

Web Crawler Report
Shreya Valaboju, Soham Mukherjee
CS 4395.001

7. In a doc: (1) describe how you created your knowledge base, include screenshots of the knowledge base, and indicate your top 10 terms;

Our top 10 terms were prominent basketball players for the Dallas Mavericks. These were:

- 1: doncic
- 2: irving
- 3: wood
- 4: bertans
- 5: powell
- 6: bullock
- 7: kleber
- 8: ntilikina
- 9: mcgee
- 10: pinson

We wanted to create a knowledge base out of these players as relevant statistics on them would probably be most relevant to someone who wants to use the chatbot. We created this knowledge base by scraping the urls for each player's profile. Here is a screenshot of the unpickled data:

```
"Luka_Doncic": {
  "name": "Luka Doncic",
  "Position": "Point guard / Shooting guard",
  "League": "NBA",
  "Born": "(1999-02-28) February 28, 1999 (age 24)Ljubljana, Slovenia",
  "Listed height": "6 ft 7 in (2.01 m)",
  "Listed weight": "230 lb (104 kg)",
  "NBA draft": "2018 / Round: 1 / Pick: 3rd overall",
  "Playing career": "2015-present",
  "2015-2018": "Real Madrid",
  "2018-present": "Dallas Mavericks",
  "Men's basketball": "Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey"
},
```

Here, we can see a JSON object for each player, which has relevant metrics and statistics such as Name, Position, League, etc. Each JSON dictionary object is printed out in the program as such:

```
Knowledge Base:
{'name': 'Luka Doncic', 'Position': 'Point guard / Shooting guard', 'League': 'NBA', 'Born': '(1999-02-28) February 28, 1999 (age 24)Ljubljana, Slovenia', 'Listed height': '6 ft 7 in (2.01 m)', 'Listed weight': '230 lb (104 kg)', 'NBA draft': '2018 / Round: 1 / Pick: 3rd overall', 'Playing career': '2015-present', '2015-2018': 'Real Madrid', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Kyrie Irving', 'Position': 'Point guard', 'League': 'NBA', 'Born': '(1992-03-23) March 23, 1992 (age 30)Melbourne, Victoria, Australia', 'Nationality': 'American'}
{'name': 'Christian Wood', 'Position': 'Power forward', 'League': 'NBA', 'Born': '(1995-09-27) September 27, 1995 (age 27)Long Beach, California U.S.', 'Listed height': '6 ft 7 in (2.01 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2015 / Round: 1 / Pick: 15th overall', 'Playing career': '2015-present', '2015-2018': 'Los Angeles Lakers', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Davis Bertans', 'Position': 'Power forward', 'League': 'NBA', 'Born': '(1992-11-12) 12 November 1992 (age 30)Valmiera, Latvia', 'Listed height': '6 ft 9 in (2.06 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2015 / Round: 1 / Pick: 15th overall', 'Playing career': '2015-present', '2015-2018': 'Dallas Mavericks', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Dwight Powell', 'Position': 'Center / Power forward', 'League': 'NBA', 'Born': '(1991-07-20) July 20, 1991 (age 31)Toronto, Ontario, Canada', 'Listed height': '6 ft 10 in (2.08 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2013 / Round: 1 / Pick: 15th overall', 'Playing career': '2013-present', '2013-2018': 'Dallas Mavericks', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Reggie Bullock', 'Position': 'Small forward / Shooting guard', 'League': 'NBA', 'Born': '(1991-03-16) March 16, 1991 (age 31)Baltimore, Maryland, U.S.', 'Listed height': '6 ft 6 in (1.98 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2013 / Round: 1 / Pick: 15th overall', 'Playing career': '2013-present', '2013-2018': 'Dallas Mavericks', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Maxi Kleber', 'Position': 'Power forward', 'League': 'NBA', 'Born': '(1992-01-29) 29 January 1992 (age 31)Wurzburg, Germany', 'Listed height': '6 ft 10 in (2.08 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2018 / Round: 1 / Pick: 3rd overall', 'Playing career': '2018-present', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Frank Ntilikina', 'Position': 'Point guard', 'League': 'NBA', 'Born': '(1998-07-28) 28 July 1998 (age 24)Ixelles, Belgium', 'Nationality': 'French', 'Listed height': '6 ft 6 in (1.98 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2018 / Round: 1 / Pick: 3rd overall', 'Playing career': '2018-present', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'JaVale McGee', 'Position': 'Center', 'League': 'NBA', 'Born': '(1988-01-19) January 19, 1988 (age 35)Flint, Michigan, U.S.', 'Listed height': '7 ft 6 in (2.29 m)', 'Listed weight': '260 lb (118 kg)', 'NBA draft': '2010 / Round: 1 / Pick: 15th overall', 'Playing career': '2010-present', '2010-2018': 'Dallas Mavericks', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
{'name': 'Theo Pinson', 'Position': 'Shooting guard / Small forward', 'League': 'NBA', 'Born': '(1995-11-05) November 5, 1995 (age 27)Greensboro, North Carolina', 'Listed height': '6 ft 6 in (1.98 m)', 'Listed weight': '215 lb (97 kg)', 'NBA draft': '2018 / Round: 1 / Pick: 3rd overall', 'Playing career': '2018-present', '2018-present': 'Dallas Mavericks', 'Men's basketball': 'Medals~Men's basketball~Representing Slovenia~EuroBasket~2017 Turkey'}
```

