NLP PROJECT FINAL REPORT

TEAM IRUN

텍스트, 장치, 게이지이(가) 표시된 사진

자동 생성된 설명

20172864 서정현

20172848 정석우

20170364 한태균

TABLE OF CONTENTS

1. PROJECT OUTLINE

1-a) What is Irun?

1. LOGIC
2. DATASET

3-a) Tweets

3-b) Match

1. SOURSE CODE & MODULES

4-a) Api\_football\_request module

4-b) VSM module

4-c) IRun module

4-d) Server\_main module

1. RESULT & EVALUATION

5-a) Interval Version

PROJECT OUTLINE

What is Irun?

Irun is a model that predicts the player’s score based on the tweets mentioned by a particular soccer player. While the existing scoring system was calculated through objective indicators such as attack points and defense in the on-the-ball situation, the Iron Model calculates scores of collective intelligence that reflect the thoughts of the anonymous group. In this respect, we will be able to present a new rating standard and show a rating on the overall game as well as the on-the-ball situation.

LOGIC

For each football match, all twitter replies mentioned by each player are collected and classified into 6 sections according to the player's rating. (0.0-6.5, 6.5-7.0, 7.0-7.5, 7.5-8.0, 8.0-8.5, 8.5-10.0) Similarly, the query collects twitter replies from a match in which a player was mentioned. Construct a Vector Space Model according to the TF-IDF value and calculate the cosine similarity between each document and query. Each calculated cosine similarity is normalized and multiplied by the weight of each document to derive the IRun Rating.

