

CSE 535: Information Retrieval

PROJECT C REPORT

Team: Fantastic Four

Naved Merchant	50168637
Khushbu Mittal	50169099
Prithvi Shah	50169672
Bhavika Jain	50170168

Introduction

In this project, we present a multilingual faceted search system: **Sportify**. As the name suggests, Sportify is a search system for social media data on sports. The user can query the database with a simple and easy-to-use interface and interact with the fetched results in an intuitive manner.

The system scours a database of over 10000 tweets on various sports ranging from Baseball, Basketball, Cricket, Formula 1, NFL, Soccer and Tennis in multiple languages namely English, French, German, Hindi and Spanish. Sportify includes several intelligent features for extracting and presenting results to the user beyond merely the simple query.

User Interface

Sportify has a very clean and simple UI. The homepage directly presents a search bar so that the user can start off querying instantly.

Search Anything about Any Sport!

Go!

The result page is partitioned into a search bar and an analytics option on the top, a filter column on the left, and a scrollable section on the remainder of the page. The tweets are presented in a card style format which are scrollable sideways.

Sportify

Category Filter

soccer

tennis

cricket

basketball

baseball

football

Related Topics

world

cricket

cup

international

twenty20

india

world

Sportsnet

LIVE NOW: @Bundesliga_EN on Sportsnet -- VfL Wolfsburg vs. Borussia Dortmund -- Sportsnet World

Tennis Weblog

Novak Djokovic gewinnt die ATP World Tour Finals 2015
[#tennis](https://t.co/nCyCu603IX)
[#barclaysatpworldtourfinalslondon](https://t.co/cXs2EvTpLz) #andymurray

Cricket News

Tendulkar predicts competitive Cricket World Cup: World Cup winner and ICC Cricket World Cup 2015 Ambassador S...
<http://t.co/cXs2EvTpLz>

International Cricket CouncilCricket World Cup2007 Cricket World Cup

Cricket Zimbabwe

T20 World Cup: The Dutch leave the World T20 with a scalp they can be proud of. <http://t.co/sKj1nzY3qW>

Cricket News

World cycling champions line-up for criterium champs: World champions and Olympic hopefuls will line-up for th...
<https://t.co/5DTTrhrwpoT>

ChampionThe Hopefuls

Cricket Zimbabwe

Bangladesh finding the going tough in the World Cup. Aussie and Bangladesh condemned out of the World Cup...
<http://t.co/3zBui0Unpf>

World cup competitionWorldWorld Cup

The tweet card has the User Name as the header followed by the tweet. Each card can be clicked on to expand and view the additional content such as images or URL link summaries present in the tweet.

Football Burp

NEW: Ryan Bennett apologises to Piers Morgan for pushing Alexis Sánchez into cameraman's pit <https://t.co/WgyCLZuscD>
#ncfc #afc

Football Burp

NEW: Ryan Bennett apologises to Piers Morgan for pushing Alexis Sánchez into cameraman's pit <https://t.co/WgyCLZuscD> #ncfc #afc

Links:

[Ryan Bennett apologises to Piers Morgan Sánchez push](#)

Football Burp uses cookies to ensure you get the best experience on our website. More Info × By Football Burp PUBLISHED: Norwich City's Ryan Bennett apologised to Piers Morgan on Twitter for pushing Arsenal 's Alexis Sánchez into a cameraman's pit during their 1-1 draw at Carrow Road on Sunday.

Football Fever

3 players of the year! Messi, Ronaldo or Neymar Who will win the Balon D'or? #BallondOr2015 #BalondeOro <https://t.co/zMvVNJ3HKI>



The column on the left has two portions. The top portion has a category filter to search within specific sports and the bottom portion consists of topics related to the query.

Framework

Foundation-for-Apps framework is a powerful framework that can be used to make single page web applications. It has various things built in features like cards, and a beautiful animation framework called MotionUI that makes animation of components of the web framework extremely easy.

The following are some features of the Foundation framework:

1. UI elements.

Foundation provides some UI elements that were used in our project. For example: The cards for showing the tweets, and the blocklist used on the sidebar. Both of these are UI elements that come with foundation for Apps framework, and can be customized using the Sass mixins provided to us.

