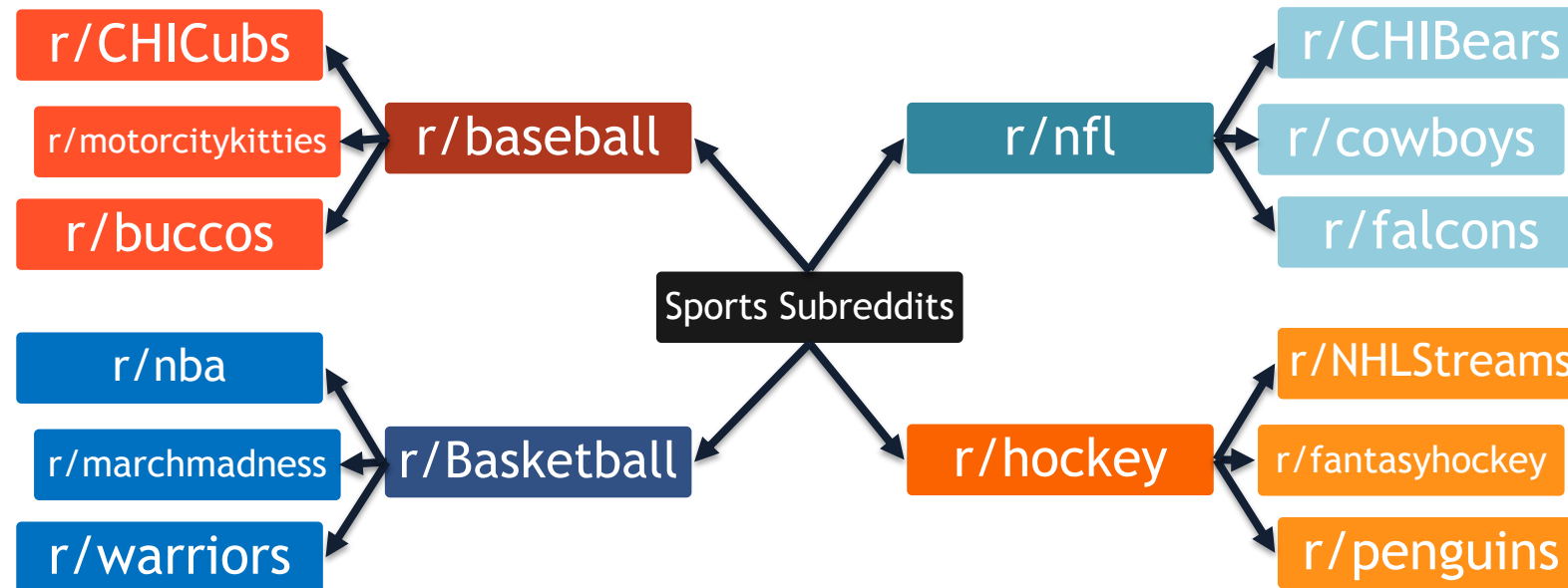
Reddit Background

- Social media and news aggregation site
- Users share various forms of content including links, photos, videos, text posts, etc.
- Popular posts gain traction through accumulating upvotes and comments
- Content is broken down into sections called 'subreddits' like the following:
 - r/python
 - r/todayilearned
 - r/UIUC



Subreddit Relationships

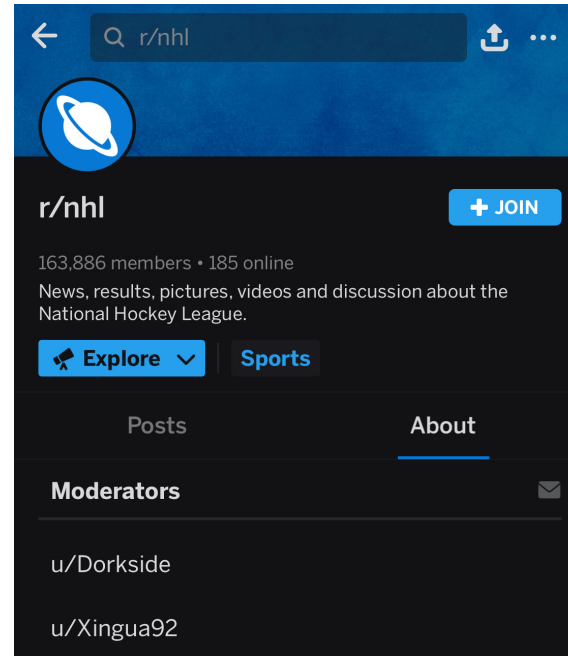
- Communities are broken down into increasingly specific subreddits
- Subreddits are wide ranging and may have unclear names
- Question: how can a user find similar, ongoing discussions on different posts?



