```python
import requests
from bs4 import BeautifulSoup

Paragraphs have been saved to paragraphs.txt


# URL of the webpage to scrape
url = "https://www.footballhistory.org/"

# Output file name
output_file = "paragraphs.txt"



try:
    # Send a GET request to the webpage
    response = requests.get(url)
    response.raise_for_status()  # Raise an error for bad responses

    # Parse the HTML content using BeautifulSoup
    soup = BeautifulSoup(response.text, 'html.parser')

    # Find all paragraph elements
    paragraphs = soup.find_all('p')

    # Open the file for writing
    with open(output_file, "w", encoding="utf-8") as file:
        for p in paragraphs:
            text = p.get_text(strip=True)
            file.write(f"{text}\n\n")

    print(f"Paragraphs have been saved to {output_file}")
except requests.exceptions.RequestException as e:
    print(f"Error fetching the webpage: {e}")
except Exception as e:
    print(f"An error occurred: {e}")
```

```python
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt
```

```python
path = "../WebScraping/paragraphs.txt"
with open(path, 'r', encoding='utf-8') as file:
    football_history = file.read()
```

```python
football_history
```

"\n\nWelcome to FootballHistory.org, a website about football history including competitions, teams and players.\n\nFootball (or soccer as the game is called in some parts of the world) has a long history. Football in its current form arose in England in the middle of the 19th century. But alternative versions of the game existed much earlier and are a part of the football history.\n\nThe first known examples of a team game involving a ball, which was made out of a rock, occurred in old Mesoamerican cultures for over 3,000 years ago. It was by the Aztecs calledTchatali, although various versions of the game were spread over large regions. In some ritual occasions, the ball would symbolize the sun and the captain of the losing team would be sacrificed to the gods. A unique feature of the Mesoamerican ball game versions was a bouncing ball made of rubber â\x80\x93 no other early culture had access to rubber.The first known ball game which also involved kicking took place In China in the 3rd and 2nd century BC under the namecuju. Cuju was played with a round ball (stitched leather with fur or feathers inside) on an area of a square. A modified form of this game later spread to Japan and was by the name ofkemaripracticed under ceremonial forms.Perhaps even older than cuju wasMarn Gook, played by Aboriginal Australians and according to white emigrants in the 1800s,   a ball game that primarily involving kicking. The ball was made by encased leaves or roots. The rules are mostly unknown, but as with many other early versions of the game, a chief feature seems to have been to keep the ball in the air.Other variety of ball games had been known from Ancient Greece. The ball was made by shreds of leather filled with hair (the first documents of balls filledÂ\xa0with air are from the 7th century). Ball games had, however, a low status and was not included at the Panhellenic Games. In the Ancient Rome, games with balls were not included in the entertainment on the big arenas (amphitheaters), but occurred in exercises in the military by the name ofHarpastum. It was the Roman culture that would bring football to the British island (Britannica). It is, however, uncertain in which degree the British people were influenced by this variety and in which degree they had developed their own variants.\n\nThe most admitted story tells that the game was developed in England in the 12th century. In this century, games that resembled football were played on meadows and roads in England. Besides from kicks, the game involved also punches of the ball with the fist. This early form of football was also much more rough and violent than the modern way of

playing.An important feature of the forerunners to football was that the games involved plenty of people and took place over large areas in towns (an equivalent was played in Florence from the 16th century where it was calledCalcio). The rampage of these games would cause damage on the town and sometimes death to the participants. These would be among the reasons for the proclamations against the game that finally was forbidden for several centuries. But the football-like games would return to the streets of London in the 17th century. It would be forbidden again in 1835, but at this stage the game had been established in the public schools.\n\nIt took, however, long time until the features of todayâ\x80\x99s football had been taken into practice. For a long time there was no clear distinction between football and rugby. There were also many variations concerning the size of the ball, the number of players and the length of a match.\n\nThe game was often played in schools and two of the predominant schools were Rugby and Eton. At Rugby the rules included the possibility to take up the ball with the hands and the game we today know as rugby has its origin from here. At Eton on the other hand the ball was played exclusively with the feet and this game can be seen as a close predecessor to the modern football. The game in Rugby was called â\x80\x9cthe running gameâ\x80\x9d while the game in Eton was called â\x80\x9cthe dribbling gameâ\x80\x9d.\n\nAn attempt to create proper rules for the game was done at a meeting in Cambridge in 1848, but a finalÂ\xa0solutionÂ\xa0to all questions of rules was not achieved. Another important event in the history of football came about in 1863 in London when the first Football association was formed in England. It was decided that carrying the ball with the hands wasn't allowed. The meeting alsoÂ\xa0resulted in a standardization of the size and weight of the ball. A consequence of the London meeting was that the game was divided into two codes: association football and rugby.The game would, however, continue to develop for a long time and there was still much flexibility concerning the rules. For one thing, the number of players on the pitch could vary. Neither were uniforms used to distinguish the appearance of the teams. It was also common with players wearing caps â\x80\x93 the header was yet to be a part of the game yet. Further reading:The development of football rules.Another important difference at this stage could be noticed between English and Scottish teams. Whereas the English teams preferred to run forward with the ball in a more rugby fashion, the Scottish chose to pass the ball between their players. It would be the Scottish approach that soon became predominant.The sport was at first an entertainment for the British working class. Unprecedented amounts of spectators, up to 30,000, would see the big matches in the late 19th century. The game would soon expand by British peoples who traveled to other parts of the world and as a result to the British colonization efforts. Especially in South America and India would the interest in football become big.\n\nFootball clubs have existed since the 15th century, but unorganized and without official status. It is therefore hard to decide which the first football club was. Some historians suggest that it was the Foot-Ball Club formed 1824 in Edinburgh. Early clubs were often formed by former school students and the first of this kind was formed in Sheffield in 1855. The oldest among professional football clubs is the English clubNotts Countythat was formed in 1862 and still exists today.An important step for the emergence of teams was the industrialization that led

to larger groups of people meeting at places such as factories, pubs and churches. Football teams were established in the larger cities and the new railroads could bring them to other cities.In the beginning, football was dominated by public school teams, but later, teams consisting by workers would make up the majority. Another change was successively taking place when some clubs became willing to pay the best players to join their team. This would be the start of a long period of transition, not without friction, in which the game would develop to a professional level.The motivation behind paying players was not only to win more matches. In the 1880s the interest in the game has moved ahead to a level that tickets were sold to the matches. And finally, in 1885 professional football was legalized and three years later theFootball Leaguewas established. During the first season, 12 clubs joined the league, but soon more clubs became interested and the competition would consequently expand into more divisions.For a long time, the British teams would be dominant. After some decades, clubs from Prague, Budapest and Sienna would be the primarily contenders to the British dominance.As with many things in history, women were for a long time excluded from participating in games. It was not before the late 19th century that women started to play football. The first official women's game took place in Inverness in 1888.\n\nOther milestones were now to follow.Football Association Challenge Cup (FA Cup)became the first important competition when it was run in 1871. The following year, a match between two national teams was played for the first time. The match that involved England and Scotland ended 0-0 and was followed by 4,000 people at Hamilton Crescent (the picture shows illustrations from this occasion).\n\nTwelve years later, in 1883, the first international tournament took place and included four national teams: England, Ireland, Scotland and Wales.\n\nFootball was for a long time a British phenomenon, but it gradually spread to other European countries. The first game that took place outside Europe occurred in Argentina in 1867, but it was foreign British workers who were involved and not Argentinean citizens.\n\nThe Fédération Internationale de Football Association (FIFA) was founded in 1904 and a foundation act was signed by representatives from France, Belgium, Denmark, the Netherlands, Spain, Sweden and Switzerland. England and the other British countriesÂ\xa0did not join FIFA from the start, they had invented the game and saw no reason to subordinate to an association. Still, they joined in the following year, but would not partake in the World Cup until 1950.Domestic leagues occurred in many countries. The first was, as already mentioned, the English Football League which was established in 1888. The leagues would by time expand by more divisions, which were based on team performance.In 1908 would football for the first time be included as an official sport in the Olympic Games. Until the first FIFA World Cup was played in 1930, the Olympic Games football tournament would rank as the most prestigious on a national level. Women's football was not added until 1996.\n\nAs in many other sports the white male was predominant for a long time. In football black players started being present relatively early and in comparison with, for example, tennis, football has traditionally been known as a sport with a mix of black and white players.\n\nIn Britain,Andrew Watsonis known to be the first black player, and he played in the Scottish club Queenâ\x80\x99s Park in the 1880s.\n\nFew other sports show examples of passion to that extent as football. The arenas

are flocked by shearing people; and in front of television even more are watching carefully and sometimes with great enthusiasm.\n\nAlready in the late 19th century, Goodison Park was built in England in purpose of hosting football games. In 1894, the FA Cup final between Notts County and Bolton Wanderers was attended by 37,000 people. A milestone in the development of football stadiums is the construction of MaracanÃ£ Stadium. In the year of 1950 the imposing stadium in Rio de Janeiro was ready for almost 200,000 people. No other sport has seen stadiums of that capacity built to host its games.There have been two different traditions of fan culture on the arenas: the British and the South American. The British fans adopted the tradition of singing, the repertoire was inspired from pub and working songs among other areas. The South Americans on the other hand would adopt the carnival style which included firecrackers and fireworks, and also the modern phenomena of Bengali fires. Fans in other countries have later adopted a mixture of these traditions.\n\nNo other sport event besides the Summer Olympic Games can today measure itself with theFIFA World Cup. The first edition of the FIFA World Cup was played in 1930 in Uruguay and has since then returned every fourth year (with two exceptions due to the Second World War). In 1991 the first World Cup for women was held in China and has since then also returned every fourth year.Today the biggest global tournament for clubs is theChampions League(played since 1992), the formerEuropean Cup(1955â\x80\x931991).\n\nIn the late 19th century, only a few national football teams existed; England and Scotland had the first active teams that played games against each other in the 1870s. Today there are 211 national associations included in the FÃ©dÃ©ration Internationale de Football Association (FIFA), the world governing body of the sport. Another proof of the globalization could be seen in the increase of nations participating in the World Cup qualifiers: from 32 in 1934 to over 200 in 2014.The world regions have been divided into six confederations: ConfÃ©dÃ©ration Africaine de Football (CAF), Asian Football Confederation (AFC), Union des Associations EuropÃ©ennes de Football (UEFA), The Confederation of North, Central America and Caribbean Association Football (CONCACAF), Oceania Football Confederation (OFC), and ConfederaciÃ³n Sudamericana de FÃºtbol (CONMEBOL).Football is definitely a global sport and without comparison the biggest in the world. A quote from David Goldblatt's bookThe Ball is Roundpresent one answer:\n\nIt offers the spotlight for individual brilliance while relishing the defiance and heart of collective endeavor. It has staged tragedy and comedy, epic and pantomime, unsophisticated music hall and inaccessible experimental performances. It does imperious triumph, lucky escapes, impossible comebacks and stubborn stalemates. It captures the brilliance of unpredictability, the uncertainty of the human heart and human skill, of improvisation and chance.\n\nIn most parts of the world,footballis used as the name for the "chess of the green pitch", the biggest sport in the world. In the United States and Canada, however,socceris used instead as a distinction from American football. A more formal name sometimes used is association football, but in popular speech, it is either football or soccer.\n\n›Formations and playing systems in football›The Evolution of Football Shoes›The football field and its dimensions›History of football stadiums›The name of football in various languages›Collection of links to other sites about football on the web\n\nReferences:The National

EncyclopediaHistory of Football: The Beautiful Game(2002 Documentary Series)The Ball is Round: A Global History of Footballâ\x80\x93 David Goldblatt (2008)http://www.fifa.com/about-fifa/who-we-are/the-game/https://en.wikipedia.org/wiki/Oldest_football_clubshttp://www.fifa.com/about-fifa/who-we-are/history/http://spartacus-educational.com/Fblack.htmhttp://spartacus-educational.com/Fstadiums.htmhttp://www.fifa.com/associations/Image sources:Andrea Scoto â\x80\x93 ItiIllustration da Francesco. Bertelli. PaduaWilliam Ralston (1848â\x80\x931911) â\x80\x93 Scanned from The Official History of The Football Association by Bryon Butler â\x80\x93 Queen Ane PressDocumentary filmAlegria do Povo(1963)\n\n"

Tokenizing the text document content

```
tokenizer = tf.keras.preprocessing.text.Tokenizer()
tokenizer.fit_on_texts([football_history])

tokenizer.get_config()
```

```
{'num_words': None,
 'filters': '!"#$%&()*+,-./:;<=>?@[\\]^_`{|}~\t\n',
 'lower': True,
 'split': ' ',
 'char_level': False,
 'oov_token': None,
 'document_count': 1,
 'word_counts': '{"welcome": 1, "to": 43, "footballhistory": 1, "org": 2,
"a": 46, "website": 1, "about": 5, "football": 55, "history": 8, "including":
1, "competitions": 1, "teams": 13, "and": 68, "players": 9, "or": 4,
"soccer": 2, "as": 15, "the": 218, "game": 30, "is": 11, "called": 3, "in":
98, "some": 5, "parts": 3, "of": 74, "world": 14, "has": 8, "long": 9, "its":
4, "current": 1, "form": 3, "arose": 1, "england": 9, "middle": 1, "19th": 5,
"century": 12, "but": 13, "alternative": 1, "versions": 4, "existed": 3,
"much": 3, "earlier": 1, "are": 8, "part": 2, "first": 20, "known": 5,
"examples": 2, "team": 4, "involving": 2, "ball": 23, "which": 9, "was": 58,
"made": 4, "out": 1, "rock": 1, "occurred": 4, "old": 1, "mesoamerican": 2,
"cultures": 1, "for": 21, "over": 4, "3": 1, "000": 5, "years": 3, "ago": 1,
"it": 20, "by": 18, "aztecs": 1, "calledtchatali": 1, "although": 1,
"various": 2, "were": 14, "spread": 3, "large": 2, "regions": 2, "ritual": 1,
"occasions": 1, "would": 23, "symbolize": 1, "sun": 1, "captain": 1,
"losing": 1, "be": 13, "sacrificed": 1, "gods": 1, "unique": 1, "feature": 3,
"bouncing": 1, "rubber": 2, "\\u00e2\\u0080\\u0093": 5, "no": 5, "other": 18,
"early": 5, "culture": 3, "had": 8, "access": 1, "also": 7, "involved": 5,
"kicking": 2, "took": 6, "place": 6, "china": 2, "3rd": 1, "2nd": 1, "bc": 1,
"under": 2, "namecuju": 1, "cuju": 2, "played": 12, "with": 17, "round": 2,
"stitched": 1, "leather": 2, "fur": 1, "feathers": 1, "inside": 1, "on": 11,
"an": 8, "area": 1, "square": 1, "modified": 1, "this": 10, "later": 5,
"japan": 1, "name": 5, "ofkemaripracticed": 1, "ceremonial": 1, "forms": 1,
"perhaps": 1, "even": 2, "older": 1, "than": 2, "wasmarn": 1, "gook": 1,
"aboriginal": 1, "australians": 1, "according": 1, "white": 3, "emigrants":
1, "1800s": 1, "that": 18, "primarily": 2, "encased": 1, "leaves": 1,
```