2. State Based Flow

Foundation provides an easy way for making states and providing navigation between them. For example in our project we used 2 states, a home state for the welcome screen and the other called the search state that takes a parameter to take the item for the search.

3. Animation

Foundation also provides animations for the UI elements. We can apply an animation for the state transitions, so that when a state enters or exits, it shows a nice animation. Animation can also be applied to other elements such as cards or lists to give a smoother effect to the UI.

4. Block Lists

Block Lists are a component that can be provided to show various types of lists. It can be shown with check boxes and various other components.

5. Modals

Modals are dialog boxes that can be used to show dialog boxes on top of the current UI. We have used dialog boxes for showing a loading screen as well as showing more information about the tweet.

Interaction with the Backend

The backend is a Solr instance running on an AWS server, which is accessible via http. The interaction has some problems like cross origin resource sharing (CORS), where the app could not access the database since it was hosted in a different place than the current server. This was resolved by allowing CORS by configuring it in solrconfig.xml.

The interaction was as simple as an AJAX request, and it is made even simpler using Angular JS, since it is used by our application.

The \$http service provided by angular lets us perform HTTP GET and POST requests, that can send the search request to the Solr instance. We had to set it a GET type request and also added wt=JSON so that we could process the JSON natively in javascript. The \$http.get() method returns a promise that can be used to get the response of the query.

Content Tagging

In order to highlight key topics or important entities within a result card, we use the Alchemy concept tagging API. The service utilizes sophisticated NLP techniques to analyze the data and tag information in a manner similar to human-based tagging. AlchemyAPI's concept tagging API is capable of abstraction, understanding how concepts relate and tagging accordingly (e.g. "My favorite brands are BMW, Ferrari, and Porsche." = "Automotive Industry").

The tags appear in green labels at the bottom of each result card (subject to them having any relevant content to be tagged). Each of these tags are clickable and fire a query to fetch all the results pertaining to that particular tag. In addition to this, we have provided a special syntax to explicitly search for a content tag directly from the search bar using: "!tag", where 'tag' is the particular content tag to be searched.

Tennis Headlines
Reuters - UK - Tennis: WRAPUP 1-Tennis-Inspired Murray leads Britain to Davis Cup title
Tennis Davis Cup United Kingdom

Faceted Search

Category Filters: The indexed data has been categorized on the basis of sport type. After performing the search operation, our system fetches the category from the retrieved results and dynamically displays the categories on the left side of the UI. The user is provided with the feature of further filtering the result set by selecting one or more categories from the panel. By de-selecting the categories, previously fetched results can be restored.

Category Filter	
cricket	✓
soccer	
football	
tennis	
baseball	✓
basketball	

San Diego Padres
Congrats to @GoAztecs football on making it to the @MountainWest Conference Championship! Date change to #BWBF: <https://t.co/H34ce2XTVO>
The Football League White Pony Deftones

New York Yankees
#Yankees season ticket holders got to play football at The

Related Topics: The lower left panel in the user interface corresponds to the topics that can be of great significance to the users. These topics are high frequency topics appearing in the result set that may be closely related to queried term. When a user clicks on any of those terms, the search box gets populated with the corresponding term and a new search operation is performed.

Related Topics	Stadium at the Legacy Club Open House. #MyYankeesLegacy https://t.co/QefoWBv0fw American football
football	
american	
league	
national	Cricket News Time change for Wellington Phoenix match this weekend:

Ranking & Cross Lingual IR

For processing Search Queries, the primary Request Handler provided with Solr is "SearchHandler". We have created a custom handler: "trinity", for which the default values for query parameters have been specified in solrconfig.xml. We use the phrase slop and minimal match parameters to score and rank the documents and display them in order of relevance to the query. The search model plus the inclusion of content tags and hashtags in the queried fields enables us to retrieve results across multiple languages for a particular query.