DATASET

Tweets

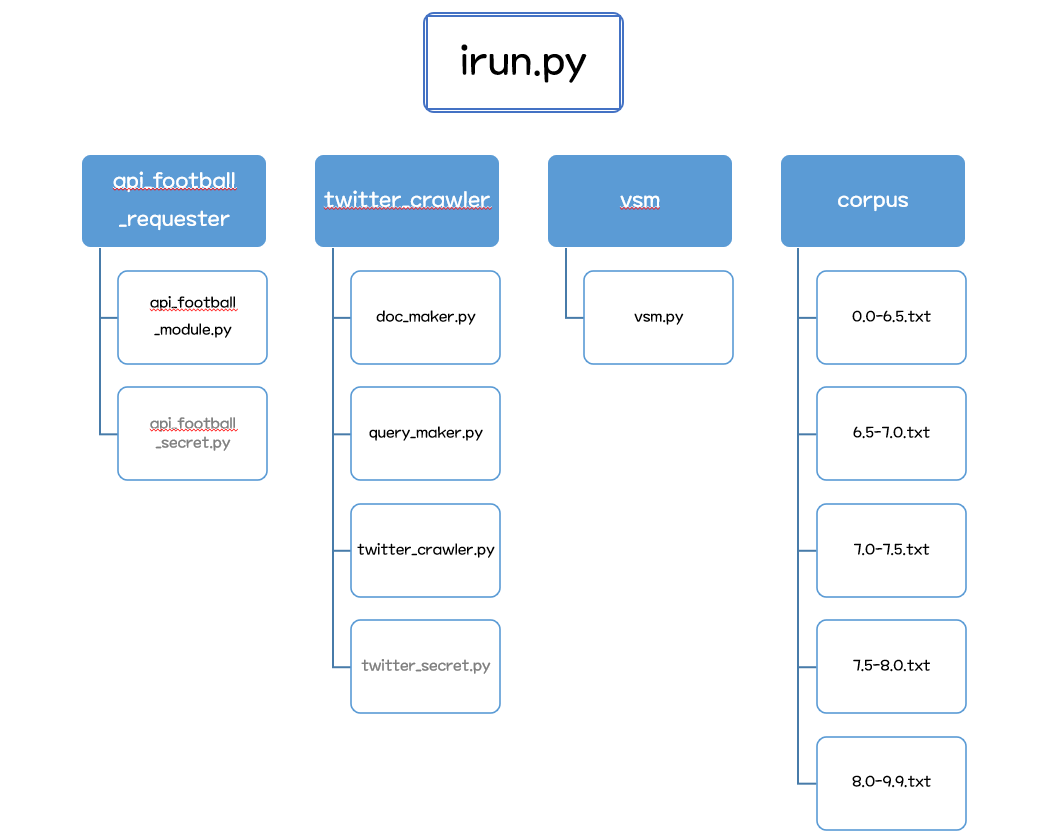
테이블이(가) 표시된 사진

자동 생성된 설명

Match

1. England Premier league
2. Germany Bundesliga
3. Spain La Liga
4. Italy SERIE A
5. USA MLS

SOURSE CODE & MODULES



Api\_football\_request module

Api\_football\_request module collects fixture data by using Football-API at RAPIDAPI. Because twitter api supports only collecting tweets within recent 7 days, we have decided to collect each 7 days’ fixture data, which includes players’ rating and full name. These information is used to sort tweets by players’ rating.

Twitter crawler module

twitter\_cralwer.py implements a get\_replies function that returns all replies related to player name and match date input through the tweepy api. doc\_maker.py and query\_maker.py can compose a corpus or query using this function.

VSM module

It is a module that reads corpus documents and stopwords to create a tf-idf dictionary. The lnc model is applied and dictionaries are generated in advance when imported from the irun module. When an input query comes in, a query vector is generated, and a cosine similarity is obtained by integrating the document vector with l2-norm. Returns the cosine similarity.

IRun module

As the top-level module, the Irun Module parses the query file, sends it to the vsm module, and obtains cosine similarity. The result is calculated with a specific score that is the last result through the result converter module.

