

Names: Ghaida Alshiddi & Bushra Rahman

Class: CS 4395.001

Assignment: Web Crawler

Knowledge Base

This program is a web crawler that started with an [encyclopedia.com](https://www.encyclopedia.com) link about the FIFA World Cup. From there, it crawled the web for related URLs, scraped and cleaned text off those links, and calculated the most frequently occurring terms. From those terms, 10 were chosen, and a knowledge base was created of those terms. The knowledge base is a pickled Python dictionary where the top 10 terms are the keys and each of their values is a list of sentences containing the term. Here are screenshots of the knowledge base.

Example output for “world”:

```
Knowledge Base:
{'world': ['Soccer is played in every corner of the world and inspires its own frenzy every four years with the convening of the international championship known as the World Cup.', 'A testament to the status of soccer as the world's most popular sport, FIFA is unquestionably the most powerful and all-encompassing world sports body.', 'The first World Cup championship took place in Uruguay in 1930, and, other than the intervention of World War II, the World Cup has been held every four years since that time.FIFA is responsible for many aspects of the governance of international soccer, including the sanction of a multitude of age group championships for both men and women, the maintenance of a world-ranking system, the resolution of issues pertaining to the nationality of players, and the annual review of the rules of the game.', 'The greatest ongoing responsibility undertaken by this organization is the organizing of the World Cup. Every member country
```

Example output for “cup”:

```
NEARBY TERMS FIBA fib.", 'cup': ['https://www.encyclopedia.com/sports/sports-fitness-recreation-and-leisure-magazines/fifa-world-cup-soccer "FIFA: World Cup', 'Soccer ."', 'Retrieved February 23, 2023 from Encyclopedia.com: https://www.encyclopedia.com/sports/sports-fitness-recreation-and-leisure-magazines/fifa-world-cup-soccer Learn more about citation styles Citation styles Encyclopedia.com gives you the ability to cite reference entries and articles according to common styles from the Modern Language Association (MLA), The Chicago Manual of Style, and the American Psychological Association (APA).', 'https://www.encyclopedia.com/media/encyclopedias-almanacs-transcripts-and-maps/world-cup "World Cup ."', 'Retrieved February 22, 2023 from Encyclopedia.com: https://www.encyclopedia.com/media/encyclopedias-almanacs-transcripts-and-maps/world-cup Learn more about citation styles Citation styles Encyclopedia.com gives you the ability to cite reference entries and articles according to common styles from the Modern Language Association (MLA), The Chicago Manual of Style, and the American Psychological Association (APA).', 'There are also world cup competitions in other sports, notably rugby, cricket and hockey.', 'https://www.encyclopedia.com/environment/encyclopedias-almanacs-transcripts-and-maps/world-cup "World Cup ."', 'Retrieved February 22, 2023 from Encyclopedia.com: https://www.encyclopedia.com/environment/encyclopedias-almanacs-transcripts-and-maps/world-cup Learn more about citation styles Citation styles Encyclopedia.com gives you the ability to cite reference entries and articles according to common styles from the Modern Language Association (MLA), The Chicago Manual of Style, and the American Psychological Association (APA).', 'He rode the bench for the first two games to recuperate from a knee injury he had suffered in a qualifying game.', 'Doctors refused him permission to play, but Brazil was so strong it won the cup without him.', 'Wall Street Journal (January 3, 2001): B.8A0ther2002 FIFA World Cup Korea Japan, Luis Figo Profile, RealMadrid.com, http://fifaworldcup.yahoo.com (May 31, 2002).', 'CNN-Sports Illustrated, http://sportsillustrated.cnn.com/soccer/world/2002/world_cup/news/2002/06/04/us_portugal_gamer/ (June 5, 2002).Ball, Phil.', 'USA Today, http://www.usatoday.com/sports/soccer/cup2002/games_2002-06-05-usaportugal.htm (June 6, 2002).Sketch by Paul Burton Notable Sports Figures Burton, Paul × Cite this article Pick a style below, and copy the text for your bibliography.', 'Some batsmen are required to occupy the bowler, especially if partnered with an adept batsman.', 'Rimet was the organization's first genuinely influential leader and its longest serving president in the 20th century, occupying the post of president until 1954.'],
```

Example output for “soccer”:

```
20th century, occupying the post of president until 1954.'], 'soccer': ['There is no competition on Earth that mobilizes such passion or that is as inclusive as the qualification for this event.The World Cup is entirely controlled through the international governing body for the sport of soccer, FIFA, the French language acronym for the Federation Internationale de Football Association.', 'A testament to the status of soccer as the world's most popular sport, FIFA is unquestionably the most powerful and all-encompassing world sports body.', 'Unlike other global sports such as basketball, where there is a lack of international unanimity regarding the rules of the sport, FIFA, through both custom and stature, reigns supreme.The first organized soccer governance occurred in 1863, with the founding of the Football Association in England.', 'The first World Cup championship took place in Uruguay in 1930, and, other than the intervention of World War II, the World Cup has been held every four years since that time.FIFA is responsible for many aspects of the governance of international soccer, including the
```