"roots": 1, "rules": 6, "mostly": 1, "unknown": 1, "many": 5, "chief": 1, "seems": 1, "have": 5, "been": 7, "keep": 1, "air": 2, "variety": 2, "games": 15, "from": 15, "ancient": 2, "greece": 1, "shreds": 1, "filled": 1, "hair": 1, "documents": 1, "balls": 2, "filled\\u00e2\\u00a0with": 1, "7th": 1, "however": 5, "low": 1, "status": 2, "not": 10, "included": 7, "at": 9, "panhellenic": 1, "rome": 1, "entertainment": 2, "big": 3, "arenas": 3, "amphitheaters": 1, "exercises": 1, "military": 1, "ofharpastum": 1, "roman": 1, "bring": 2, "british": 12, "island": 1, "britannica": 1, "uncertain": 1, "degree": 2, "people": 7, "influenced": 1, "they": 3, "developed": 2, "their": 3, "own": 1, "variants": 1, "most": 3, "admitted": 1, "story": 1, "tells": 1, "12th": 1, "resembled": 1, "meadows": 1, "roads": 1, "besides": 2, "kicks": 1, "punches": 1, "fist": 1, "more": 8, "rough": 1, "violent": 1, "modern": 3, "way": 1, "playing": 2, "important": 5, "forerunners": 1, "plenty": 1, "areas": 2, "towns": 1, "equivalent": 1, "florence": 1, "16th": 1, "where": 1, "calledcalcio": 1, "rampage": 1, "these": 3, "cause": 1, "damage": 1, "town": 1, "sometimes": 3, "death": 1, "participants": 1, "among": 3, "reasons": 1, "proclamations": 1, "against": 2, "finally": 2, "forbidden": 2, "several": 1, "centuries": 1, "like": 1, "return": 1, "streets": 1, "london": 3, "17th": 1, "again": 1, "1835": 1, "stage": 2, "established": 4, "public": 2, "schools": 3, "time": 10, "until": 4, "features": 1, "today\\u00e2\\u0080\\u0099s": 1, "taken": 1, "into": 4, "practice": 1, "there": 5, "clear": 1, "distinction": 2, "between": 5, "rugby": 7, "variations": 1, "concerning": 2, "size": 2, "number": 2, "length": 1, "match": 3, "often": 2, "two": 5, "predominant": 3, "eton": 3, "possibility": 1, "take": 1, "up": 3, "hands": 2, "we": 3, "today": 5, "know": 1, "origin": 1, "here": 1, "hand": 2, "exclusively": 1, "feet": 1, "can": 2, "seen": 3, "close": 1, "predecessor": 1, "\\u00e2\\u0080\\u009cthe": 2, "running": 1, "game\\u00e2\\u0080\\u009d": 2, "while": 2, "dribbling": 1, "attempt": 1, "create": 1, "proper": 1, "done": 1, "meeting": 4, "cambridge": 1, "1848": 1, "final\\u00e2\\u00a0solution\\u00e2\\u00a0to": 1, "all": 1, "questions": 1, "achieved": 1, "another": 4, "event": 2, "came": 1, "1863": 1, "when": 3, "association": 9, "formed": 5, "decided": 1, "carrying": 1, "wasn\\'t": 1, "allowed": 1, "also\\u00e2\\u00a0resulted": 1, "standardization": 1, "weight": 1, "consequence": 1, "divided": 2, "codes": 1, "continue": 1, "develop": 2, "still": 3, "flexibility": 1, "one": 2, "thing": 1, "pitch": 1, "could": 4, "vary": 1, "neither": 1, "uniforms": 1, "used": 4, "distinguish": 1, "appearance": 1, "common": 1, "wearing": 1, "caps": 1, "header": 1, "yet": 2, "further": 1, "reading": 1, "development": 2, "difference": 1, "noticed": 1, "english": 4, "scottish": 4, "whereas": 1, "preferred": 1, "run": 2, "forward": 1, "fashion": 1, "chose": 1, "pass": 1, "approach": 1, "soon": 3, "became": 4, "sport": 8, "working": 2, "class": 1, "unprecedented": 1, "amounts": 1, "spectators": 1, "30": 1, "see": 1, "matches": 3, "late": 4, "expand": 3, "peoples": 1, "who": 4, "traveled": 1, "result": 1, "colonization": 1, "efforts": 1, "especially": 1, "south": 3, "america": 2, "india": 1, "interest": 2, "become": 1, "clubs": 8, "since": 4, "15th": 1, "unorganized": 1, "without": 3, "official": 4, "therefore": 1, "hard": 1, "decide": 1, "club": 3, "historians": 1, "suggest": 1, "foot": 1, "1824": 1, "edinburgh": 1, "former": 1, "school": 2, "students": 1, "kind": 1, "sheffield": 1, "1855": 1, "oldest": 2, "professional": 3, "clubnotts": 1,

"countythat": 1, "1862": 1, "exists": 1, "step": 1, "emergence": 1, "industrialization": 1, "led": 1, "larger": 2, "groups": 1, "places": 1, "such": 1, "factories": 1, "pubs": 1, "churches": 1, "cities": 2, "new": 1, "railroads": 1, "them": 1, "beginning": 1, "dominated": 1, "consisting": 1, "workers": 2, "make": 1, "majority": 1, "change": 1, "successively": 1, "taking": 1, "willing": 1, "pay": 1, "best": 1, "join": 2, "start": 2, "period": 1, "transition": 1, "friction": 1, "level": 3, "motivation": 1, "behind": 1, "paying": 1, "only": 2, "win": 1, "1880s": 2, "moved": 1, "ahead": 1, "tickets": 1, "sold": 1, "1885": 1, "legalized": 1, "three": 1, "thefootball": 1, "leaguewas": 1, "during": 1, "season": 1, "12": 1, "joined": 2, "league": 3, "interested": 1, "competition": 2, "consequently": 1, "divisions": 2, "dominant": 1, "after": 1, "decades": 1, "prague": 1, "budapest": 1, "sienna": 1, "contenders": 1, "dominance": 1, "things": 1, "women": 3, "excluded": 1, "participating": 2, "before": 1, "started": 2, "play": 1, "women\'s": 2, "inverness": 1, "1888": 2, "milestones": 1, "now": 1, "follow": 1, "challenge": 1, "cup": 10, "fa": 2, "1871": 1, "following": 2, "year": 5, "national": 6, "scotland": 3, "ended": 1, "0": 2, "followed": 1, "4": 1, "hamilton": 1, "crescent": 1, "picture": 1, "shows": 1, "illustrations": 1, "occasion": 1, "twelve": 1, "1883": 1, "international": 1, "tournament": 3, "four": 1, "ireland": 1, "wales": 1, "phenomenon": 1, "gradually": 1, "european": 1, "countries": 3, "outside": 1, "europe": 1, "argentina": 1, "1867": 1, "foreign": 1, "argentinean": 1, "citizens": 1, "f\\u00e3\\u00a9d\\u00e3\\u00a9ration": 2, "internationale": 2, "de": 6, "fifa": 10, "founded": 1, "1904": 1, "foundation": 1, "act": 1, "signed": 1, "representatives": 1, "france": 1, "belgium": 1, "denmark": 1, "netherlands": 1, "spain": 1, "sweden": 1, "switzerland": 1, "countries\\u00e2\\u00a0did": 1, "invented": 1, "saw": 1, "reason": 1, "subordinate": 1, "partake": 1, "1950": 2, "domestic": 1, "leagues": 2, "already": 2, "mentioned": 1, "based": 1, "performance": 1, "1908": 1, "olympic": 3, "1930": 2, "rank": 1, "prestigious": 1, "added": 1, "1996": 1, "sports": 2, "male": 1, "black": 3, "being": 1, "present": 1, "relatively": 1, "comparison": 2, "example": 1, "tennis": 1, "traditionally": 1, "mix": 1, "britain": 1, "andrew": 1, "watsonis": 1, "player": 1, "he": 1, "queen\\u00e2\\u0080\\u0099s": 1, "park": 2, "few": 2, "show": 1, "passion": 1, "extent": 1, "flocked": 1, "shearing": 1, "front": 1, "television": 1, "watching": 1, "carefully": 1, "great": 1, "enthusiasm": 1, "goodison": 1, "built": 2, "purpose": 1, "hosting": 1, "1894": 1, "final": 1, "notts": 1, "county": 1, "bolton": 1, "wanderers": 1, "attended": 1, "37": 1, "milestone": 1, "stadiums": 2, "construction": 1, "maracan\\u00e3\\u00a3": 1, "stadium": 2, "imposing": 1, "rio": 1, "janeiro": 1, "ready": 1, "almost": 1, "200": 2, "capacity": 1, "host": 1, "different": 1, "traditions": 2, "fan": 1, "american": 2, "fans": 2, "adopted": 2, "tradition": 1, "singing": 1, "repertoire": 1, "inspired": 1, "pub": 1, "songs": 1, "americans": 1, "adopt": 1, "carnival": 1, "style": 1, "firecrackers": 1, "fireworks": 1, "phenomena": 1, "bengali": 1, "fires": 1, "mixture": 1, "summer": 1, "measure": 1, "itself": 1, "thefifa": 1, "edition": 1, "uruguay": 1, "then": 2, "returned": 2, "every": 2, "fourth": 2, "exceptions": 1, "due": 1, "second": 1, "war": 1, "1991": 1, "held": 1, "biggest": 3, "global": 3, "thechampions": 1, "1992": 1, "formereuropean": 1, "1955\\u00e2\\u0080\\u00931991": 1, "active": 1, "each": 1, "1870s": 1, "211": 1, "associations": 3, "governing": 1, "body": 1, "proof": 1,

"globalization": 1, "increase": 1, "nations": 1, "qualifiers": 1, "32": 1, "1934": 1, "2014": 1, "six": 1, "confederations": 1, "conf\\u00e3\\u00a9d\\u00e3\\u00a9ration": 1, "africaine": 1, "caf": 1, "asian": 1, "confederation": 3, "afc": 1, "union": 1, "des": 1, "europ\\u00e3\\u00a9ennes": 1, "uefa": 1, "north": 1, "central": 1, "caribbean": 1, "concacaf": 1, "oceania": 1, "ofc": 1, "confederaci\\u00e3\\u00b3n": 1, "sudamericana": 1, "f\\u00e3\\u00babtbol": 1, "conmebol": 1, "definitely": 1, "quote": 1, "david": 2, "goldblatt\'s": 1, "bookthe": 1, "roundpresent": 1, "answer": 1, "offers": 1, "spotlight": 1, "individual": 1, "brilliance": 2, "relishing": 1, "defiance": 1, "heart": 2, "collective": 1, "endeavor": 1, "staged": 1, "tragedy": 1, "comedy": 1, "epic": 1, "pantomime": 1, "unsophisticated": 1, "music": 1, "hall": 1, "inaccessible": 1, "experimental": 1, "performances": 1, "does": 1, "imperious": 1, "triumph": 1, "lucky": 1, "escapes": 1, "impossible": 1, "comebacks": 1, "stubborn": 1, "stalemates": 1, "captures": 1, "unpredictability": 1, "uncertainty": 1, "human": 2, "skill": 1, "improvisation": 1, "chance": 1, "footballis": 1, "\\u201cchess": 1, "green": 1, "pitch\\u201d": 1, "united": 1, "states": 1, "canada": 1, "socceris": 1, "instead": 1, "formal": 1, "popular": 1, "speech": 1, "either": 1, "\\u203aformations": 1, "systems": 1, "football\\u203athe": 1, "evolution": 1, "shoes\\u203athe": 1, "field": 1, "dimensions\\u203ahistory": 1, "stadiums\\u203athe": 1, "languages\\u203acollection": 1, "links": 1, "sites": 1, "web": 1, "references": 1, "encyclopediahistory": 1, "beautiful": 1, "2002": 1, "documentary": 1, "series": 1, "football\\u00e2\\u0080\\u0093": 1, "goldblatt": 1, "2008": 1, "http": 2, "www": 3, "com": 5, "https": 1, "en": 1, "wikipedia": 1, "wiki": 1, "clubshttp": 1, "spartacus": 2, "educational": 2, "fblack": 1, "htmhttp": 2, "fstadiums": 1, "image": 1, "sources": 1, "andrea": 1, "scoto": 1, "itiillustration": 1, "da": 1, "francesco": 1, "bertelli": 1, "paduawilliam": 1, "ralston": 1, "1848\\u00e2\\u0080\\u00931911": 1, "scanned": 1, "bryon": 1, "butler": 1, "queen": 1, "ane": 1, "pressdocumentary": 1, "filmalegria": 1, "do": 1, "povo": 1, "1963": 1}',
 'word_docs': '{"denmark": 1, "island": 1, "taken": 1, "origin": 1, "encyclopediahistory": 1, "white": 1, "htmhttp": 1, "tradition": 1, "cambridge": 1, "three": 1, "amounts": 1, "africaine": 1, "maracan\\u00e3\\u00a3": 1, "television": 1, "paduawilliam": 1, "ceremonial": 1, "dominant": 1, "an": 1, "subordinate": 1, "club": 1, "team": 1, "part": 1, "rules": 1, "illustrations": 1, "football\\u203athe": 1, "sun": 1, "asian": 1, "english": 1, "common": 1, "became": 1, "larger": 1, "confederations": 1, "developed": 1, "sweden": 1, "chance": 1, "beautiful": 1, "rio": 1, "foot": 1, "some": 1, "phenomena": 1, "do": 1, "second": 1, "leaguewas": 1, "first": 1, "definitely": 1, "tickets": 1, "de": 1, "foreign": 1, "active": 1, "2008": 1, "features": 1, "dribbling": 1, "at": 1, "win": 1, "official": 1, "1930": 1, "1824": 1, "hosting": 1, "sienna": 1, "fires": 1, "globalization": 1, "football": 1, "such": 1, "each": 1, "japan": 1, "cup": 1, "either": 1, "especially": 1, "football\\u00e2\\u0080\\u0093": 1, "transition": 1, "forward": 1, "enthusiasm": 1, "17th": 1, "ago": 1, "entertainment": 1, "12": 1, "being": 1, "citizens": 1, "documentary": 1, "matches": 1, "unknown": 1, "1908": 1, "air": 1, "ball": 1, "decades": 1, "tournament": 1, "black": 1, "1871": 1, "world": 1, "show": 1, "reason": 1, "1894": 1, "1996": 1, "india":

