**Uniwersytet Rzeszowski**

**Wydział Nauk Ścisłych i Technicznych**

**Instytut Informatyki**

**Krzysztof Motas**

***Analiza piłkarzy nominowanych do Złotej Piłki 2024 – grafowa baza danych***

**Pierwszy projekt zaliczeniowy z Sieci semantycznych**

Prowadzący: dr inż. Wojciech Kozioł

Rzeszów 2025

Spis treści

[1. Opis bazy danych 3](#_Toc196582419)

[2. Źródła danych 3](#_Toc196582420)

[3. Struktura bazy danych 3](#_Toc196582421)

[4. Nietrywialne zapytania w języku Cypher 6](#_Toc196582422)

[5. Możliwości wykorzystania bazy danych 6](#_Toc196582423)

[6. Podsumowanie i możliwości rozwoju bazy danych 6](#_Toc196582424)

1. Opis bazy danych

Baza danych zawiera szczegółowe informacje o piłkarzach nominowanych do Złotej Piłki 2024, jednej z najbardziej prestiżowych nagród indywidualnych w światowym futbolu. Dane obejmują statystyki meczowe, przebieg kariery klubowej, osiągnięcia indywidualne oraz rozszerzone informacje biograficzne zawodników.

**Zakres przechowywanych danych:**

* Statystyki z sezonu 2023/2024, obejmujące m.in. gole, asysty, expected goals (xG), expected assists (xAG) oraz inne wskaźniki.
* Dane biograficzne, takie jak data urodzenia, wzrost, narodowość, aktualny klub, pozycja na boisku oraz preferowana noga.
* Informacje o przebiegu kariery: historia transferów, szczegóły kontraktowe, agenci reprezentujący zawodników oraz sponsorzy sprzętu sportowego.
* Osiągnięcia indywidualne, w tym zdobyte nagrody oraz wyniki głosowania w plebiscycie Złotej Piłki 2024 (punkty i miejsce w rankingu).
* Relacje pomiędzy zawodnikami a klubami, krajami, agentami, sponsorami oraz miejscami urodzenia, odzwierciedlające sieciowe powiązania w karierach piłkarzy.

1. Źródła danych

Baza danych została utworzona na podstawie danych pochodzących z trzech głównych źródeł, obejmujących zarówno statystyki meczowe, informacje biograficzne i transferowe zawodników, jak i oficjalne wyniki głosowania w plebiscycie Złotej Piłki 2024.

* **Kaggle:** <https://www.kaggle.com/datasets/willianoliveiragibin/ballon-dor-2024>

Zbiór zawierający szczegółowe statystyki meczowe nominowanych zawodników za sezon 2023/2024 w największych europejskich ligach (m.in. gole, asysty, xG, xAG, progresje piłki).

* **Transfermarkt:** <https://www.transfermarkt.pl/>

Dane pobrane za pomocą dynamicznego web scrapingu, obejmujące historię transferową, dane kontraktowe, nagrody, a także profil zawodnika (m.in. wzrost, pozycja, preferowana noga, agenci, outfitterzy).

* **Wikipedia:** <https://en.wikipedia.org/wiki/2024_Ballon_d%27Or>

Oficjalne dane dotyczące liczby punktów i miejsc zajętych przez zawodników w  głosowaniu Ballon d'Or 2024.

1. Struktura bazy danych

Baza danych zawiera informacje, które zostały odwzorowane w formie grafu, w którym poszczególne elementy reprezentowane są jako węzły, a relacje między nimi jako krawędzie. Baza umożliwia analizę karier zawodników, ich osiągnięć oraz transferów w ujęciu sieciowym.