Example output for “rugby”:

```
of millions, soccer is "the beautiful game."'], 'rugby': ['There are also world cup competitions in other sports, notably rugby, cricket and hockey.', 'The rules of rugby were not entirely formalized until approximately 1845, and by 1871, an association known as the Rugby Football Union was created and the rules of the sport, were codified.Rugby, or a game very similar to it, was first exported to North America by the British soldiers of the Quebec military garrison in the 1860s.', 'Through games played against McGill University of Montreal, rugby became popular with the universities of the northeastern United States.', 'The American form of rugby evolved once more into the sport now known as American football.', 'Rugby, cricket, cross-country running, and soccer (in its organized form) represent four great English sports exports of the late 1880s.It is said that the essential difference between soccer and rugby may be stated as a credo, that "soccer is a gentleman's game played by toughs, and rugby is a tough's game, played by gentlemen."', 'Rugby permits and encourages significant physical contact, and one of the noteworthy features of the sport is that unlike American football and its specialized play, every player on the rugby field must have a basic command of all physical aspects of the sport: running, tackling, passing, kicking, and carrying the ball.The basic rules of rugby are that the game is most commonly played with 15 players to a side; the game of seven per side employs the same rules of play regarding scoring and physical contact, with necessary modifications.', 'The rugby field (pitch) is 110 yd (100 m) from goal to goal, and a maximum of 75 yd (68 m) wide; the area behind the goal is a maximum of 24 yd deep (20 m).', 'Due to the nature of the sport and the fact that the players wear very little protective equipment, natural grass surfaces are preferred.The rugby ball is a rounded oblong shape, approximately 12 in (between 280 cm and 300 cm) long and 26 in (620 cm) in circumference; the chief objects of the game are to either advance the ball across, the opponent's goal line for a "try," worth four points, or to kick the ball through the uprights of the opponent's goal for either a penalty, worth three points, or a drop goal, also worth three points.', 'The shape of a rugby ball makes it conducive to the dropkick.The 15 players play positions that are divided into two general groups, determined by their respective roles on the field.', 'The player responsible for much of the coordination of the rugby team's attack is the "scrum half"; the fastest player on a rugby team is often the wing, who plays on the outside of the team formation.The players are permitted to advance the ball towards the opposing goal line by running with the ball, kicking the ball, and passing the ball, so long as their teammate is behind them, as no player can be ahead of a teammate with the ball, the concept known as being onside.', 'The line of scrummage, the imaginary boundary between the two scrums, became the well-known expression in the line play in American football that is similar to the formation of the two rugby scrums.When the ball goes out of bounds, the ball is returned to the field of play through a formation known as a "line out," where the forwards from each team form a line facing the in-bounding player, who throws the ball down the line, where the forwards attempt to either catch the ball or tip it to one of the teammates.Although a rough and often fiercely physical game, rugby has strict rules regarding tackling.', 'If ejected, a player may not be replaced for the balance of the game.As with other English sport exports, rugby first flourished in the countries of the former British Empire.', 'In addition to American football, Australian Rules Football is also a derivative of rugby.', 'Rugby League, in contrast to the international game that is sometime referred to Rugby Union, is a form of professional rugby popular in Great Britain, with 13 players per side and a modified scoring system for tries, penalties, and convert kicks.Rugby first became prominent as an international sport through elite regional championships.', 'Most of the world's competitive players tend to belong to rugby club teams, affiliated with a university or an adult volunteer-based amateur organization.The Tri Nations Cup became an annual fixture between the rugby powers of Australia, New Zealand, and South Africa in 1996.', 'In addition to these formal regional
```

Example output for “basketball”:

```
the international championships.", 'Invented by Canadian James Naismith in 1891 in Springfield, Massachusetts, basketball was a pure creation and not a derivative game or a long-evolving sport in the manner of soccer, cricket, rugby, and baseball.'], 'basketball': ['Unlike other global sports such as basketball, where there is a lack of international unanimity regarding the rules of the sport, FIFA, through both custom and stature, reigns supreme.The first organized soccer governance occurred in 1863, with the founding of the Football Association in England.', 'In the 1950s baseball was still supreme, American football began its rise to prominence, and in the 1960s ice hockey and
```

Example output for “game”:

```
Language Association (MLA), The Chicago Manual of Style, and the American Psychological Association (APA).'], 'game': ['FIFA: World Cup Soccer | Encyclopedia.com Skip to main content EXPLORE EXPLORE Earth and Environment History Literature and the Arts Medicine People Philosophy and Religion Places Plants and Animals Science and Technology Social Sciences and the Law Sports and Everyday Life Additional References Articles Daily Sports Sports fitness recreation and leisure magazines FIFA: World Cup Soccer FIFA: World Cup Soccer gale views updated FIFA: World Cup SoccerSoccer, known as football, outside of North America, is the "beautiful game" to its multitudes of fans.', "FIFA's membership and authority over the game increased each year thereafter.", 'The
```