1, "milestones": 1, "all": 1, "can": 1, "pitch": 1, "wales": 1, "1992": 1, "europ\\u00e3\\u00a9ennes": 1, "mostly": 1, "answer": 1, "own": 1, "number": 1, "standardization": 1, "excluded": 1, "feathers": 1, "rock": 1, "plenty": 1, "round": 1, "place": 1, "which": 1, "was": 1, "sometimes": 1, "started": 1, "by": 1, "american": 1, "scanned": 1, "spread": 1, "fur": 1, "included": 1, "performance": 1, "2nd": 1, "notts": 1, "extent": 1, "degree": 1, "see": 1, "who": 1, "ready": 1, "whereas": 1, "sports": 1, "singing": 1, "run": 1, "development": 1, "from": 1, "leagues": 1, "belgium": 1, "body": 1, "military": 1, "built": 1, "central": 1, "arose": 1, "parts": 1, "ireland": 1, "canada": 1, "\\u00e2\\u0080\\u0093": 1, "start": 1, "however": 1, "students": 1, "had": 1, "goldblatt\'s": 1, "milestone": 1, "play": 1, "as": 1, "bring": 1, "violent": 1, "arenas": 1, "gods": 1, "uniforms": 1, "traveled": 1, "without": 1, "fa": 1, "in": 1, "variants": 1, "\\u00e2\\u0080\\u009cthe": 1, "phenomenon": 1, "comparison": 1, "caps": 1, "best": 1, "people": 1, "occasion": 1, "australians": 1, "confederaci\\u00e3\\u00b3n": 1, "https": 1, "heart": 1, "exists": 1, "ralston": 1, "colonization": 1, "hamilton": 1, "america": 1, "not": 1, "fans": 1, "led": 1, "scotland": 1, "joined": 1, "public": 1, "also": 1, "nations": 1, "big": 1, "1955\\u00e2\\u0080\\u00931991": 1, "challenge": 1, "1880s": 1, "participants": 1, "even": 1, "access": 1, "edinburgh": 1, "there": 1, "women": 1, "length": 1, "systems": 1, "consequence": 1, "like": 1, "uruguay": 1, "late": 1, "footballhistory": 1, "wearing": 1, "emergence": 1, "war": 1, "other": 1, "change": 1, "impossible": 1, "could": 1, "rugby": 1, "ane": 1, "amphitheaters": 1, "com": 1, "done": 1, "firecrackers": 1, "questions": 1, "union": 1, "moved": 1, "olympic": 1, "decide": 1, "sources": 1, "take": 1, "often": 1, "ended": 1, "2014": 1, "professional": 1, "argentinean": 1, "passion": 1, "distinguish": 1, "clubs": 1, "fstadiums": 1, "including": 1, "32": 1, "quote": 1, "field": 1, "mix": 1, "important": 1, "act": 1, "early": 1, "bookthe": 1, "on": 1, "preferred": 1, "exceptions": 1, "between": 1, "made": 1, "used": 1, "proof": 1, "panhellenic": 1, "reading": 1, "pub": 1, "prestigious": 1, "19th": 1, "sacrificed": 1, "followed": 1, "thing": 1, "flexibility": 1, "outside": 1, "dominated": 1, "where": 1, "with": 1, "eton": 1, "return": 1, "before": 1, "unsophisticated": 1, "britannica": 1, "pay": 1, "countries": 1, "queen\\u00e2\\u0080\\u0099s": 1, "further": 1, "until": 1, "series": 1, "1848": 1, "comebacks": 1, "a": 1, "codes": 1, "later": 1, "1963": 1, "1855": 1, "offers": 1, "stage": 1, "proclamations": 1, "when": 1, "established": 1, "sheffield": 1, "shearing": 1, "towns": 1, "en": 1, "defiance": 1, "greece": 1, "improvisation": 1, "switzerland": 1, "almost": 1, "create": 1, "most": 1, "divided": 1, "thefootball": 1, "3rd": 1, "variety": 1, "been": 1, "janeiro": 1, "taking": 1, "chief": 1, "railroads": 1, "watsonis": 1, "clubshttp": 1, "bengali": 1, "make": 1, "founded": 1, "form": 1, "national": 1, "they": 1, "association": 1, "players": 1, "comedy": 1, "uefa": 1, "wasmarn": 1, "website": 1, "imperious": 1, "have": 1, "historians": 1, "contenders": 1, "tells": 1, "hands": 1, "predecessor": 1, "prague": 1, "0": 1, "stadiums": 1, "london": 1, "teams": 1, "took": 1, "result": 1, "participating": 1, "fan": 1, "fireworks": 1, "lucky": 1, "size": 1, "inspired": 1, "shoes\\u203athe": 1, "pressdocumentary": 1, "1863": 1, "traditions": 1, "stitched": 1, "spotlight": 1, "measure": 1, "kicking": 1, "under": 1, "green": 1, "close": 1, "david": 1, "conf\\u00e3\\u00a9d\\u00e3\\u00a9ration": 1, "primarily": 1,

"become": 1, "evolution": 1, "came": 1, "player": 1, "qualifiers": 1, "name": 1, "then": 1, "wanderers": 1, "adopted": 1, "added": 1, "is": 1, "versions": 1, "roundpresent": 1, "over": 1, "against": 1, "host": 1, "sudamericana": 1, "povo": 1, "encased": 1, "hair": 1, "motivation": 1, "welcome": 1, "unpredictability": 1, "wikipedia": 1, "brilliance": 1, "footballis": 1, "formereuropean": 1, "four": 1, "occurred": 1, "variations": 1, "representatives": 1, "known": 1, "2002": 1, "allowed": 1, "www": 1, "various": 1, "would": 1, "new": 1, "wiki": 1, "attended": 1, "roman": 1, "adopt": 1, "shows": 1, "interested": 1, "exercises": 1, "hand": 1, "regions": 1, "feature": 1, "south": 1, "stadium": 1, "americans": 1, "caf": 1, "besides": 1, "3": 1, "internationale": 1, "inside": 1, "forerunners": 1, "16th": 1, "afc": 1, "no": 1, "while": 1, "saw": 1, "carefully": 1, "itiillustration": 1, "goldblatt": 1, "them": 1, "construction": 1, "captures": 1, "than": 1, "queen": 1, "friction": 1, "filled\\u00e2\\u00a0with": 1, "shreds": 1, "ofc": 1, "seen": 1, "bouncing": 1, "relishing": 1, "educational": 1, "current": 1, "associations": 1, "37": 1, "that": 1, "calledcalcio": 1, "spain": 1, "way": 1, "long": 1, "roots": 1, "keep": 1, "interest": 1, "after": 1, "flocked": 1, "different": 1, "thechampions": 1, "much": 1, "playing": 1, "soccer": 1, "bc": 1, "andrea": 1, "due": 1, "image": 1, "hard": 1, "its": 1, "summer": 1, "earlier": 1, "admitted": 1, "florence": 1, "know": 1, "leaves": 1, "soon": 1, "out": 1, "aboriginal": 1, "modified": 1, "legalized": 1, "european": 1, "resembled": 1, "to": 1, "uncertain": 1, "modern": 1, "peoples": 1, "britain": 1, "repertoire": 1, "leather": 1, "web": 1, "school": 1, "many": 1, "edition": 1, "alternative": 1, "cause": 1, "match": 1, "rank": 1, "symbolize": 1, "returned": 1, "forbidden": 1, "formal": 1, "forms": 1, "expand": 1, "involved": 1, "sport": 1, "individual": 1, "7th": 1, "scoto": 1, "tennis": 1, "bolton": 1, "successively": 1, "countythat": 1, "proper": 1, "again": 1, "women\\'s": 1, "skill": 1, "distinction": 1, "appearance": 1, "stadiums\\u203athe": 1, "schools": 1, "invented": 1, "filmalegria": 1, "into": 1, "the": 1, "willing": 1, "netherlands": 1, "old": 1, "pitch\\u201d": 1, "status": 1, "1848\\u00e2\\u0080\\u00931911": 1, "scottish": 1, "games": 1, "aztecs": 1, "death": 1, "competition": 1, "biggest": 1, "workers": 1, "instead": 1, "goodison": 1, "da": 1, "org": 1, "rome": 1, "also\\u00e2\\u00a0resulted": 1, "exclusively": 1, "losing": 1, "kicks": 1, "francesco": 1, "captain": 1, "area": 1, "working": 1, "century": 1, "formed": 1, "areas": 1, "attempt": 1, "\\u203aformations": 1, "des": 1, "game": 1, "since": 1, "seems": 1, "global": 1, "imposing": 1, "competitions": 1, "involving": 1, "namecuju": 1, "event": 1, "1885": 1, "music": 1, "cities": 1, "predominant": 1, "andrew": 1, "today": 1, "1862": 1, "divisions": 1, "gradually": 1, "north": 1, "twelve": 1, "meadows": 1, "concerning": 1, "picture": 1, "square": 1, "rough": 1, "does": 1, "f\\u00e3\\u00a9d\\u00e3\\u00a9ration": 1, "stubborn": 1, "http": 1, "this": 1, "final": 1, "clear": 1, "noticed": 1, "links": 1, "inaccessible": 1, "these": 1, "step": 1, "ritual": 1, "fourth": 1, "here": 1, "ofkemaripracticed": 1, "fifa": 1, "follow": 1, "itself": 1, "punches": 1, "season": 1, "approach": 1, "1991": 1, "gook": 1, "middle": 1, "traditionally": 1, "ancient": 1, "consequently": 1, "two": 1, "decided": 1, "years": 1, "finally": 1, "4": 1, "conmebol": 1, "popular": 1, "rubber": 1, "cultures": 1, "uncertainty": 1, "butler": 1, "tragedy": 1, "style": 1,

"neither": 1, "bryon": 1, "churches": 1, "history": 1, "held": 1, "roads": 1, "unprecedented": 1, "therefore": 1, "sites": 1, "places": 1, "centuries": 1, "meeting": 1, "000": 1, "calledtchatali": 1, "we": 1, "league": 1, "of": 1, "endeavor": 1, "budapest": 1, "signed": 1, "their": 1, "fashion": 1, "but": 1, "oldest": 1, "europe": 1, "1867": 1, "although": 1, "culture": 1, "during": 1, "consisting": 1, "example": 1, "former": 1, "older": 1, "pass": 1, "performances": 1, "male": 1, "triumph": 1, "perhaps": 1, "level": 1, "1904": 1, "reasons": 1, "countries\\u00e2\\u00a0did": 1, "concacaf": 1, "dimensions\\u203ahistory": 1, "behind": 1, "\\u201cchess": 1, "low": 1, "pantomime": 1, "following": 1, "12th": 1, "states": 1, "groups": 1, "spartacus": 1, "difference": 1, "rampage": 1, "town": 1, "200": 1, "british": 1, "england": 1, "great": 1, "still": 1, "suggest": 1, "mesoamerican": 1, "domestic": 1, "carnival": 1, "increase": 1, "ofharpastum": 1, "watching": 1, "called": 1, "china": 1, "vary": 1, "weight": 1, "staged": 1, "according": 1, "up": 1, "1888": 1, "chose": 1, "now": 1, "things": 1, "speech": 1, "dominance": 1, "time": 1, "header": 1, "game\\u00e2\\u0080\\u009d": 1, "1870s": 1, "epic": 1, "yet": 1, "are": 1, "he": 1, "escapes": 1, "bertelli": 1, "argentina": 1, "industrialization": 1, "emigrants": 1, "possibility": 1, "front": 1, "were": 1, "1835": 1, "f\\u00e3\\u00batbol": 1, "six": 1, "experimental": 1, "partake": 1, "30": 1, "confederation": 1, "more": 1, "foundation": 1, "france": 1, "played": 1, "develop": 1, "inverness": 1, "international": 1, "kind": 1, "factories": 1, "unorganized": 1, "hall": 1, "already": 1, "clubnotts": 1, "sold": 1, "county": 1, "relatively": 1, "only": 1, "purpose": 1, "about": 1, "united": 1, "1950": 1, "beginning": 1, "year": 1, "final\\u00e2\\u00a0solution\\u00e2\\u00a0to": 1, "achieved": 1, "spectators": 1, "few": 1, "documents": 1, "has": 1, "crescent": 1, "mentioned": 1, "fblack": 1, "another": 1, "socceris": 1, "governing": 1, "be": 1, "present": 1, "wasn\'t": 1, "equivalent": 1, "unique": 1, "1934": 1, "majority": 1, "pubs": 1, "for": 1, "cuju": 1, "one": 1, "paying": 1, "existed": 1, "fist": 1, "references": 1, "damage": 1, "stalemates": 1, "running": 1, "and": 1, "among": 1, "based": 1, "thefifa": 1, "large": 1, "ahead": 1, "balls": 1, "every": 1, "continue": 1, "filled": 1, "influenced": 1, "15th": 1, "it": 1, "1800s": 1, "park": 1, "caribbean": 1, "languages\\u203acollection": 1, "practice": 1, "streets": 1, "211": 1, "songs": 1, "collective": 1, "period": 1, "oceania": 1, "feet": 1, "several": 1, "examples": 1, "story": 1, "capacity": 1, "1883": 1, "class": 1, "mixture": 1, "human": 1, "carrying": 1, "occasions": 1, "or": 1, "join": 1, "today\\u00e2\\u0080\\u0099s": 1, "efforts": 1}',
 'index_docs': '{"571": 1, "336": 1, "381": 1, "389": 1, "771": 1, "113": 1, "258": 1, "639": 1, "401": 1, "512": 1, "444": 1, "688": 1, "629": 1, "611": 1, "794": 1, "302": 1, "520": 1, "47": 1, "580": 1, "138": 1, "86": 1, "155": 1, "58": 1, "545": 1, "760": 1, "281": 1, "690": 1, "97": 1, "427": 1, "99": 1, "212": 1, "686": 1, "180": 1, "574": 1, "744": 1, "772": 1, "631": 1, "463": 1, "62": 1, "651": 1, "804": 1, "663": 1, "514": 1, "13": 1, "706": 1, "508": 1, "60": 1, "560": 1, "671": 1, "778": 1, "379": 1, "396": 1, "41": 1, "505": 1, "103": 1, "234": 1, "464": 1, "618": 1, "525": 1, "653": 1, "678": 1, "6": 1, "481": 1, "672": 1, "300": 1, "35": 1, "757": 1, "453": 1, "776": 1, "500": 1, "437": 1, "615": 1, "376": 1, "274": 1, "177": 1, "517": 1, "592": 1, "562": 1, "774": 1, "134": 1, "317": 1, "586": 1, "172": 1, "10":