Server\_main module

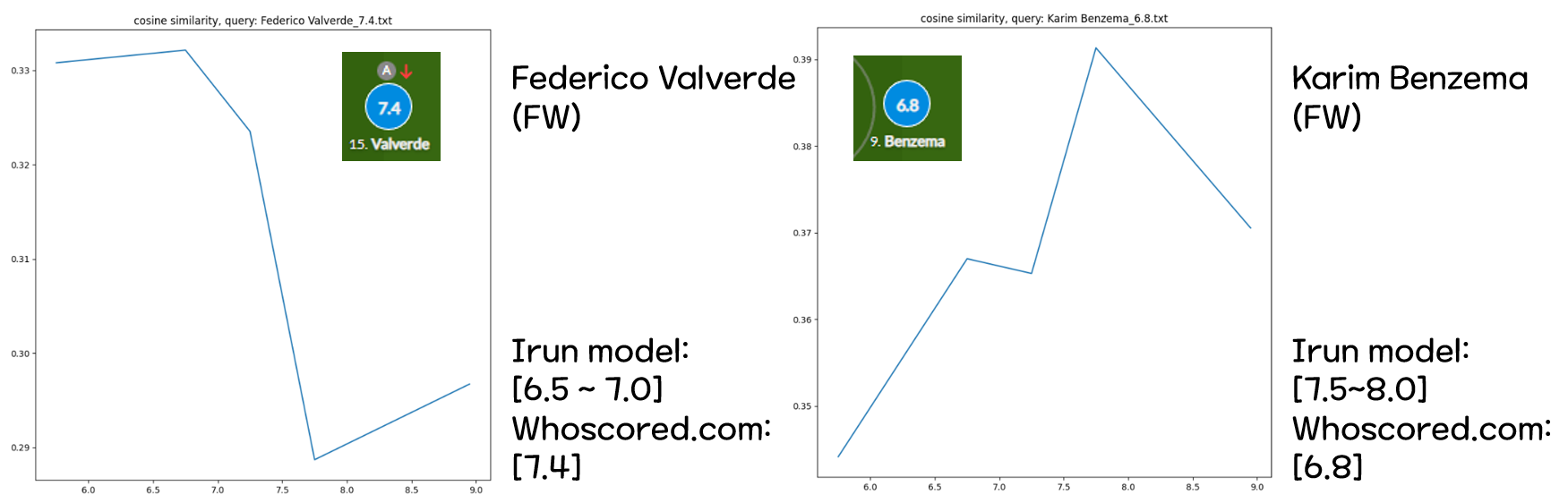
텍스트이(가) 표시된 사진

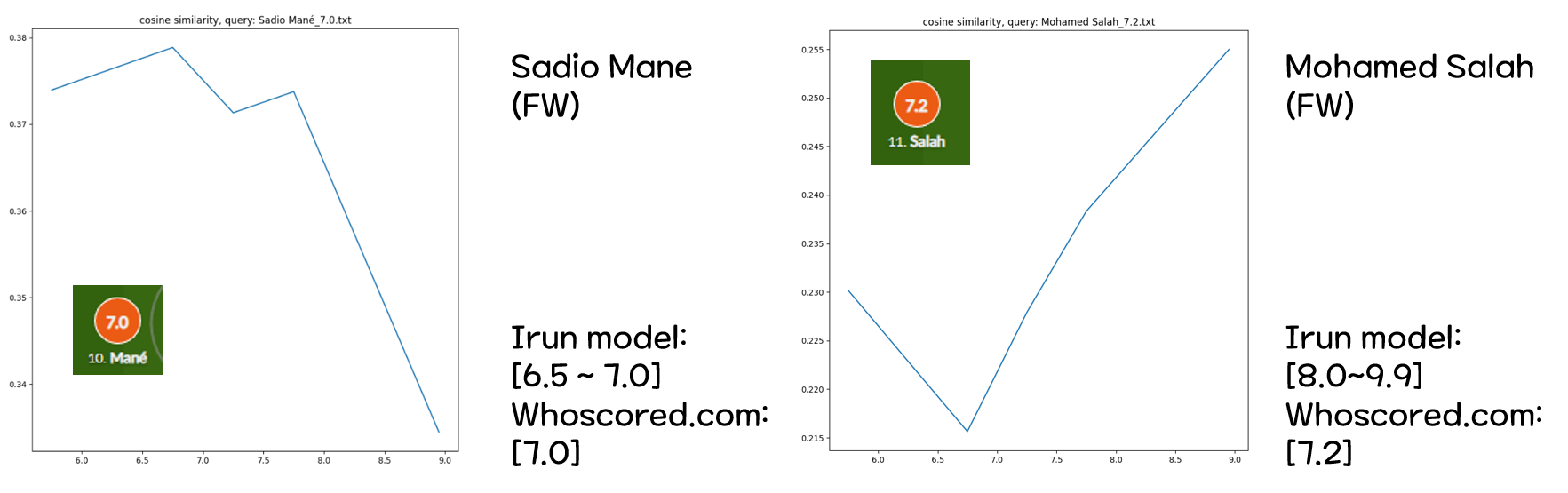
자동 생성된 설명

Irun project aims to serve irun rating to users, which represents tweeter’s reaction to the players’ play and recent trends. To serve rating easily and comfortably to users, we made a simple web api to serve IRun rating. We have used python’s Flask framework to develop api. Irun rating can be made with HTTP Post request to /api/rating with body data that includes query and player’s name. Example query input could be like this one: “His pass is the worst. Even my 5-year-old child can pass well than him”. Then Irun server module returns with expected player’s rating range and specific rating.

Result & EVALUATION

With graph.





MAP Analysis

테이블이(가) 표시된 사진

자동 생성된 설명

Calculated each query into irun rating module and got irun rating rank. Then we compared with actual rating from whoscored.com and futmob.com. We set the highest and idel case value of MAP to 0.2, which means irun module returned relevant (exact match with other rating websites) document as 1st document. There were two ideal case that was close to futmob.com’s rating, but not at whoscored.com rating overally.