* 1. **Węzły**

W bazie danych wyróżniono następujące typy węzłów:

* **Player** – zawodnik. Jego właściwości to np. imię i nazwisko, data urodzenia, wzrost, aktualny klub, punkty zdobyte w głosowaniu, miejsce w rankingu Ballon d'Or, statystyki sezonowe (gole, asysty, xG, xAG itp.).
* **Club** – klub piłkarski
* **Country** – kraj, reprezentujący obywatelstwo zawodnika
* **PlaceOfBirth** – miejsce urodzenia zawodnika
* **Position** – pozycja, na której zawodnik występuje na boisku
* **Foot** – dominująca noga zawodnika (lewa lub prawa)
* **Outfitter** – sponsor sprzętu sportowego zawodnika
* **Agent** – agent piłkarski reprezentujący zawodnika
* **Award** – nagroda. Jego właściwości to tytuł nagrody, rok zdobycia, klub (jeżeli dotyczy)
  1. **Relacje**

Węzły są połączone za pomocą następujących typów relacji:

* **(:Player)-[:BORN\_IN]->(:PlaceOfBirth)** – miejsce urodzenia zawodnika
* **(:Player)-[:HAS\_CITIZENSHIP]->(:Country)** – obywatelstwo zawodnika
* **(:Player)-[:PLAYS\_AS]->(:Position)** – pozycja na boisku
* **(:Player)-[:PLAYS\_WITH]->(:Foot)** – preferowana noga
* **(:Player)-[:SPONSORED\_BY]->(:Outfitter)** – sponsor sprzętu
* **(:Player)-[:REPRESENTED\_BY]->(:Agent)** – agent reprezentujący zawodnika
* **(:Player)-[:WON]->(:Award)** – zdobyte nagrody
* **(:Player)-[:TRANSFERRED]->(:Club)** – historia transferów
* **(:Player)-[:STARTED\_CAREER\_AT]->(:Club)** – pierwszy klub w karierze zawodnika

Relacje często zawierają dodatkowe właściwości, takie jak np. sezon transferu, data transferu, wartość rynkowa lub kwota transferu (dla TRANSFERRED).

* 1. **Liczebność bazy danych**

Liczebność poszczególnych węzłów i relacji przedstawia się następująco:

**Węzły:**

* Award: 580
* Club: 162
* Player: 30
* PlaceOfBirth: 27
* Agent: 23
* Country: 18
* Position: 11
* Outfitter: 6
* Foot: 2

**Całkowita liczba węzłów:** 859

**Relacje:**

* WON: 824
* TRANSFERRED: 221
* HAS\_CITIZENSHIP: 43
* BORN\_IN: 30
* PLAYS\_AS: 30
* PLAYS\_WITH: 30
* STARTED\_CAREER\_AT: 30
* SPONSORED\_BY: 29
* REPRESENTED\_BY: 27

**Całkowita liczba relacji:** 1264

* 1. **Ogólny wygląd bazy danych**

Na rysunku nr 1 przedstawiono pełną strukturę bazy danych w formie grafu. Układ grafu przypomina rozległą sieć skupisk (klastrów), gdzie zawodnicy znajdują się w centrum swoich powiązań, a wokół nich rozmieszczone są powiązane podmioty.