1, "522": 1, "144": 1, "147": 1, "537": 1, "22": 1, "605": 1, "579": 1,
"619": 1, "590": 1, "454": 1, "533": 1, "404": 1, "196": 1, "421": 1, "552":
1, "668": 1, "694": 1, "316": 1, "711": 1, "340": 1, "192": 1, "414": 1,
"529": 1, "295": 1, "270": 1, "356": 1, "167": 1, "57": 1, "40": 1, "5": 1,
"121": 1, "223": 1, "15": 1, "244": 1, "797": 1, "110": 1, "294": 1, "53": 1,
"585": 1, "290": 1, "621": 1, "607": 1, "179": 1, "447": 1, "101": 1, "633":
1, "435": 1, "235": 1, "640": 1, "206": 1, "205": 1, "21": 1, "232": 1,
"570": 1, "676": 1, "333": 1, "239": 1, "697": 1, "265": 1, "105": 1, "551":
1, "751": 1, "66": 1, "216": 1, "74": 1, "467": 1, "46": 1, "708": 1, "627":
1, "531": 1, "19": 1, "178": 1, "353": 1, "115": 1, "285": 1, "424": 1,
"449": 1, "137": 1, "226": 1, "2": 1, "341": 1, "197": 1, "553": 1, "236": 1,
"429": 1, "498": 1, "54": 1, "546": 1, "309": 1, "702": 1, "779": 1, "253":
1, "474": 1, "795": 1, "451": 1, "541": 1, "208": 1, "33": 1, "245": 1,
"478": 1, "143": 1, "219": 1, "188": 1, "51": 1, "680": 1, "114": 1, "670":
1, "536": 1, "218": 1, "368": 1, "169": 1, "288": 1, "465": 1, "76": 1,
"142": 1, "385": 1, "759": 1, "416": 1, "373": 1, "660": 1, "100": 1, "260":
1, "428": 1, "476": 1, "664": 1, "16": 1, "493": 1, "735": 1, "95": 1, "55":
1, "801": 1, "331": 1, "82": 1, "400": 1, "649": 1, "405": 1, "692": 1,
"506": 1, "146": 1, "460": 1, "787": 1, "387": 1, "193": 1, "538": 1, "684":
1, "139": 1, "561": 1, "606": 1, "425": 1, "50": 1, "785": 1, "262": 1,
"682": 1, "707": 1, "763": 1, "598": 1, "75": 1, "566": 1, "68": 1, "709": 1,
"31": 1, "436": 1, "661": 1, "77": 1, "87": 1, "96": 1, "677": 1, "329": 1,
"432": 1, "643": 1, "588": 1, "63": 1, "284": 1, "539": 1, "420": 1, "419":
1, "556": 1, "489": 1, "361": 1, "18": 1, "127": 1, "374": 1, "530": 1,
"724": 1, "337": 1, "497": 1, "145": 1, "604": 1, "431": 1, "91": 1, "775":
1, "402": 1, "736": 1, "7": 1, "417": 1, "70": 1, "806": 1, "470": 1, "712":
1, "187": 1, "370": 1, "131": 1, "90": 1, "469": 1, "609": 1, "357": 1,
"780": 1, "716": 1, "321": 1, "743": 1, "575": 1, "634": 1, "398": 1, "118":
1, "201": 1, "513": 1, "289": 1, "173": 1, "52": 1, "632": 1, "495": 1,
"318": 1, "486": 1, "601": 1, "783": 1, "652": 1, "491": 1, "563": 1, "106":
1, "59": 1, "116": 1, "42": 1, "37": 1, "721": 1, "695": 1, "306": 1, "261":
1, "731": 1, "73": 1, "461": 1, "526": 1, "344": 1, "194": 1, "394": 1,
"523": 1, "228": 1, "240": 1, "123": 1, "24": 1, "56": 1, "450": 1, "222": 1,
"638": 1, "650": 1, "733": 1, "191": 1, "642": 1, "762": 1, "802": 1, "408":
1, "243": 1, "293": 1, "713": 1, "656": 1, "163": 1, "165": 1, "747": 1,
"393": 1, "251": 1, "687": 1, "171": 1, "455": 1, "761": 1, "407": 1, "602":
1, "681": 1, "71": 1, "247": 1, "624": 1, "246": 1, "589": 1, "30": 1, "85":
1, "710": 1, "89": 1, "184": 1, "636": 1, "703": 1, "805": 1, "313": 1,
"324": 1, "502": 1, "259": 1, "740": 1, "781": 1, "252": 1, "745": 1, "669":
1, "550": 1, "88": 1, "384": 1, "568": 1, "64": 1, "773": 1, "412": 1, "152":
1, "159": 1, "11": 1, "485": 1, "782": 1, "625": 1, "335": 1, "646": 1,
"544": 1, "518": 1, "332": 1, "195": 1, "161": 1, "111": 1, "136": 1, "241":
1, "645": 1, "689": 1, "181": 1, "273": 1, "230": 1, "296": 1, "355": 1,
"360": 1, "691": 1, "67": 1, "199": 1, "578": 1, "613": 1, "790": 1, "777":
1, "487": 1, "628": 1, "739": 1, "170": 1, "800": 1, "501": 1, "326": 1,
"322": 1, "701": 1, "130": 1, "287": 1, "715": 1, "257": 1, "264": 1, "150":
1, "626": 1, "17": 1, "362": 1, "573": 1, "354": 1, "38": 1, "315": 1, "320":
1, "209": 1, "521": 1, "608": 1, "637": 1, "667": 1, "108": 1, "182": 1,
"154": 1, "291": 1, "788": 1, "662": 1, "786": 1, "459": 1, "84": 1, "655":
1, "268": 1, "342": 1, "359": 1, "388": 1, "314": 1, "133": 1, "269": 1,

"308": 1, "299": 1, "511": 1, "555": 1, "346": 1, "8": 1, "338": 1, "119": 1, "448": 1, "599": 1, "641": 1, "168": 1, "769": 1, "210": 1, "72": 1, "659": 1, "267": 1, "364": 1, "125": 1, "587": 1, "280": 1, "248": 1, "186": 1, "754": 1, "303": 1, "135": 1, "69": 1, "49": 1, "714": 1, "327": 1, "789": 1, "596": 1, "623": 1, "494": 1, "472": 1, "399": 1, "377": 1, "224": 1, "742": 1, "189": 1, "426": 1, "765": 1, "124": 1, "577": 1, "803": 1, "92": 1, "1": 1, "496": 1, "572": 1, "271": 1, "748": 1, "176": 1, "796": 1, "98": 1, "20": 1, "275": 1, "367": 1, "220": 1, "148": 1, "214": 1, "753": 1, "616": 1, "791": 1, "153": 1, "330": 1, "413": 1, "391": 1, "283": 1, "349": 1, "792": 1, "282": 1, "297": 1, "207": 1, "27": 1, "80": 1, "183": 1, "397": 1, "758": 1, "693": 1, "9": 1, "102": 1, "319": 1, "149": 1, "630": 1, "263": 1, "157": 1, "292": 1, "200": 1, "510": 1, "725": 1, "213": 1, "126": 1, "600": 1, "79": 1, "473": 1, "221": 1, "554": 1, "696": 1, "547": 1, "347": 1, "190": 1, "543": 1, "298": 1, "352": 1, "730": 1, "229": 1, "737": 1, "255": 1, "32": 1, "620": 1, "383": 1, "434": 1, "767": 1, "727": 1, "120": 1, "475": 1, "278": 1, "250": 1, "390": 1, "301": 1, "36": 1, "535": 1, "657": 1, "350": 1, "516": 1, "441": 1, "665": 1, "307": 1, "266": 1, "597": 1, "174": 1, "519": 1, "78": 1, "409": 1, "109": 1, "185": 1, "540": 1, "705": 1, "755": 1, "162": 1, "272": 1, "741": 1, "799": 1, "720": 1, "648": 1, "423": 1, "798": 1, "484": 1, "43": 1, "666": 1, "348": 1, "443": 1, "458": 1, "768": 1, "480": 1, "372": 1, "93": 1, "65": 1, "276": 1, "129": 1, "141": 1, "3": 1, "718": 1, "524": 1, "567": 1, "117": 1, "438": 1, "25": 1, "211": 1, "557": 1, "559": 1, "277": 1, "112": 1, "515": 1, "490": 1, "595": 1, "466": 1, "305": 1, "440": 1, "729": 1, "591": 1, "732": 1, "304": 1, "140": 1, "564": 1, "369": 1, "576": 1, "699": 1, "764": 1, "503": 1, "746": 1, "328": 1, "723": 1, "227": 1, "345": 1, "750": 1, "479": 1, "256": 1, "433": 1, "363": 1, "366": 1, "242": 1, "29": 1, "39": 1, "614": 1, "132": 1, "462": 1, "158": 1, "582": 1, "647": 1, "679": 1, "334": 1, "612": 1, "104": 1, "164": 1, "422": 1, "415": 1, "719": 1, "310": 1, "128": 1, "225": 1, "439": 1, "534": 1, "528": 1, "756": 1, "527": 1, "34": 1, "430": 1, "198": 1, "673": 1, "722": 1, "204": 1, "45": 1, "603": 1, "734": 1, "793": 1, "558": 1, "477": 1, "311": 1, "386": 1, "610": 1, "23": 1, "378": 1, "704": 1, "685": 1, "728": 1, "581": 1, "446": 1, "151": 1, "48": 1, "565": 1, "569": 1, "28": 1, "202": 1, "532": 1, "549": 1, "468": 1, "482": 1, "457": 1, "726": 1, "233": 1, "471": 1, "509": 1, "622": 1, "594": 1, "217": 1, "617": 1, "61": 1, "749": 1, "231": 1, "488": 1, "81": 1, "403": 1, "406": 1, "445": 1, "238": 1, "325": 1, "44": 1, "542": 1, "583": 1, "784": 1, "94": 1, "752": 1, "675": 1, "26": 1, "593": 1, "411": 1, "358": 1, "286": 1, "683": 1, "492": 1, "483": 1, "12": 1, "166": 1, "203": 1, "504": 1, "107": 1, "351": 1, "770": 1, "365": 1, "738": 1, "395": 1, "4": 1, "122": 1, "584": 1, "658": 1, "160": 1, "507": 1, "175": 1, "249": 1, "418": 1, "323": 1, "339": 1, "456": 1, "14": 1, "312": 1, "237": 1, "698": 1, "766": 1, "382": 1, "375": 1, "674": 1, "644": 1, "717": 1, "499": 1, "700": 1, "392": 1, "371": 1, "156": 1, "343": 1, "635": 1, "548": 1, "442": 1, "654": 1, "254": 1, "410": 1, "279": 1, "83": 1, "215": 1, "380": 1, "452": 1}',
 'index_word': '{"1": "the", "2": "in", "3": "of", "4": "and", "5": "was", "6": "football", "7": "a", "8": "to", "9": "game", "10": "ball", "11": "would", "12": "for", "13": "first", "14": "it", "15": "by", "16": "other", "17": "that", "18": "with", "19": "as", "20": "games", "21": "from", "22": "world", "23": "were", "24": "teams", "25": "but", "26": "be", "27":

"century", "28": "played", "29": "british", "30": "is", "31": "on", "32": "this", "33": "not", "34": "time", "35": "cup", "36": "fifa", "37": "players", "38": "long", "39": "england", "40": "which", "41": "at", "42": "association", "43": "history", "44": "has", "45": "are", "46": "had", "47": "an", "48": "more", "49": "sport", "50": "clubs", "51": "also", "52": "been", "53": "included", "54": "people", "55": "rugby", "56": "took", "57": "place", "58": "rules", "59": "national", "60": "de", "61": "about", "62": "some", "63": "19th", "64": "known", "65": "000", "66": "\\u00e2\\u0080\\u0093", "67": "no", "68": "early", "69": "involved", "70": "later", "71": "name", "72": "many", "73": "have", "74": "however", "75": "important", "76": "there", "77": "between", "78": "two", "79": "today", "80": "formed", "81": "year", "82": "com", "83": "or", "84": "its", "85": "versions", "86": "team", "87": "made", "88": "occurred", "89": "over", "90": "established", "91": "until", "92": "into", "93": "meeting", "94": "another", "95": "could", "96": "used", "97": "english", "98": "scottish", "99": "became", "100": "late", "101": "who", "102": "since", "103": "official", "104": "called", "105": "parts", "106": "form", "107": "existed", "108": "much", "109": "years", "110": "spread", "111": "feature", "112": "culture", "113": "white", "114": "big", "115": "arenas", "116": "they", "117": "their", "118": "most", "119": "modern", "120": "these", "121": "sometimes", "122": "among", "123": "london", "124": "schools", "125": "match", "126": "predominant", "127": "eton", "128": "up", "129": "we", "130": "seen", "131": "when", "132": "still", "133": "soon", "134": "matches", "135": "expand", "136": "south", "137": "without", "138": "club", "139": "professional", "140": "level", "141": "league", "142": "women", "143": "scotland", "144": "tournament", "145": "countries", "146": "olympic", "147": "black", "148": "biggest", "149": "global", "150": "associations", "151": "confederation", "152": "www", "153": "org", "154": "soccer", "155": "part", "156": "examples", "157": "involving", "158": "mesoamerican", "159": "various", "160": "large", "161": "regions", "162": "rubber", "163": "kicking", "164": "china", "165": "under", "166": "cuju", "167": "round", "168": "leather", "169": "even", "170": "than", "171": "primarily", "172": "air", "173": "variety", "174": "ancient", "175": "balls", "176": "status", "177": "entertainment", "178": "bring", "179": "degree", "180": "developed", "181": "besides", "182": "playing", "183": "areas", "184": "against", "185": "finally", "186": "forbidden", "187": "stage", "188": "public", "189": "distinction", "190": "concerning", "191": "size", "192": "number", "193": "often", "194": "hands", "195": "hand", "196": "can", "197": "\\u00e2\\u0080\\u009cthe", "198": "game\\u00e2\\u0080\\u009d", "199": "while", "200": "event", "201": "divided", "202": "develop", "203": "one", "204": "yet", "205": "development", "206": "run", "207": "working", "208": "america", "209": "interest", "210": "school", "211": "oldest", "212": "larger", "213": "cities", "214": "workers", "215": "join", "216": "start", "217": "only", "218": "1880s", "219": "joined", "220": "competition", "221": "divisions", "222": "participating", "223": "started", "224": "women\\'s", "225": "1888", "226": "fa", "227": "following", "228": "0", "229": "f\\u00e3\\u00a9d\\u00e3\\u00a9ration", "230": "internationale", "231": "1950", "232": "leagues", "233": "already", "234": "1930", "235": "sports", "236": "comparison", "237": "park", "238": "few", "239": "built", "240": "stadiums", "241": "stadium", "242": "200", "243": "traditions", "244":

