Project name : **Premier League Scout**

Team members : **Zhihao Zheng**
Meiyue Song
Kemi Peng
Junchao Kang

Football, also known as soccer, which played by over 200 million players from more than 200 different countries, is the world's most popular sport. There are around 3.5 billion fans all around the Europe, Africa, Asia, Americas, etc.

Though football fans can easily get any specific information about matches, players or clubs by googling the query, most of the results will focus on the most recent news and it's hard to summarize or group these unstructured data. For instance, it's easy to search for the key player in the most recent match of Manchester United FC vs Liverpool FC , however, it's hard to get all the key players in every match between Manchester United FC and Liverpool FC in history. Most football fans enjoy predictions and the latter kind of query is important to fans to predict whether the team they support will have a big win or face a struggling time and hope Pele the 'jinx' is not on your side.

In this project, we'll focus on converting unstructured data (news, reports, wiki pages, etc) into structured data and processing queries about Barclays Premier League (English Premier League).

We'll grab basic data from wikipedia and more data from related web pages (e.g. www.skysports.com). First, we'll write a web crawler to get related web pages. Since they unstructured natural language data, we'll apply Information Extraction methods to raw web pages to extract certain information we want and store them as structured data for further processing.

There're three basic types queries we'll work on. The first is player based query which can get players who meet certain requirements from users (age, position, nationality, etc). The second is club based query which can search certain clubs using founded date, nickname, winning percentage and etc. The third part is match based queries which can search certain match information of clubs given club names, referee name, player names, etc. Having these basic types of queries, we can process most of the query which meet users requirements and can let users get information which google can't help them as is said earlier.