Example output for “players”:

```
hundreds of millions, soccer is "the beautiful game."'], 'players': ['The first World Cup championship took place in Uruguay in 1930, and, other than the intervention of World War II, the World Cup has been held every four years since that time.FIFA is responsible for many aspects of the governance of international soccer, including the sanction of a multitude of age group championships for both men and women, the maintenance of a world-ranking system, the resolution of issues, pertaining to the nationality of players, and the annual review of the rules of the game.', 'However, the competition did not materialize for more than twenty years because of conflict among national federations over whether to allow only amateur players to compete, as in the Olympic Games, or to accept professional players, who were becoming more prevalent in Europe.', 'Finally, FIFA agreed to include professional players and to hold the tournament every four years, alternating with the Olympic Games.FIFA selected Uruguay to host the first ever World Cup finals in 1930.', 'Thus, the U.S. teams that competed in the World Cups of, 1930, 1934, and 1950 consisted largely of immigrant players.After World War II, soccer became the most popular sport on the planet.', 'Rugby permits and encourages,
```

Example output for “cricket”:

```
Basketball; International federations; National Collegiate Athletic Association (NCAA).'], 'cricket': ['There are also world cup competitions in other sports, notably rugby, cricket and hockey.', 'Rugby, cricket, cross-country running, and soccer (in its organized form) represent four great English sports exports of the late 1880s.It is said that the essential difference between soccer and rugby may be stated as a credo, that "soccer is a gentleman's game played by toughs, and rugby is a tough's game, played by gentlemen."', 'Although baseball derived from Britain's cricket and rounders, boosters utilized the Abner Doubleday myth to promote baseball as a native-born sport, nothing less than the "national pastime."', 'Among those who are a part of more action-packed athletic traditions, cricket is variously seen as a boring, tedious game.', 'To its hundreds of millions of fans united in an intense global following, cricket is a truly international sport, one of immeasurable subtlety, sportsmanship, and athletic skill.', 'To play cricket as a representative of one's country is to achieve celebrity status in countries, such as India, Pakistan, the various West Indies, and South Africa.Cricket was a product of the English countryside, an ancient game that was played in a formalized fashion at least as early as the 1500s.', 'The rules of the game were first codified in 1744; the format of cricket has been only modified, as opposed to, being subjected to wholesale reconstruction, since that time.', 'As with American baseball, the fundamental distinction between cricket and virtually all other sports is the fact that traditionally there was no time limit imposed on play; the length of the game was determined by how long it took one team to retire the other side in their turn at bat, known as their "innings."', 'Ironically, the first international cricket match was played between two countries with a more limited, current cricket tradition, when Canada played the United States in 1844.Cricket is a predominately, but not exclusively, male sport.', 'Women's cricket enjoys a
```

Example output for “football”:

```

pure creation and not a derivative game or a long-evolving sport in the manner of soccer, cricket, rugby, and baseball.'], 'football': ['FIFA: World Cup Soccer | >
Encyclopedia.com Skip to main content EXPLORE EXPLORE Earth and Environment History Literature and the Arts Medicine People Philosophy and >
Religion Places Plants and Animals Science and Technology Social Sciences and the Law Sports and Everyday Life Additional References Articles Daily Sports >
Sports fitness recreation and leisure magazines FIFA: World Cup Soccer FIFA: World Cup Soccer gale views updated FIFA: World Cup SoccerSoccer, known >
as football outside of North America, is the "beautiful game" to its multitudes of fans.', 'Unlike sports such as American football or baseball, FIFA is extremely >
particular about the nature and quality of the surfaces on which its competitions are held.', 'More From encyclopedia.com World Cup , The World Cup of >
Football, or soccer as the game is called in the United States, is the most popular sporting event in the world.', "Among those who are a part of more action-packed,

```

Example output for “sports”:

```

or soccer as the game is called in the United States, is the most popular sporting event in the world.'], 'sports': ["A testament to the status of soccer as the >
world's most popular sport, FIFA is unquestionably the most powerful and all-encompassing world sports body.", 'Unlike other global sports such as basketball, where, >
there is a lack of international unanimity regarding the rules of the sport, FIFA, through both custom and stature, reigns supreme.The first organized soccer >
governance occurred in 1863, with the founding of the Football Association in England.', 'Unlike sports such as American football or baseball, FIFA is extremely >
particular about the nature and quality of the surfaces on which its competitions are held.', 'https://www.encyclopedia,

```

Chatbot Dialog

The following is an example conversation with a chatbot built using this knowledge base:

Hello, I’m a chatbot, and I like talking about sports! Especially soccer. What’s your name?

I’m Ghaida

Nice to meet you Ghaida. Do you like watching sports?

Yes, I like watching soccer.

I like soccer too! Do you enjoy watching the FIFA World Cups?

Yes! I’ve been watching them since I was a kid.

Me too! Did you watch the last World Cup?

Yes, I’m so happy that Argentina won!

Me too! Other than soccer, do you have any favorite sports or games?

Umm, not really. But I do like to play basketball.

Wow, that’s fun! I also like basketball and many other sports such as rugby, cricket, and football! Do you have any favorite basketball players?

I don’t really watch basketball games, but I’d say Michael Jordan.

Same! Is there anything else you’d like to ask me?

No

Nice chatting with you, Ghaida. See you soon!

Bye