"american", "245": "fans", "246": "adopted", "247": "then", "248": "returned", "249": "every", "250": "fourth", "251": "david", "252": "brilliance", "253": "heart", "254": "human", "255": "http", "256": "spartacus", "257": "educational", "258": "htmhttp", "259": "welcome", "260": "footballhistory", "261": "website", "262": "including", "263": "competitions", "264": "current", "265": "arose", "266": "middle", "267": "alternative", "268": "earlier", "269": "out", "270": "rock", "271": "old", "272": "cultures", "273": "3", "274": "ago", "275": "aztecs", "276": "calledtchatali", "277": "although", "278": "ritual", "279": "occasions", "280": "symbolize", "281": "sun", "282": "captain", "283": "losing", "284": "sacrificed", "285": "gods", "286": "unique", "287": "bouncing", "288": "access", "289": "3rd", "290": "2nd", "291": "bc", "292": "namecuju", "293": "stitched", "294": "fur", "295": "feathers", "296": "inside", "297": "area", "298": "square", "299": "modified", "300": "japan", "301": "ofkemaripracticed", "302": "ceremonial", "303": "forms", "304": "perhaps", "305": "older", "306": "wasmarn", "307": "gook", "308": "aboriginal", "309": "australians", "310": "according", "311": "emigrants", "312": "1800s", "313": "encased", "314": "leaves", "315": "roots", "316": "mostly", "317": "unknown", "318": "chief", "319": "seems", "320": "keep", "321": "greece", "322": "shreds", "323": "filled", "324": "hair", "325": "documents", "326": "filled\\u00e2\\u00a0with", "327": "7th", "328": "low", "329": "panhellenic", "330": "rome", "331": "amphitheaters", "332": "exercises", "333": "military", "334": "ofharpastum", "335": "roman", "336": "island", "337": "britannica", "338": "uncertain", "339": "influenced", "340": "own", "341": "variants", "342": "admitted", "343": "story", "344": "tells", "345": "12th", "346": "resembled", "347": "meadows", "348": "roads", "349": "kicks", "350": "punches", "351": "fist", "352": "rough", "353": "violent", "354": "way", "355": "forerunners", "356": "plenty", "357": "towns", "358": "equivalent", "359": "florence", "360": "16th", "361": "where", "362": "calledcalcio", "363": "rampage", "364": "cause", "365": "damage", "366": "town", "367": "death", "368": "participants", "369": "reasons", "370": "proclamations", "371": "several", "372": "centuries", "373": "like", "374": "return", "375": "streets", "376": "17th", "377": "again", "378": "1835", "379": "features", "380": "today\\u00e2\\u0080\\u0099s", "381": "taken", "382": "practice", "383": "clear", "384": "variations", "385": "length", "386": "possibility", "387": "take", "388": "know", "389": "origin", "390": "here", "391": "exclusively", "392": "feet", "393": "close", "394": "predecessor", "395": "running", "396": "dribbling", "397": "attempt", "398": "create", "399": "proper", "400": "done", "401": "cambridge", "402": "1848", "403": "final\\u00e2\\u00a0solution\\u00e2\\u00a0to", "404": "all", "405": "questions", "406": "achieved", "407": "came", "408": "1863", "409": "decided", "410": "carrying", "411": "wasn\\'t", "412": "allowed", "413": "also\\u00e2\\u00a0resulted", "414": "standardization", "415": "weight", "416": "consequence", "417": "codes", "418": "continue", "419": "flexibility", "420": "thing", "421": "pitch", "422": "vary", "423": "neither", "424": "uniforms", "425": "distinguish", "426": "appearance", "427": "common", "428": "wearing", "429": "caps", "430": "header", "431": "further", "432": "reading", "433": "difference", "434": "noticed", "435": "whereas", "436": "preferred", "437": "forward", "438": "fashion", "439": "chose", "440": "pass", "441": "approach", "442": "class", "443":

"unprecedented", "444": "amounts", "445": "spectators", "446": "30", "447": "see", "448": "peoples", "449": "traveled", "450": "result", "451": "colonization", "452": "efforts", "453": "especially", "454": "india", "455": "become", "456": "15th", "457": "unorganized", "458": "therefore", "459": "hard", "460": "decide", "461": "historians", "462": "suggest", "463": "foot", "464": "1824", "465": "edinburgh", "466": "former", "467": "students", "468": "kind", "469": "sheffield", "470": "1855", "471": "clubnotts", "472": "countythat", "473": "1862", "474": "exists", "475": "step", "476": "emergence", "477": "industrialization", "478": "led", "479": "groups", "480": "places", "481": "such", "482": "factories", "483": "pubs", "484": "churches", "485": "new", "486": "railroads", "487": "them", "488": "beginning", "489": "dominated", "490": "consisting", "491": "make", "492": "majority", "493": "change", "494": "successively", "495": "taking", "496": "willing", "497": "pay", "498": "best", "499": "period", "500": "transition", "501": "friction", "502": "motivation", "503": "behind", "504": "paying", "505": "win", "506": "moved", "507": "ahead", "508": "tickets", "509": "sold", "510": "1885", "511": "legalized", "512": "three", "513": "thefootball", "514": "leaguewas", "515": "during", "516": "season", "517": "12", "518": "interested", "519": "consequently", "520": "dominant", "521": "after", "522": "decades", "523": "prague", "524": "budapest", "525": "sienna", "526": "contenders", "527": "dominance", "528": "things", "529": "excluded", "530": "before", "531": "play", "532": "inverness", "533": "milestones", "534": "now", "535": "follow", "536": "challenge", "537": "1871", "538": "ended", "539": "followed", "540": "4", "541": "hamilton", "542": "crescent", "543": "picture", "544": "shows", "545": "illustrations", "546": "occasion", "547": "twelve", "548": "1883", "549": "international", "550": "four", "551": "ireland", "552": "wales", "553": "phenomenon", "554": "gradually", "555": "european", "556": "outside", "557": "europe", "558": "argentina", "559": "1867", "560": "foreign", "561": "argentinean", "562": "citizens", "563": "founded", "564": "1904", "565": "foundation", "566": "act", "567": "signed", "568": "representatives", "569": "france", "570": "belgium", "571": "denmark", "572": "netherlands", "573": "spain", "574": "sweden", "575": "switzerland", "576": "countries\\u00e2\\u00a0did", "577": "invented", "578": "saw", "579": "reason", "580": "subordinate", "581": "partake", "582": "domestic", "583": "mentioned", "584": "based", "585": "performance", "586": "1908", "587": "rank", "588": "prestigious", "589": "added", "590": "1996", "591": "male", "592": "being", "593": "present", "594": "relatively", "595": "example", "596": "tennis", "597": "traditionally", "598": "mix", "599": "britain", "600": "andrew", "601": "watsonis", "602": "player", "603": "he", "604": "queen\\u00e2\\u0080\\u0099s", "605": "show", "606": "passion", "607": "extent", "608": "flocked", "609": "shearing", "610": "front", "611": "television", "612": "watching", "613": "carefully", "614": "great", "615": "enthusiasm", "616": "goodison", "617": "purpose", "618": "hosting", "619": "1894", "620": "final", "621": "notts", "622": "county", "623": "bolton", "624": "wanderers", "625": "attended", "626": "37", "627": "milestone", "628": "construction", "629": "maracan\\u00e3\\u00a3", "630": "imposing", "631": "rio", "632": "janeiro", "633": "ready", "634": "almost", "635": "capacity", "636": "host", "637": "different", "638": "fan", "639": "tradition", "640": "singing", "641": "repertoire", "642": "inspired", "643":

"pub", "644": "songs", "645": "americans", "646": "adopt", "647": "carnival", "648": "style", "649": "firecrackers", "650": "fireworks", "651": "phenomena", "652": "bengali", "653": "fires", "654": "mixture", "655": "summer", "656": "measure", "657": "itself", "658": "thefifa", "659": "edition", "660": "uruguay", "661": "exceptions", "662": "due", "663": "second", "664": "war", "665": "1991", "666": "held", "667": "thechampions", "668": "1992", "669": "formereuropean", "670": "1955\\u00e2\\u0080\\u00931991", "671": "active", "672": "each", "673": "1870s", "674": "211", "675": "governing", "676": "body", "677": "proof", "678": "globalization", "679": "increase", "680": "nations", "681": "qualifiers", "682": "32", "683": "1934", "684": "2014", "685": "six", "686": "confederations", "687": "conf\\u00e3\\u00a9d\\u00e3\\u00a9ration", "688": "africaine", "689": "caf", "690": "asian", "691": "afc", "692": "union", "693": "des", "694": "europ\\u00e3\\u00a9ennes", "695": "uefa", "696": "north", "697": "central", "698": "caribbean", "699": "concacaf", "700": "oceania", "701": "ofc", "702": "confederaci\\u00e3\\u00b3n", "703": "sudamericana", "704": "f\\u00e3\\u00batbol", "705": "conmebol", "706": "definitely", "707": "quote", "708": "goldblatt\\'s", "709": "bookthe", "710": "roundpresent", "711": "answer", "712": "offers", "713": "spotlight", "714": "individual", "715": "relishing", "716": "defiance", "717": "collective", "718": "endeavor", "719": "staged", "720": "tragedy", "721": "comedy", "722": "epic", "723": "pantomime", "724": "unsophisticated", "725": "music", "726": "hall", "727": "inaccessible", "728": "experimental", "729": "performances", "730": "does", "731": "imperious", "732": "triumph", "733": "lucky", "734": "escapes", "735": "impossible", "736": "comebacks", "737": "stubborn", "738": "stalemates", "739": "captures", "740": "unpredictability", "741": "uncertainty", "742": "skill", "743": "improvisation", "744": "chance", "745": "footballis", "746": "\\u201cchess", "747": "green", "748": "pitch\\u201d", "749": "united", "750": "states", "751": "canada", "752": "socceris", "753": "instead", "754": "formal", "755": "popular", "756": "speech", "757": "either", "758": "\\u203aformations", "759": "systems", "760": "football\\u203athe", "761": "evolution", "762": "shoes\\u203athe", "763": "field", "764": "dimensions\\u203ahistory", "765": "stadiums\\u203athe", "766": "languages\\u203acollection", "767": "links", "768": "sites", "769": "web", "770": "references", "771": "encyclopediahistory", "772": "beautiful", "773": "2002", "774": "documentary", "775": "series", "776": "football\\u00e2\\u0080\\u0093", "777": "goldblatt", "778": "2008", "779": "https", "780": "en", "781": "wikipedia", "782": "wiki", "783": "clubshttp", "784": "fblack", "785": "fstadiums", "786": "image", "787": "sources", "788": "andrea", "789": "scoto", "790": "itiillustration", "791": "da", "792": "francesco", "793": "bertelli", "794": "paduawilliam", "795": "ralston", "796": "1848\\u00e2\\u0080\\u00931911", "797": "scanned", "798": "bryon", "799": "butler", "800": "queen", "801": "ane", "802": "pressdocumentary", "803": "filmalegria", "804": "do", "805": "povo", "806": "1963"}',
 'word_index': '{"the": 1, "in": 2, "of": 3, "and": 4, "was": 5, "football": 6, "a": 7, "to": 8, "game": 9, "ball": 10, "would": 11, "for": 12, "first": 13, "it": 14, "by": 15, "other": 16, "that": 17, "with": 18, "as": 19, "games": 20, "from": 21, "world": 22, "were": 23, "teams": 24, "but": 25, "be": 26, "century": 27, "played": 28, "british": 29, "is": 30, "on": 31,

"this": 32, "not": 33, "time": 34, "cup": 35, "fifa": 36, "players": 37, "long": 38, "england": 39, "which": 40, "at": 41, "association": 42, "history": 43, "has": 44, "are": 45, "had": 46, "an": 47, "more": 48, "sport": 49, "clubs": 50, "also": 51, "been": 52, "included": 53, "people": 54, "rugby": 55, "took": 56, "place": 57, "rules": 58, "national": 59, "de": 60, "about": 61, "some": 62, "19th": 63, "known": 64, "000": 65, "\\u00e2\\u0080\\u0093": 66, "no": 67, "early": 68, "involved": 69, "later": 70, "name": 71, "many": 72, "have": 73, "however": 74, "important": 75, "there": 76, "between": 77, "two": 78, "today": 79, "formed": 80, "year": 81, "com": 82, "or": 83, "its": 84, "versions": 85, "team": 86, "made": 87, "occurred": 88, "over": 89, "established": 90, "until": 91, "into": 92, "meeting": 93, "another": 94, "could": 95, "used": 96, "english": 97, "scottish": 98, "became": 99, "late": 100, "who": 101, "since": 102, "official": 103, "called": 104, "parts": 105, "form": 106, "existed": 107, "much": 108, "years": 109, "spread": 110, "feature": 111, "culture": 112, "white": 113, "big": 114, "arenas": 115, "they": 116, "their": 117, "most": 118, "modern": 119, "these": 120, "sometimes": 121, "among": 122, "london": 123, "schools": 124, "match": 125, "predominant": 126, "eton": 127, "up": 128, "we": 129, "seen": 130, "when": 131, "still": 132, "soon": 133, "matches": 134, "expand": 135, "south": 136, "without": 137, "club": 138, "professional": 139, "level": 140, "league": 141, "women": 142, "scotland": 143, "tournament": 144, "countries": 145, "olympic": 146, "black": 147, "biggest": 148, "global": 149, "associations": 150, "confederation": 151, "www": 152, "org": 153, "soccer": 154, "part": 155, "examples": 156, "involving": 157, "mesoamerican": 158, "various": 159, "large": 160, "regions": 161, "rubber": 162, "kicking": 163, "china": 164, "under": 165, "cuju": 166, "round": 167, "leather": 168, "even": 169, "than": 170, "primarily": 171, "air": 172, "variety": 173, "ancient": 174, "balls": 175, "status": 176, "entertainment": 177, "bring": 178, "degree": 179, "developed": 180, "besides": 181, "playing": 182, "areas": 183, "against": 184, "finally": 185, "forbidden": 186, "stage": 187, "public": 188, "distinction": 189, "concerning": 190, "size": 191, "number": 192, "often": 193, "hands": 194, "hand": 195, "can": 196, "\\u00e2\\u0080\\u009cthe": 197, "game\\u00e2\\u0080\\u009d": 198, "while": 199, "event": 200, "divided": 201, "develop": 202, "one": 203, "yet": 204, "development": 205, "run": 206, "working": 207, "america": 208, "interest": 209, "school": 210, "oldest": 211, "larger": 212, "cities": 213, "workers": 214, "join": 215, "start": 216, "only": 217, "1880s": 218, "joined": 219, "competition": 220, "divisions": 221, "participating": 222, "started": 223, "women\\'s": 224, "1888": 225, "fa": 226, "following": 227, "0": 228, "f\\u00e3\\u00a9d\\u00e3\\u00a9ration": 229, "internationale": 230, "1950": 231, "leagues": 232, "already": 233, "1930": 234, "sports": 235, "comparison": 236, "park": 237, "few": 238, "built": 239, "stadiums": 240, "stadium": 241, "200": 242, "traditions": 243, "american": 244, "fans": 245, "adopted": 246, "then": 247, "returned": 248, "every": 249, "fourth": 250, "david": 251, "brilliance": 252, "heart": 253, "human": 254, "http": 255, "spartacus": 256, "educational": 257, "htmhttp": 258, "welcome": 259, "footballhistory": 260, "website": 261, "including": 262, "competitions": 263, "current": 264, "arose": 265, "middle": 266, "alternative": 267, "earlier": 268, "out": 269, "rock": 270, "old": 271, "cultures": 272, "3":