manyatomer #Cricket News Hindi सीएम के जोर पर हुआ था सम्मान समारोह: मुद्रल: पूर्व न्यायमूर्ति मुक... https://t.co/PQzaiXIJJT https://t.co/2UI7ms2lxJ	ESPN Cricinfo Newlands track likely to favour batsmen, says curator http://goo.gl/fb/bZjYn #cricket
manyatomer #Cricket News Hindi रैना संभालेंगे विजय हजारे ट्रॉफी में उग्र की कमान: भारतीय वनडे टीम ... https://t.co/61Sz8ZJb5N https://t.co/2UI7ms2lxJ	ESPN Cricinfo The Surfer: Give Clarke a chance http://goo.gl/fb/1dfTB #cricket Arthur C. Clarke Award 2008 deaths
ESPN Cricinfo Hameed apologises for NOTW sting http://goo.gl/fb/qp0GL #cricket	ESPN Cricinfo IPL cuts 56 players from auction list http://goo.gl/fb/ql1QI #cricket 2000s American television series

Summarization

Query Summarization:

We have implemented summarization in our system for the given query using excerpts from Wikipedia articles. To identify a topic for generating a summary for a query, we used a Natural Language Processing library **-nlp_compromise** that analyses the tweet and provides a named entity. Now, using the MediaWiki API, we generate and display a summary for the derived entity at the beginning of the search results.

Q⚙

Sachin Tendulkar

Sachin Ramesh Tendulkar (/ˌsəʃɪn tənˈduːlkər/; born 24 April 1973) is a former Indian cricketer and captain, widely regarded to be the greatest cricketer of all time. He took up cricket at the age of eleven, made his Test debut on 15 November 1989 against Pakistan in Karachi at the age of sixteen, and went on to represent Mumbai domestically and India internationally for close to twenty-four years. He is the only player to have scored one hundred international centuries, the first batsman to score a double century in a One Day International, holds the record for most number of runs in both ODI and Test cricket, the only player to complete more than 30,000 runs in international cricket. In 2002 just half way through his career, Wisden Cricketers' Almanack ranked him the second greatest Test batsman of all time, behind Don Bradman, and the second greatest ODI batsman of all time, behind Viv Richards. Later in his career, Tendulkar was a part of the Indian team that won the 2011 World Cup, his first win in six World Cup appearances for India. He had previously been named "Player of the Tournament" at the 2003 edition of the tournament, held in South Africa. In 2013, he was the only Indian cricketer included in an all-time Test World XI named to mark the 150th anniversary of Wisden Cricketers' Almanack. Tendulkar received the Arjuna Award in 1994 for his outstanding sporting achievement, the Rajiv Gandhi Khel Ratna award in 1997, India's highest sporting honour, and the Padma Sh...

Full Cricket

#cricket New Style Case Cover Sachin Tendulkar God Of Cricket Compatible With Iphone ... <https://t.co/P8Dou1t3Hc>
<https://t.co/wjJ5HqlifU>

Cricket World CupSachin Tendulkar

Lobo

Que grande que #Gary #Neville pego ser él técnico del @valenciafc del @ManUtd Great #GaryNeville is the new coach of #ValenciaFCB #ManUtd

Blue Is The Color

Atleti to win the league this season :). Atleti va a ganar la liga esta temporada #GoAtleti #AupaAtleti @atletienglish @Atleti

Summarizing Information from Embedded Links:

Sportify allows users to view important information corresponding to the links embedded in any tweet. Embed.ly's Extract API has been used to summarize the contents of the embedded url. Now, when a user selects any result card, he can view the summarized url content along with the tweet-text.

R. Premadasa Stadium

×

Sharing dressing room with Tendulkar my finest moment: Ramesh Powar at <https://t.co/LfSwJID0kq> #Cricket #IndvsSA #PakvEng #SLvWI #AusvA

Links:

[Sharing dressing room with Tendulkar my finest moment: Ramesh Powar](#)

It is rare in Indian cricket these days for a player to speak openly about his feelings. Either players toe the diplomatic line to keep the Board of Control for Cricket in India (BCCI) happy or some go overboard with criticism. Ramesh Powar, playing his final first class match at Wankhede, offers a rare exception.

Graphical Analysis

Analysis of the retrieved results has been performed on the basis of languages. The count of retrieved tweets in different languages helps in analyzing the demographics. This can tell us about the popularity of any sport or a team in certain part of the world. For example, high number of Hindi tweets on Cricket can signify the popularity of Cricket in India. On clicking the button at the top-right corner of the UI, a panel displaying the charts pops-up.