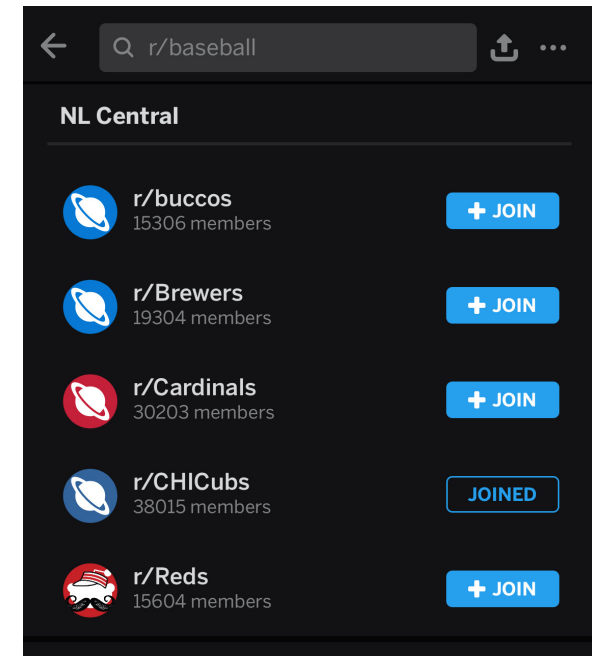
Existing Solutions

- The 'about' section may provide a list of related subreddits
 - Not every subreddit has an 'About' section
 - Reddit App vs. third party App vs. iOS vs. traditional site do not have consistent 'about' sections
 - Related subreddits may not have threads similar to the current one
 - Not all subreddits list related subreddits
- Users cross-post content to related areas

r/nhl 'About' Page



r/baseball 'About' Page



Proposed Solution

- Reddit Recommendation System:
 - Reddit bot 'u/SecureCondorBot' is mentioned in a comment on a thread/post
 - Thread content is analyzed
 - Meanwhile, bot has been pulling popular posts from the front page for content
 - Bot responds back to the user with a similar thread (post) to follow
 - The user can continue similar discussions in new contexts



Architecture

Reddit users generate ongoing discussions via comments

Pushshift.io archives Reddit comments

The PSAW module provides access via the Pushshift API to query or ingest comments

Reddit

Pushshift.io

Index (PSAW)

Inverted Index (Whoosh)

Reply(PRAW)

Analysis

Watch (PRAW)

The recommendation is submitted as a reply comment to the query

Given a query, an analysis is executed to find a similar ongoing discussion

The PRAW module provides access to the Reddit API to watch for queries



Implementation

- The recommender system is comprised of three programs.
- The Indexer
 - Uses PSAW to ingest comments.
 - Uses Whoosh to continually build the inverted index
- The Watcher
 - Uses PRAW to watch for queries directed at the bot.
 - Hands off queries to Responder
- The Responder
 - Uses PRAW to collect the ongoing discussion “around” the query to build the query document.
 - Uses Whoosh, the inverted index and Okapi to recommend a similar ongoing discussion.



Installation



Demo



Further Implementations

- Cloud integration for ease of use
 - EC2 server aggregates and indexes incoming content
 - Lambda function triggered by bot comment mention handles ranking content and returning the most relevant posts
- Reddit bot could link users to related subreddits as well
 - Similar subreddits could be compared for related content and returned to the user