273, "ago": 274, "aztecs": 275, "calledtchatali": 276, "although": 277, "ritual": 278, "occasions": 279, "symbolize": 280, "sun": 281, "captain": 282, "losing": 283, "sacrificed": 284, "gods": 285, "unique": 286, "bouncing": 287, "access": 288, "3rd": 289, "2nd": 290, "bc": 291, "namecuju": 292, "stitched": 293, "fur": 294, "feathers": 295, "inside": 296, "area": 297, "square": 298, "modified": 299, "japan": 300, "ofkemaripracticed": 301, "ceremonial": 302, "forms": 303, "perhaps": 304, "older": 305, "wasmarn": 306, "gook": 307, "aboriginal": 308, "australians": 309, "according": 310, "emigrants": 311, "1800s": 312, "encased": 313, "leaves": 314, "roots": 315, "mostly": 316, "unknown": 317, "chief": 318, "seems": 319, "keep": 320, "greece": 321, "shreds": 322, "filled": 323, "hair": 324, "documents": 325, "filled\\u00e2\\u00a0with": 326, "7th": 327, "low": 328, "panhellenic": 329, "rome": 330, "amphitheaters": 331, "exercises": 332, "military": 333, "ofharpastum": 334, "roman": 335, "island": 336, "britannica": 337, "uncertain": 338, "influenced": 339, "own": 340, "variants": 341, "admitted": 342, "story": 343, "tells": 344, "12th": 345, "resembled": 346, "meadows": 347, "roads": 348, "kicks": 349, "punches": 350, "fist": 351, "rough": 352, "violent": 353, "way": 354, "forerunners": 355, "plenty": 356, "towns": 357, "equivalent": 358, "florence": 359, "16th": 360, "where": 361, "calledcalcio": 362, "rampage": 363, "cause": 364, "damage": 365, "town": 366, "death": 367, "participants": 368, "reasons": 369, "proclamations": 370, "several": 371, "centuries": 372, "like": 373, "return": 374, "streets": 375, "17th": 376, "again": 377, "1835": 378, "features": 379, "today\\u00e2\\u0080\\u0099s": 380, "taken": 381, "practice": 382, "clear": 383, "variations": 384, "length": 385, "possibility": 386, "take": 387, "know": 388, "origin": 389, "here": 390, "exclusively": 391, "feet": 392, "close": 393, "predecessor": 394, "running": 395, "dribbling": 396, "attempt": 397, "create": 398, "proper": 399, "done": 400, "cambridge": 401, "1848": 402, "final\\u00e2\\u00a0solution\\u00e2\\u00a0to": 403, "all": 404, "questions": 405, "achieved": 406, "came": 407, "1863": 408, "decided": 409, "carrying": 410, "wasn\'t": 411, "allowed": 412, "also\\u00e2\\u00a0resulted": 413, "standardization": 414, "weight": 415, "consequence": 416, "codes": 417, "continue": 418, "flexibility": 419, "thing": 420, "pitch": 421, "vary": 422, "neither": 423, "uniforms": 424, "distinguish": 425, "appearance": 426, "common": 427, "wearing": 428, "caps": 429, "header": 430, "further": 431, "reading": 432, "difference": 433, "noticed": 434, "whereas": 435, "preferred": 436, "forward": 437, "fashion": 438, "chose": 439, "pass": 440, "approach": 441, "class": 442, "unprecedented": 443, "amounts": 444, "spectators": 445, "30": 446, "see": 447, "peoples": 448, "traveled": 449, "result": 450, "colonization": 451, "efforts": 452, "especially": 453, "india": 454, "become": 455, "15th": 456, "unorganized": 457, "therefore": 458, "hard": 459, "decide": 460, "historians": 461, "suggest": 462, "foot": 463, "1824": 464, "edinburgh": 465, "former": 466, "students": 467, "kind": 468, "sheffield": 469, "1855": 470, "clubnotts": 471, "countythat": 472, "1862": 473, "exists": 474, "step": 475, "emergence": 476, "industrialization": 477, "led": 478, "groups": 479, "places": 480, "such": 481, "factories": 482, "pubs": 483, "churches": 484, "new": 485, "railroads": 486, "them": 487, "beginning": 488, "dominated": 489, "consisting": 490, "make": 491, "majority": 492, "change": 493, "successively": 494, "taking":

495, "willing": 496, "pay": 497, "best": 498, "period": 499, "transition":
500, "friction": 501, "motivation": 502, "behind": 503, "paying": 504, "win":
505, "moved": 506, "ahead": 507, "tickets": 508, "sold": 509, "1885": 510,
"legalized": 511, "three": 512, "thefootball": 513, "leaguewas": 514,
"during": 515, "season": 516, "12": 517, "interested": 518, "consequently":
519, "dominant": 520, "after": 521, "decades": 522, "prague": 523,
"budapest": 524, "sienna": 525, "contenders": 526, "dominance": 527,
"things": 528, "excluded": 529, "before": 530, "play": 531, "inverness": 532,
"milestones": 533, "now": 534, "follow": 535, "challenge": 536, "1871": 537,
"ended": 538, "followed": 539, "4": 540, "hamilton": 541, "crescent": 542,
"picture": 543, "shows": 544, "illustrations": 545, "occasion": 546,
"twelve": 547, "1883": 548, "international": 549, "four": 550, "ireland":
551, "wales": 552, "phenomenon": 553, "gradually": 554, "european": 555,
"outside": 556, "europe": 557, "argentina": 558, "1867": 559, "foreign": 560,
"argentinean": 561, "citizens": 562, "founded": 563, "1904": 564,
"foundation": 565, "act": 566, "signed": 567, "representatives": 568,
"france": 569, "belgium": 570, "denmark": 571, "netherlands": 572, "spain":
573, "sweden": 574, "switzerland": 575, "countries\\u00e2\\u00a0did": 576,
"invented": 577, "saw": 578, "reason": 579, "subordinate": 580, "partake":
581, "domestic": 582, "mentioned": 583, "based": 584, "performance": 585,
"1908": 586, "rank": 587, "prestigious": 588, "added": 589, "1996": 590,
"male": 591, "being": 592, "present": 593, "relatively": 594, "example": 595,
"tennis": 596, "traditionally": 597, "mix": 598, "britain": 599, "andrew":
600, "watsonis": 601, "player": 602, "he": 603,
"queen\\u00e2\\u0080\\u0099s": 604, "show": 605, "passion": 606, "extent":
607, "flocked": 608, "shearing": 609, "front": 610, "television": 611,
"watching": 612, "carefully": 613, "great": 614, "enthusiasm": 615,
"goodison": 616, "purpose": 617, "hosting": 618, "1894": 619, "final": 620,
"notts": 621, "county": 622, "bolton": 623, "wanderers": 624, "attended":
625, "37": 626, "milestone": 627, "construction": 628,
"maracan\\u00e3\\u00a3": 629, "imposing": 630, "rio": 631, "janeiro": 632,
"ready": 633, "almost": 634, "capacity": 635, "host": 636, "different": 637,
"fan": 638, "tradition": 639, "singing": 640, "repertoire": 641, "inspired":
642, "pub": 643, "songs": 644, "americans": 645, "adopt": 646, "carnival":
647, "style": 648, "firecrackers": 649, "fireworks": 650, "phenomena": 651,
"bengali": 652, "fires": 653, "mixture": 654, "summer": 655, "measure": 656,
"itself": 657, "thefifa": 658, "edition": 659, "uruguay": 660, "exceptions":
661, "due": 662, "second": 663, "war": 664, "1991": 665, "held": 666,
"thechampions": 667, "1992": 668, "formereuropean": 669,
"1955\\u00e2\\u0080\\u00931991": 670, "active": 671, "each": 672, "1870s":
673, "211": 674, "governing": 675, "body": 676, "proof": 677,
"globalization": 678, "increase": 679, "nations": 680, "qualifiers": 681,
"32": 682, "1934": 683, "2014": 684, "six": 685, "confederations": 686,
"conf\\u00e3\\u00a9d\\u00e3\\u00a9ration": 687, "africaine": 688, "caf": 689,
"asian": 690, "afc": 691, "union": 692, "des": 693,
"europ\\u00e3\\u00a9ennes": 694, "uefa": 695, "north": 696, "central": 697,
"caribbean": 698, "concacaf": 699, "oceania": 700, "ofc": 701,
"confederaci\\u00e3\\u00b3n": 702, "sudamericana": 703,
"f\\u00e3\\u00batbol": 704, "conmebol": 705, "definitely": 706, "quote": 707,
"goldblatt\'s": 708, "bookthe": 709, "roundpresent": 710, "answer": 711,

"offers": 712, "spotlight": 713, "individual": 714, "relishing": 715, "defiance": 716, "collective": 717, "endeavor": 718, "staged": 719, "tragedy": 720, "comedy": 721, "epic": 722, "pantomime": 723, "unsophisticated": 724, "music": 725, "hall": 726, "inaccessible": 727, "experimental": 728, "performances": 729, "does": 730, "imperious": 731, "triumph": 732, "lucky": 733, "escapes": 734, "impossible": 735, "comebacks": 736, "stubborn": 737, "stalemates": 738, "captures": 739, "unpredictability": 740, "uncertainty": 741, "skill": 742, "improvisation": 743, "chance": 744, "footballis": 745, "\\u201cchess": 746, "green": 747, "pitch\\u201d": 748, "united": 749, "states": 750, "canada": 751, "socceris": 752, "instead": 753, "formal": 754, "popular": 755, "speech": 756, "either": 757, "\\\\u203aformations": 758, "systems": 759, "football\\u203athe": 760, "evolution": 761, "shoes\\u203athe": 762, "field": 763, "dimensions\\u203ahistory": 764, "stadiums\\u203athe": 765, "languages\\u203acollection": 766, "links": 767, "sites": 768, "web": 769, "references": 770, "encyclopediahistory": 771, "beautiful": 772, "2002": 773, "documentary": 774, "series": 775, "football\\u00e2\\u0080\\u0093": 776, "goldblatt": 777, "2008": 778, "https": 779, "en": 780, "wikipedia": 781, "wiki": 782, "clubshttp": 783, "fblack": 784, "fstadiums": 785, "image": 786, "sources": 787, "andrea": 788, "scoto": 789, "itiillustration": 790, "da": 791, "francesco": 792, "bertelli": 793, "paduawilliam": 794, "ralston": 795, "1848\\u00e2\\u0080\\u00931911": 796, "scanned": 797, "bryon": 798, "butler": 799, "queen": 800, "ane": 801, "pressdocumentary": 802, "filmalegria": 803, "do": 804, "povo": 805, "1963": 806}'}