Obraz zawierający fajerwerki, światło

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

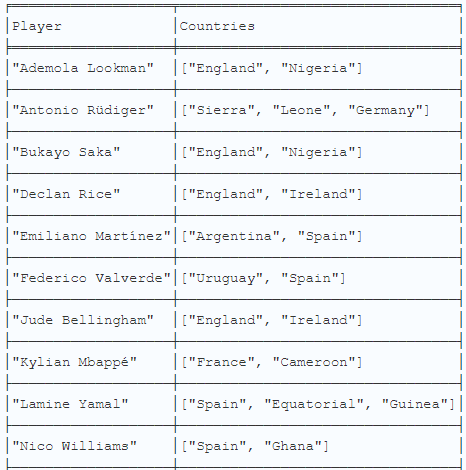
Rysunek Wygląd bazy danych

1. Nietrywialne zapytania w języku Cypher

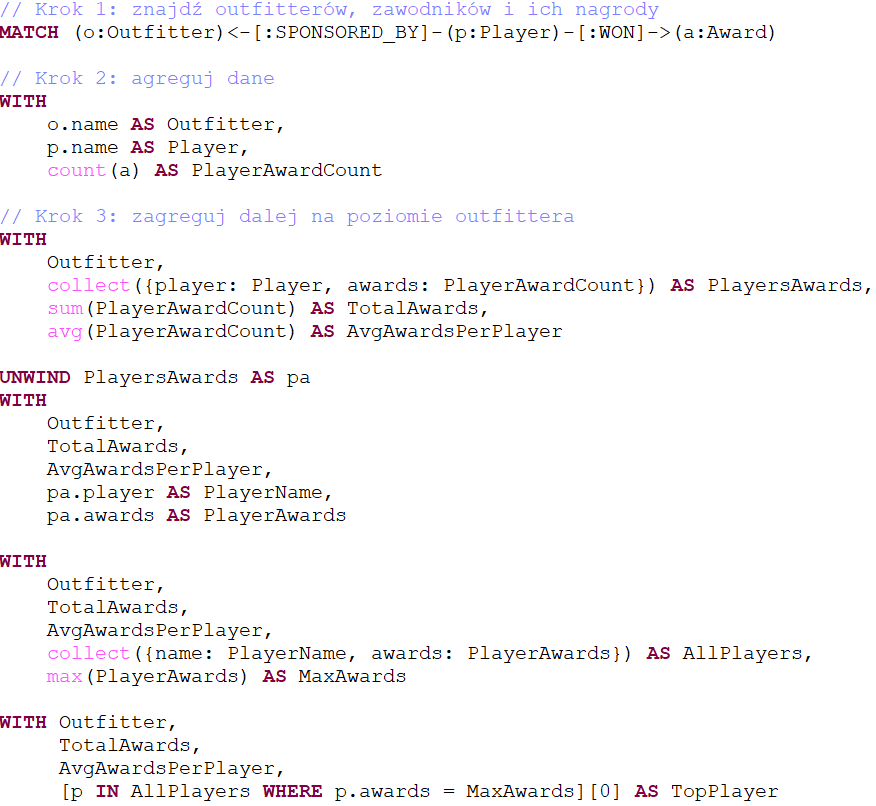
W ramach projektu przygotowano zestaw nietrywialnych zapytań w języku Cypher, umożliwiających analizę zgromadzonych danych. Zapytania te pozwalają m.in. na ocenę efektywności zawodników, analizę ścieżek kariery, wpływu agentów na sukcesy piłkarzy, a także porównywanie osiągnięć klubów oraz sponsorów sprzętu sportowego.

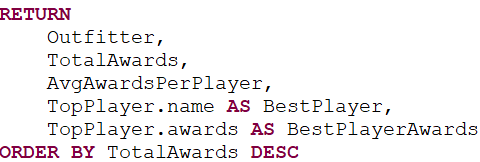
* **Wyszukaj zawodników, którzy mają więcej niż jedno obywatelstwo i przynajmniej jedno z krajów to kraj spoza Europy.**

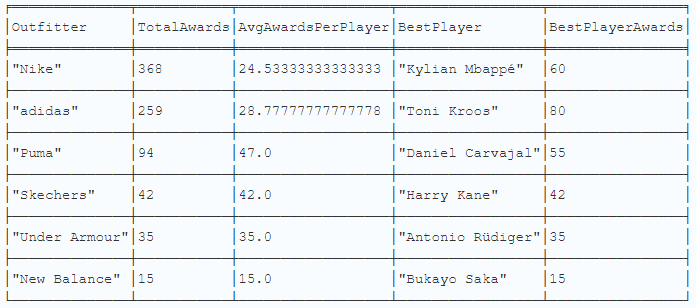




* **Wypisz sponsorów, których zawodnicy zdobyli największą liczbę nagród w karierze.**







* **Wypisz kraje, których zawodnicy charakteryzują się najwyższym średnim wzrostem, podając przy tym liczbę zawodników, minimalny i maksymalny wzrost oraz rozrzut wzrostu w danym kraju.**