Get the Each Token index
tokenizer.word_index

{'the': 1,
 'in': 2,
 'of': 3,
 'and': 4,
 'was': 5,
 'football': 6,
 'a': 7,
 'to': 8,
 'game': 9,
 'ball': 10,
 'would': 11,
 'for': 12,
 'first': 13,
 'it': 14,
 'by': 15,
 'other': 16,
 'that': 17,
 'with': 18,
 'as': 19,
 'games': 20,
 'from': 21,
 'world': 22,

```
'were': 23,
'teams': 24,
'but': 25,
'be': 26,
'century': 27,
'played': 28,
'british': 29,
'is': 30,
'on': 31,
'this': 32,
'not': 33,
'time': 34,
'cup': 35,
'fifa': 36,
'players': 37,
'long': 38,
'england': 39,
'which': 40,
'at': 41,
'association': 42,
'history': 43,
'has': 44,
'are': 45,
'had': 46,
'an': 47,
'more': 48,
'sport': 49,
'clubs': 50,
'also': 51,
'been': 52,
'included': 53,
'people': 54,
'rugby': 55,
'took': 56,
'place': 57,
'rules': 58,
'national': 59,
'de': 60,
'about': 61,
'some': 62,
'19th': 63,
'known': 64,
'000': 65,
'â\x80\x93': 66,
'no': 67,
'early': 68,
'involved': 69,
'later': 70,
'name': 71,
'many': 72,
```

```
'have': 73,
'however': 74,
'important': 75,
'there': 76,
'between': 77,
'two': 78,
'today': 79,
'formed': 80,
'year': 81,
'com': 82,
'or': 83,
'its': 84,
'versions': 85,
'team': 86,
'made': 87,
'occurred': 88,
'over': 89,
'established': 90,
'until': 91,
'into': 92,
'meeting': 93,
'another': 94,
'could': 95,
'used': 96,
'english': 97,
'scottish': 98,
'became': 99,
'late': 100,
'who': 101,
'since': 102,
'official': 103,
'called': 104,
'parts': 105,
'form': 106,
'existed': 107,
'much': 108,
'years': 109,
'spread': 110,
'feature': 111,
'culture': 112,
'white': 113,
'big': 114,
'arenas': 115,
'they': 116,
'their': 117,
'most': 118,
'modern': 119,
'these': 120,
'sometimes': 121,
'among': 122,
```

```
'london': 123,
'schools': 124,
'match': 125,
'predominant': 126,
'eton': 127,
'up': 128,
'we': 129,
'seen': 130,
'when': 131,
'still': 132,
'soon': 133,
'matches': 134,
'expand': 135,
'south': 136,
'without': 137,
'club': 138,
'professional': 139,
'level': 140,
'league': 141,
'women': 142,
'scotland': 143,
'tournament': 144,
'countries': 145,
'olympic': 146,
'black': 147,
'biggest': 148,
'global': 149,
'associations': 150,
'confederation': 151,
'www': 152,
'org': 153,
'soccer': 154,
'part': 155,
'examples': 156,
'involving': 157,
'mesoamerican': 158,
'various': 159,
'large': 160,
'regions': 161,
'rubber': 162,
'kicking': 163,
'china': 164,
'under': 165,
'cuju': 166,
'round': 167,
'leather': 168,
'even': 169,
'than': 170,
'primarily': 171,
'air': 172,
```

```
'variety': 173,
'ancient': 174,
'balls': 175,
'status': 176,
'entertainment': 177,
'bring': 178,
'degree': 179,
'developed': 180,
'besides': 181,
'playing': 182,
'areas': 183,
'against': 184,
'finally': 185,
'forbidden': 186,
'stage': 187,
'public': 188,
'distinction': 189,
'concerning': 190,
'size': 191,
'number': 192,
'often': 193,
'hands': 194,
'hand': 195,
'can': 196,
'â\x80\x9cthe': 197,
'gameâ\x80\x9d': 198,
'while': 199,
'event': 200,
'divided': 201,
'develop': 202,
'one': 203,
'yet': 204,
'development': 205,
'run': 206,
'working': 207,
'america': 208,
'interest': 209,
'school': 210,
'oldest': 211,
'larger': 212,
'cities': 213,
'workers': 214,
'join': 215,
'start': 216,
'only': 217,
'1880s': 218,
'joined': 219,
'competition': 220,
'divisions': 221,
'participating': 222,
```

```
'started': 223,
"women's": 224,
'1888': 225,
'fa': 226,
'following': 227,
'0': 228,
'fã©dã©ration': 229,
'internationale': 230,
'1950': 231,
'leagues': 232,
'already': 233,
'1930': 234,
'sports': 235,
'comparison': 236,
'park': 237,
'few': 238,
'built': 239,
'stadiums': 240,
'stadium': 241,
'200': 242,
'traditions': 243,
'american': 244,
'fans': 245,
'adopted': 246,
'then': 247,
'returned': 248,
'every': 249,
'fourth': 250,
'david': 251,
'brilliance': 252,
'heart': 253,
'human': 254,
'http': 255,
'spartacus': 256,
'educational': 257,
'htmhttp': 258,
'welcome': 259,
'footballhistory': 260,
'website': 261,
'including': 262,
'competitions': 263,
'current': 264,
'arose': 265,
'middle': 266,
'alternative': 267,
'earlier': 268,
'out': 269,
'rock': 270,
'old': 271,
'cultures': 272,
```

```
'3': 273,
'ago': 274,
'aztecs': 275,
'calledtchatali': 276,
'although': 277,
'ritual': 278,
'occasions': 279,
'symbolize': 280,
'sun': 281,
'captain': 282,
'losing': 283,
'sacrificed': 284,
'gods': 285,
'unique': 286,
'bouncing': 287,
'access': 288,
'3rd': 289,
'2nd': 290,
'bc': 291,
'namecuju': 292,
'stitched': 293,
'fur': 294,
'feathers': 295,
'inside': 296,
'area': 297,
'square': 298,
'modified': 299,
'japan': 300,
'ofkemaripracticed': 301,
'ceremonial': 302,
'forms': 303,
'perhaps': 304,
'older': 305,
'wasmarn': 306,
'gook': 307,
'aboriginal': 308,
'australians': 309,
'according': 310,
'emigrants': 311,
'1800s': 312,
'encased': 313,
'leaves': 314,
'roots': 315,
'mostly': 316,
'unknown': 317,
'chief': 318,
'seems': 319,
'keep': 320,
'greece': 321,
'shreds': 322,
```

```
'filled': 323,
'hair': 324,
'documents': 325,
'filledâ\xa0with': 326,
'7th': 327,
'low': 328,
'panhellenic': 329,
'rome': 330,
'amphitheaters': 331,
'exercises': 332,
'military': 333,
'ofharpastum': 334,
'roman': 335,
'island': 336,
'britannica': 337,
'uncertain': 338,
'influenced': 339,
'own': 340,
'variants': 341,
'admitted': 342,
'story': 343,
'tells': 344,
'12th': 345,
'resembled': 346,
'meadows': 347,
'roads': 348,
'kicks': 349,
'punches': 350,
'fist': 351,
'rough': 352,
'violent': 353,
'way': 354,
'forerunners': 355,
'plenty': 356,
'towns': 357,
'equivalent': 358,
'florence': 359,
'16th': 360,
'where': 361,
'calledcalcio': 362,
'rampage': 363,
'cause': 364,
'damage': 365,
'town': 366,
'death': 367,
'participants': 368,
'reasons': 369,
'proclamations': 370,
'several': 371,
'centuries': 372,
```

```
'like': 373,
'return': 374,
'streets': 375,
'17th': 376,
'again': 377,
'1835': 378,
'features': 379,
'todayâ\x80\x99s': 380,
'taken': 381,
'practice': 382,
'clear': 383,
'variations': 384,
'length': 385,
'possibility': 386,
'take': 387,
'know': 388,
'origin': 389,
'here': 390,
'exclusively': 391,
'feet': 392,
'close': 393,
'predecessor': 394,
'running': 395,
'dribbling': 396,
'attempt': 397,
'create': 398,
'proper': 399,
'done': 400,
'cambridge': 401,
'1848': 402,
'finalâ\xa0solutionâ\xa0to': 403,
'all': 404,
'questions': 405,
'achieved': 406,
'came': 407,
'1863': 408,
'decided': 409,
'carrying': 410,
"wasn't": 411,
'allowed': 412,
'alsoâ\xa0resulted': 413,
'standardization': 414,
'weight': 415,
'consequence': 416,
'codes': 417,
'continue': 418,
'flexibility': 419,
'thing': 420,
'pitch': 421,
'vary': 422,
```

```
'neither': 423,
'uniforms': 424,
'distinguish': 425,
'appearance': 426,
'common': 427,
'wearing': 428,
'caps': 429,
'header': 430,
'further': 431,
'reading': 432,
'difference': 433,
'noticed': 434,
'whereas': 435,
'preferred': 436,
'forward': 437,
'fashion': 438,
'chose': 439,
'pass': 440,
'approach': 441,
'class': 442,
'unprecedented': 443,
'amounts': 444,
'spectators': 445,
'30': 446,
'see': 447,
'peoples': 448,
'traveled': 449,
'result': 450,
'colonization': 451,
'efforts': 452,
'especially': 453,
'india': 454,
'become': 455,
'15th': 456,
'unorganized': 457,
'therefore': 458,
'hard': 459,
'decide': 460,
'historians': 461,
'suggest': 462,
'foot': 463,
'1824': 464,
'edinburgh': 465,
'former': 466,
'students': 467,
'kind': 468,
'sheffield': 469,
'1855': 470,
'clubnotts': 471,
'countythat': 472,
```

```
'1862': 473,
'exists': 474,
'step': 475,
'emergence': 476,
'industrialization': 477,
'led': 478,
'groups': 479,
'places': 480,
'such': 481,
'factories': 482,
'pubs': 483,
'churches': 484,
'new': 485,
'railroads': 486,
'them': 487,
'beginning': 488,
'dominated': 489,
'consisting': 490,
'make': 491,
'majority': 492,
'change': 493,
'successively': 494,
'taking': 495,
'willing': 496,
'pay': 497,
'best': 498,
'period': 499,
'transition': 500,
'friction': 501,
'motivation': 502,
'behind': 503,
'paying': 504,
'win': 505,
'moved': 506,
'ahead': 507,
'tickets': 508,
'sold': 509,
'1885': 510,
'legalized': 511,
'three': 512,
'thefootball': 513,
'leaguewas': 514,
'during': 515,
'season': 516,
'12': 517,
'interested': 518,
'consequently': 519,
'dominant': 520,
'after': 521,
'decades': 522,
```

```
'prague': 523,
'budapest': 524,
'sienna': 525,
'contenders': 526,
'dominance': 527,
'things': 528,
'excluded': 529,
'before': 530,
'play': 531,
'inverness': 532,
'milestones': 533,
'now': 534,
'follow': 535,
'challenge': 536,
'1871': 537,
'ended': 538,
'followed': 539,
'4': 540,
'hamilton': 541,
'crescent': 542,
'picture': 543,
'shows': 544,
'illustrations': 545,
'occasion': 546,
'twelve': 547,
'1883': 548,
'international': 549,
'four': 550,
'ireland': 551,
'wales': 552,
'phenomenon': 553,
'gradually': 554,
'european': 555,
'outside': 556,
'europe': 557,
'argentina': 558,
'1867': 559,
'foreign': 560,
'argentinean': 561,
'citizens': 562,
'founded': 563,
'1904': 564,
'foundation': 565,
'act': 566,
'signed': 567,
'representatives': 568,
'france': 569,
'belgium': 570,
'denmark': 571,
'netherlands': 572,
```

```
'spain': 573,
'sweden': 574,
'switzerland': 575,
'countriesâ\xa0did': 576,
'invented': 577,
'saw': 578,
'reason': 579,
'subordinate': 580,
'partake': 581,
'domestic': 582,
'mentioned': 583,
'based': 584,
'performance': 585,
'1908': 586,
'rank': 587,
'prestigious': 588,
'added': 589,
'1996': 590,
'male': 591,
'being': 592,
'present': 593,
'relatively': 594,
'example': 595,
'tennis': 596,
'traditionally': 597,
'mix': 598,
'britain': 599,
'andrew': 600,
'watsonis': 601,
'player': 602,
'he': 603,
'queenâ\x80\x99s': 604,
'show': 605,
'passion': 606,
'extent': 607,
'flocked': 608,
'shearing': 609,
'front': 610,
'television': 611,
'watching': 612,
'carefully': 613,
'great': 614,
'enthusiasm': 615,
'goodison': 616,
'purpose': 617,
'hosting': 618,
'1894': 619,
'final': 620,
'notts': 621,
'county': 622,
```

```
'bolton': 623,
'wanderers': 624,
'attended': 625,
'37': 626,
'milestone': 627,
'construction': 628,
'maracanã£': 629,
'imposing': 630,
'rio': 631,
'janeiro': 632,
'ready': 633,
'almost': 634,
'capacity': 635,
'host': 636,
'different': 637,
'fan': 638,
'tradition': 639,
'singing': 640,
'repertoire': 641,
'inspired': 642,
'pub': 643,
'songs': 644,
'americans': 645,
'adopt': 646,
'carnival': 647,
'style': 648,
'firecrackers': 649,
'fireworks': 650,
'phenomena': 651,
'bengali': 652,
'fires': 653,
'mixture': 654,
'summer': 655,
'measure': 656,
'itself': 657,
'thefifa': 658,
'edition': 659,
'uruguay': 660,
'exceptions': 661,
'due': 662,
'second': 663,
'war': 664,
'1991': 665,
'held': 666,
'thechampions': 667,
'1992': 668,
'formereuropean': 669,
'1955â\x80\x931991': 670,
'active': 671,
'each': 672,
```

```
'1870s': 673,
'211': 674,
'governing': 675,
'body': 676,
'proof': 677,
'globalization': 678,
'increase': 679,
'nations': 680,
'qualifiers': 681,
'32': 682,
'1934': 683,
'2014': 684,
'six': 685,
'confederations': 686,
'confã©dã©ration': 687,
'africaine': 688,
'caf': 689,
'asian': 690,
'afc': 691,
'union': 692,
'des': 693,
'europã©ennes': 694,
'uefa': 695,
'north': 696,
'central': 697,
'caribbean': 698,
'concacaf': 699,
'oceania': 700,
'ofc': 701,
'confederaciã³n': 702,
'sudamericana': 703,
'fãºtbol': 704,
'conmebol': 705,
'definitely': 706,
'quote': 707,
"goldblatt's": 708,
'bookthe': 709,
'roundpresent': 710,
'answer': 711,
'offers': 712,
'spotlight': 713,
'individual': 714,
'relishing': 715,
'defiance': 716,
'collective': 717,
'endeavor': 718,
'staged': 719,
'tragedy': 720,
'comedy': 721,
'epic': 722,
```

```
'pantomime': 723,
'unsophisticated': 724,
'music': 725,
'hall': 726,
'inaccessible': 727,
'experimental': 728,
'performances': 729,
'does': 730,
'imperious': 731,
'triumph': 732,
'lucky': 733,
'escapes': 734,
'impossible': 735,
'comebacks': 736,
'stubborn': 737,
'stalemates': 738,
'captures': 739,
'unpredictability': 740,
'uncertainty': 741,
'skill': 742,
'improvisation': 743,
'chance': 744,
'footballis': 745,
'"chess': 746,
'green': 747,
'pitch"': 748,
'united': 749,
'states': 750,
'canada': 751,
'socceris': 752,
'instead': 753,
'formal': 754,
'popular': 755,
'speech': 756,
'either': 757,
'›formations': 758,
'systems': 759,
'football›the': 760,
'evolution': 761,
'shoes›the': 762,
'field': 763,
'dimensions›history': 764,
'stadiums›the': 765,
'languages›collection': 766,
'links': 767,
'sites': 768,
'web': 769,
'references': 770,
'encyclopediahistory': 771,
'beautiful': 772,
```

```
 '2002': 773,
 'documentary': 774,
 'series': 775,
 'footballâ\x80\x93': 776,
 'goldblatt': 777,
 '2008': 778,
 'https': 779,
 'en': 780,
 'wikipedia': 781,
 'wiki': 782,
 'clubshttp': 783,
 'fblack': 784,
 'fstadiums': 785,
 'image': 786,
 'sources': 787,
 'andrea': 788,
 'scoto': 789,
 'itiillustration': 790,
 'da': 791,
 'francesco': 792,
 'bertelli': 793,
 'paduawilliam': 794,
 'ralston': 795,
 '1848â\x80\x931911': 796,
 'scanned': 797,
 'bryon': 798,
 'butler': 799,
 'queen': 800,
 'ane': 801,
 'pressdocumentary': 802,
 'filmalegria': 803,
 'do': 804,
 'povo': 805,
 '1963': 806}

input_sequences = []
for sentence in football_history.split('\n'):
    token_list = tokenizer.texts_to_sequences([sentence])[0]
    for i in range(1, len(token_list)):
        sequence = token_list[:i+1]
        input_sequences.append(sequence)

input_sequences[:5]

[[259, 8],
 [259, 8, 260],
 [259, 8, 260, 153],
 [259, 8, 260, 153, 7],
 [259, 8, 260, 153, 7, 261]]
```

Get the Maximum length array size

```python
max_length = max([len(input_sequence) for input_sequence in input_sequences])
max_length
```

337

add padding

```python
input_sequences =
np.array(tf.keras.preprocessing.sequence.pad_sequences(input_sequences,
maxlen=max_length, padding='pre'))
```

divide dataset into feature set (X) and labels (y)

```python
X = input_sequences[:, :-1]
y = input_sequences[:, -1]
```

number of Unique words

```python
num_classes = len(tokenizer.word_index)+1
```

one-hot encoding labels

```python
y = np.array(tf.keras.utils.to_categorical(y, num_classes=num_classes))
```

model building

```python
model = tf.keras.models.Sequential()
model.add(tf.keras.layers.Embedding(num_classes, 80, input_length=max_length-1))
model.add(tf.keras.layers.LSTM(100))
model.add(tf.keras.layers.Dense(num_classes, activation='softmax'))


# num_classes = total unique words or tokens in your dataset
# 80 = embedding dimension
# max_length-1 = input_length
# 100 = number of neurons in LSTM layer
# softmax = activation function for multi-class classification

model.summary()
```

Model: "sequential_1"