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, numer, Równolegle, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, numer, Równolegle, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

* **Wypisz zawodników, którzy rozegrali najwięcej minut, ale mieli minimalny lub żaden udział w golach i asystach. Zapytanie pozwala wykryć mało efektywnych ofensywnie graczy lub zidentyfikować zawodników o defensywnym profilu.**

**Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

* **Znajdź zawodników z największą liczbą kluczowych podań (key passes) per 90 minut, co pozwala określić, kto jest najlepszym kreatorem gry.**

**Obraz zawierający tekst, Czcionka, zrzut ekranu, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, numer, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

* **Wypisz zawodników o najbardziej wszechstronnym profilu: ci, którzy łączą skuteczność ofensywną, kreatywność, progresję piłki, grę w defensywie oraz dryblingi. Zapytanie pokazuje TOP kompletne profile piłkarskie.**

**Obraz zawierający tekst, zrzut ekranu, Czcionka, dokument

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, numer, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

* **Wypisz top 5 zawodników, którzy najczęściej wykonują progresywne podania oraz utrzymują wysoką skuteczność podań. Pokazuje kluczowych graczy odpowiedzialnych za kontrolę tempa gry.**

**Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

* **Wypisz zawodników, którzy zdobyli najwięcej nagród grając w najmniejszej liczbie klubów w karierze.**

**Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, zrzut ekranu, numer, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

* **Wypisz zawodników, których liczba zdobytych bramek i asyst (G+A) jest niższa niż przewidywana suma Expected Goals i Expected Assists (xG + xA). Zapytanie pokazuje także ile gracz "powinien mieć" (xG i xA osobno) oraz ile faktycznie ma (gole i asysty).**

**Obraz zawierający tekst, Czcionka, zrzut ekranu, algebra

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, numer, zrzut ekranu, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

* **Wypisz zawodników, którzy zdobyli nagrody w sezonie 23/24 i posortuj ich według efektywności — czyli liczby minut przypadających na jedną nagrodę. Zapytanie pokazuje także listę zdobytych nagród oraz łączną liczbę minut zawodnika. Im niższa wartość minut na nagrodę, tym bardziej zawodnik był doceniany mimo mniejszej liczby minut na boisku.**

**Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający tekst, zrzut ekranu, linia, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

1. Możliwości wykorzystania bazy danych

Baza danych umożliwia analizę i porównywanie statystyk zawodników nominowanych do  Złotej Piłki 2024. Pozwala badać wpływ transferów oraz zmian klubów na osiągnięcia zawodników w sezonie. Umożliwia śledzenie przebiegu kariery zawodników w ujęciu czasowym.  
Na podstawie zgromadzonych danych możliwe jest przewidywanie przyszłych wyników i  rankingów. Baza daje możliwość analizy powiązań pomiędzy agentami a sukcesami reprezentowanych zawodników. Pozwala na wizualizację ścieżek kariery zawodników w postaci grafu. Może być również podstawą do budowy aplikacji rekomendacyjnych, np. wskazujących zawodników najlepiej pasujących do określonego klubu.

1. Podsumowanie i możliwości rozwoju bazy danych

Projekt umożliwia szczegółową analizę danych o czołowych piłkarzach sezonu 2023/2024, obejmującą zarówno ich osiągnięcia, jak i przebieg kariery. Dzięki integracji danych statystycznych, biograficznych i transferowych w modelu grafowym możliwe jest przeprowadzanie analiz oraz zapytań.

W ramach dalszego rozwoju bazy danych możliwe jest:

* Automatyczne aktualizowanie danych o zawodnikach wraz z pojawianiem się nowych sezonów.
* Rozszerzenie bazy o dane dotyczące drużyn narodowych oraz występów w międzynarodowych turniejach, takich jak Euro 2024 czy Copa América.
* Dodanie informacji o historii kontuzji zawodników.
* Integracja z zewnętrznymi źródłami danych, takimi jak API FIFA czy UEFA.