_____

Layer (type)	Output Shape	Param #
===========================	====================	=========
embedding_1 (Embedding)	(None, 336, 80)	64560
lstm_1 (LSTM)	(None, 100)	72400
dense_1 (Dense)	(None, 807)	81507

```
===============================================================
Total params: 218467 (853.39 KB)
Trainable params: 218467 (853.39 KB)
Non-trainable params: 0 (0.00 Byte)
_____

# compile the model
model.compile(loss='categorical_crossentropy',
              optimizer='adam',
              metrics=['accuracy'])

# model training
epoch = 100
history = model.fit(X, y, epochs=epoch)

Epoch 1/100
74/74 [==============================] - 30s 247ms/step - loss: 6.2029 -
accuracy: 0.0851
Epoch 2/100
74/74 [==============================] - 19s 254ms/step - loss: 5.7092 -
accuracy: 0.0911
Epoch 3/100
74/74 [==============================] - 19s 253ms/step - loss: 5.6232 -
accuracy: 0.0911
Epoch 4/100
74/74 [==============================] - 23s 314ms/step - loss: 5.5409 -
accuracy: 0.0911
Epoch 5/100
74/74 [==============================] - 24s 320ms/step - loss: 5.4467 -
accuracy: 0.0911
Epoch 6/100
74/74 [==============================] - 25s 336ms/step - loss: 5.3268 -
accuracy: 0.1064
Epoch 7/100
74/74 [==============================] - 25s 335ms/step - loss: 5.1849 -
accuracy: 0.1281
Epoch 8/100
74/74 [==============================] - 34s 462ms/step - loss: 5.0496 -
accuracy: 0.1417
Epoch 9/100
74/74 [==============================] - 26s 355ms/step - loss: 4.9251 -
accuracy: 0.1468
Epoch 10/100
74/74 [==============================] - 24s 330ms/step - loss: 4.8075 -
accuracy: 0.1540
Epoch 11/100
74/74 [==============================] - 26s 351ms/step - loss: 4.6916 -
accuracy: 0.1587
Epoch 12/100
74/74 [==============================] - 21s 278ms/step - loss: 4.5798 -
```

```
accuracy: 0.1689
Epoch 13/100
74/74 [==============================] - 23s 306ms/step - loss: 4.4623 -
accuracy: 0.1834
Epoch 14/100
74/74 [==============================] - 24s 319ms/step - loss: 4.3518 -
accuracy: 0.1970
Epoch 15/100
74/74 [==============================] - 24s 327ms/step - loss: 4.2364 -
accuracy: 0.2140
Epoch 16/100
74/74 [==============================] - 19s 261ms/step - loss: 4.1199 -
accuracy: 0.2268
Epoch 17/100
74/74 [==============================] - 24s 319ms/step - loss: 4.0108 -
accuracy: 0.2345
Epoch 18/100
74/74 [==============================] - 21s 280ms/step - loss: 3.8955 -
accuracy: 0.2477
Epoch 19/100
74/74 [==============================] - 24s 318ms/step - loss: 3.7838 -
accuracy: 0.2613
Epoch 20/100
74/74 [==============================] - 18s 247ms/step - loss: 3.6759 -
accuracy: 0.2715
Epoch 21/100
74/74 [==============================] - 18s 241ms/step - loss: 3.5685 -
accuracy: 0.2838
Epoch 22/100
74/74 [==============================] - 20s 277ms/step - loss: 3.4579 -
accuracy: 0.3013
Epoch 23/100
74/74 [==============================] - 22s 294ms/step - loss: 3.3513 -
accuracy: 0.3136
Epoch 24/100
74/74 [==============================] - 21s 280ms/step - loss: 3.2451 -
accuracy: 0.3247
Epoch 25/100
74/74 [==============================] - 21s 285ms/step - loss: 3.1392 -
accuracy: 0.3455
Epoch 26/100
74/74 [==============================] - 19s 261ms/step - loss: 3.0391 -
accuracy: 0.3634
Epoch 27/100
74/74 [==============================] - 18s 247ms/step - loss: 2.9336 -
accuracy: 0.3779
Epoch 28/100
74/74 [==============================] - 27s 365ms/step - loss: 2.8364 -
accuracy: 0.4106
Epoch 29/100
```

```
74/74 [==============================] - 24s 315ms/step - loss: 2.7351 -
accuracy: 0.4260
Epoch 30/100
74/74 [==============================] - 21s 278ms/step - loss: 2.6387 -
accuracy: 0.4468
Epoch 31/100
74/74 [==============================] - 22s 292ms/step - loss: 2.5451 -
accuracy: 0.4647
Epoch 32/100
74/74 [==============================] - 19s 257ms/step - loss: 2.4504 -
accuracy: 0.4936
Epoch 33/100
74/74 [==============================] - 24s 321ms/step - loss: 2.3601 -
accuracy: 0.5115
Epoch 34/100
74/74 [==============================] - 20s 271ms/step - loss: 2.2713 -
accuracy: 0.5387
Epoch 35/100
74/74 [==============================] - 24s 322ms/step - loss: 2.1819 -
accuracy: 0.5562
Epoch 36/100
74/74 [==============================] - 19s 260ms/step - loss: 2.0973 -
accuracy: 0.5779
Epoch 37/100
74/74 [==============================] - 21s 278ms/step - loss: 2.0184 -
accuracy: 0.5902
Epoch 38/100
74/74 [==============================] - 19s 258ms/step - loss: 1.9334 -
accuracy: 0.6132
Epoch 39/100
74/74 [==============================] - 19s 261ms/step - loss: 1.8582 -
accuracy: 0.6196
Epoch 40/100
74/74 [==============================] - 22s 302ms/step - loss: 1.7838 -
accuracy: 0.6511
Epoch 41/100
74/74 [==============================] - 21s 282ms/step - loss: 1.7096 -
accuracy: 0.6600
Epoch 42/100
74/74 [==============================] - 18s 242ms/step - loss: 1.6413 -
accuracy: 0.6809
Epoch 43/100
74/74 [==============================] - 18s 245ms/step - loss: 1.5721 -
accuracy: 0.6940
Epoch 44/100
74/74 [==============================] - 24s 319ms/step - loss: 1.5071 -
accuracy: 0.7115
Epoch 45/100
74/74 [==============================] - 23s 313ms/step - loss: 1.4439 -
accuracy: 0.7204
```
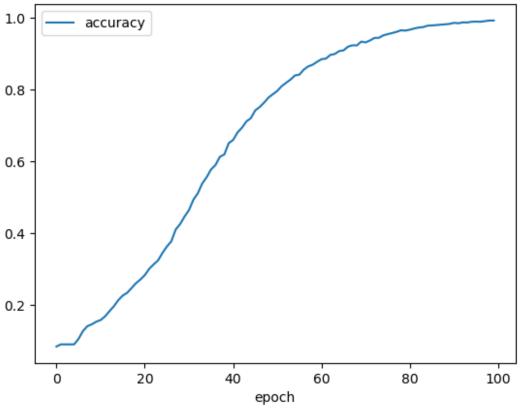
```
Epoch 46/100
74/74 [==============================] - 21s 286ms/step - loss: 1.3808 -
accuracy: 0.7421
Epoch 47/100
74/74 [==============================] - 19s 259ms/step - loss: 1.3266 -
accuracy: 0.7515
Epoch 48/100
74/74 [==============================] - 20s 273ms/step - loss: 1.2693 -
accuracy: 0.7638
Epoch 49/100
74/74 [==============================] - 19s 252ms/step - loss: 1.2176 -
accuracy: 0.7779
Epoch 50/100
74/74 [==============================] - 18s 248ms/step - loss: 1.1696 -
accuracy: 0.7872
Epoch 51/100
74/74 [==============================] - 20s 273ms/step - loss: 1.1202 -
accuracy: 0.7966
Epoch 52/100
74/74 [==============================] - 23s 316ms/step - loss: 1.0722 -
accuracy: 0.8098
Epoch 53/100
74/74 [==============================] - 24s 321ms/step - loss: 1.0303 -
accuracy: 0.8191
Epoch 54/100
74/74 [==============================] - 26s 346ms/step - loss: 0.9875 -
accuracy: 0.8281
Epoch 55/100
74/74 [==============================] - 19s 261ms/step - loss: 0.9455 -
accuracy: 0.8396
Epoch 56/100
74/74 [==============================] - 21s 287ms/step - loss: 0.9081 -
accuracy: 0.8417
Epoch 57/100
74/74 [==============================] - 19s 263ms/step - loss: 0.8692 -
accuracy: 0.8557
Epoch 58/100
74/74 [==============================] - 23s 307ms/step - loss: 0.8335 -
accuracy: 0.8647
Epoch 59/100
74/74 [==============================] - 23s 315ms/step - loss: 0.8041 -
accuracy: 0.8694
Epoch 60/100
74/74 [==============================] - 23s 312ms/step - loss: 0.7657 -
accuracy: 0.8774
Epoch 61/100
74/74 [==============================] - 21s 284ms/step - loss: 0.7371 -
accuracy: 0.8847
Epoch 62/100
74/74 [==============================] - 17s 230ms/step - loss: 0.7071 -
```

```
accuracy: 0.8864
Epoch 63/100
74/74 [==============================] - 17s 230ms/step - loss: 0.6773 -
accuracy: 0.8966
Epoch 64/100
74/74 [==============================] - 21s 280ms/step - loss: 0.6504 -
accuracy: 0.8991
Epoch 65/100
74/74 [==============================] - 21s 280ms/step - loss: 0.6258 -
accuracy: 0.9072
Epoch 66/100
74/74 [==============================] - 19s 255ms/step - loss: 0.5992 -
accuracy: 0.9094
Epoch 67/100
74/74 [==============================] - 19s 257ms/step - loss: 0.5736 -
accuracy: 0.9196
Epoch 68/100
74/74 [==============================] - 17s 225ms/step - loss: 0.5491 -
accuracy: 0.9234
Epoch 69/100
74/74 [==============================] - 17s 227ms/step - loss: 0.5278 -
accuracy: 0.9230
Epoch 70/100
74/74 [==============================] - 23s 315ms/step - loss: 0.5079 -
accuracy: 0.9336
Epoch 71/100
74/74 [==============================] - 18s 244ms/step - loss: 0.4866 -
accuracy: 0.9315
Epoch 72/100
74/74 [==============================] - 19s 259ms/step - loss: 0.4664 -
accuracy: 0.9370
Epoch 73/100
74/74 [==============================] - 19s 252ms/step - loss: 0.4447 -
accuracy: 0.9438
Epoch 74/100
74/74 [==============================] - 25s 335ms/step - loss: 0.4275 -
accuracy: 0.9443
Epoch 75/100
74/74 [==============================] - 19s 256ms/step - loss: 0.4093 -
accuracy: 0.9511
Epoch 76/100
74/74 [==============================] - 18s 245ms/step - loss: 0.3949 -
accuracy: 0.9545
Epoch 77/100
74/74 [==============================] - 23s 307ms/step - loss: 0.3771 -
accuracy: 0.9574
Epoch 78/100
74/74 [==============================] - 42s 574ms/step - loss: 0.3611 -
accuracy: 0.9609
Epoch 79/100
```

```
74/74 [==============================] - 49s 659ms/step - loss: 0.3460 -
accuracy: 0.9651
Epoch 80/100
74/74 [==============================] - 36s 484ms/step - loss: 0.3317 -
accuracy: 0.9643
Epoch 81/100
74/74 [==============================] - 31s 422ms/step - loss: 0.3176 -
accuracy: 0.9668
Epoch 82/100
74/74 [==============================] - 25s 336ms/step - loss: 0.3045 -
accuracy: 0.9702
Epoch 83/100
74/74 [==============================] - 23s 305ms/step - loss: 0.2920 -
accuracy: 0.9728
Epoch 84/100
74/74 [==============================] - 24s 323ms/step - loss: 0.2793 -
accuracy: 0.9740
Epoch 85/100
74/74 [==============================] - 26s 353ms/step - loss: 0.2683 -
accuracy: 0.9779
Epoch 86/100
74/74 [==============================] - 24s 319ms/step - loss: 0.2569 -
accuracy: 0.9787
Epoch 87/100
74/74 [==============================] - 24s 332ms/step - loss: 0.2442 -
accuracy: 0.9796
Epoch 88/100
74/74 [==============================] - 26s 351ms/step - loss: 0.2372 -
accuracy: 0.9809
Epoch 89/100
74/74 [==============================] - 20s 269ms/step - loss: 0.2259 -
accuracy: 0.9817
Epoch 90/100
74/74 [==============================] - 19s 263ms/step - loss: 0.2151 -
accuracy: 0.9830
Epoch 91/100
74/74 [==============================] - 21s 282ms/step - loss: 0.2075 -
accuracy: 0.9860
Epoch 92/100
74/74 [==============================] - 22s 294ms/step - loss: 0.1995 -
accuracy: 0.9847
Epoch 93/100
74/74 [==============================] - 21s 281ms/step - loss: 0.1889 -
accuracy: 0.9872
Epoch 94/100
74/74 [==============================] - 22s 291ms/step - loss: 0.1803 -
accuracy: 0.9868
Epoch 95/100
74/74 [==============================] - 23s 309ms/step - loss: 0.1723 -
accuracy: 0.9889
```

```
Epoch 96/100
74/74 [==============================] - 21s 285ms/step - loss: 0.1661 -
accuracy: 0.9894
Epoch 97/100
74/74 [==============================] - 18s 245ms/step - loss: 0.1593 -
accuracy: 0.9889
Epoch 98/100
74/74 [==============================] - 19s 253ms/step - loss: 0.1523 -
accuracy: 0.9906
Epoch 99/100
74/74 [==============================] - 18s 248ms/step - loss: 0.1457 -
accuracy: 0.9923
Epoch 100/100
74/74 [==============================] - 18s 249ms/step - loss: 0.1391 -
accuracy: 0.9923
```

```python
plt.plot(history.history['accuracy'],label="accuracy")
plt.legend()
plt.title('model accuracy')
plt.xlabel('epoch')
plt.show()
```
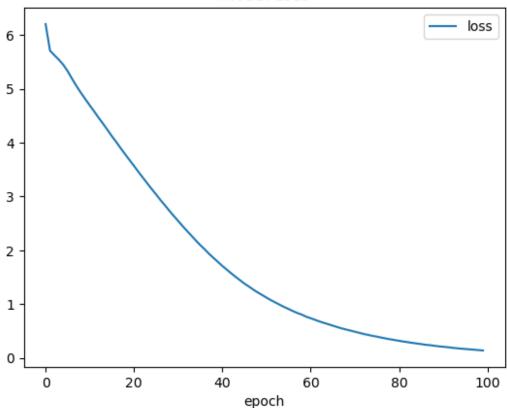


```python
plt.plot(history.history['loss'],label="loss")
plt.legend()
```

```
plt.title('model Loss')
plt.xlabel('epoch')
plt.show()
```

model Loss

save the model
```
model.save("Next_word_Prediction.h5")
```

```
c:\Users\Admin\anaconda3\lib\site-packages\keras\src\engine\training.py:3103:
UserWarning: You are saving your model as an HDF5 file via `model.save()`.
This file format is considered legacy. We recommend using instead the native
Keras format, e.g. `model.save('my_model.keras')`.
  saving_api.save_model(
```

Load the save model
```
final_model = tf.keras.models.load_model("./Next_word_Prediction.h5")
```

Prediction Phase
```
no_of_words_predict = 6
input_text = "Football clubs"
for i in range(no_of_words_predict):
    token_list = tokenizer.texts_to_sequences([input_text])[0]
    token_list = tf.keras.preprocessing.sequence.pad_sequences([token_list],
maxlen=max_length-1, padding='pre')
    predicted = np.argmax(final_model.predict(token_list), axis=-1)
```

```python
        output_word = ""
        for word, index in tokenizer.word_index.items():
            if index == predicted:
                output_word = word
                break
        input_text += " " + output_word

print(input_text)
```

```
1/1 [==============================] - 1s 1s/step
1/1 [==============================] - 0s 75ms/step
1/1 [==============================] - 0s 95ms/step
1/1 [==============================] - 0s 91ms/step
1/1 [==============================] - 0s 85ms/step
1/1 [==============================] - 0s 81ms/step
Football clubs have existed since the 15th century
```