

Baza danych Curling Masters - dokumentacja

Bogdan Banasiak, Paulina Muszyńska, Karolina Rakus, Joanna Wantoła

26 czerwca 2022

1 Dokumentacja

1.1 Spis użytych technologii

- Visual Studio Code
- Python 3.9
- Jupyter
- RStudio
- R
- sql
- diagram.io
- Excel

1.2 Lista plików

- main.ipynb → wygenerowanie danych do tabel 'address', 'matches', 'finance', 'equipment' i zapis do .csv, wczytanie danych z plików .csv, logowanie do bazy, umieszczenie danych na serwerze
- db_schema.sql → wygenerowanie schematu bazy danych
- curling_masters.vuerd.json - program ERD DIAGRAM
- folder 'dane':
 - Address.csv → dane do tabeli address
 - Employees.csv → dane do tabeli employees
 - Equipment.csv → dane do tabeli equipment
 - Matches.csv → dane do tabeli matches
 - Person.csv → dane do tabeli person

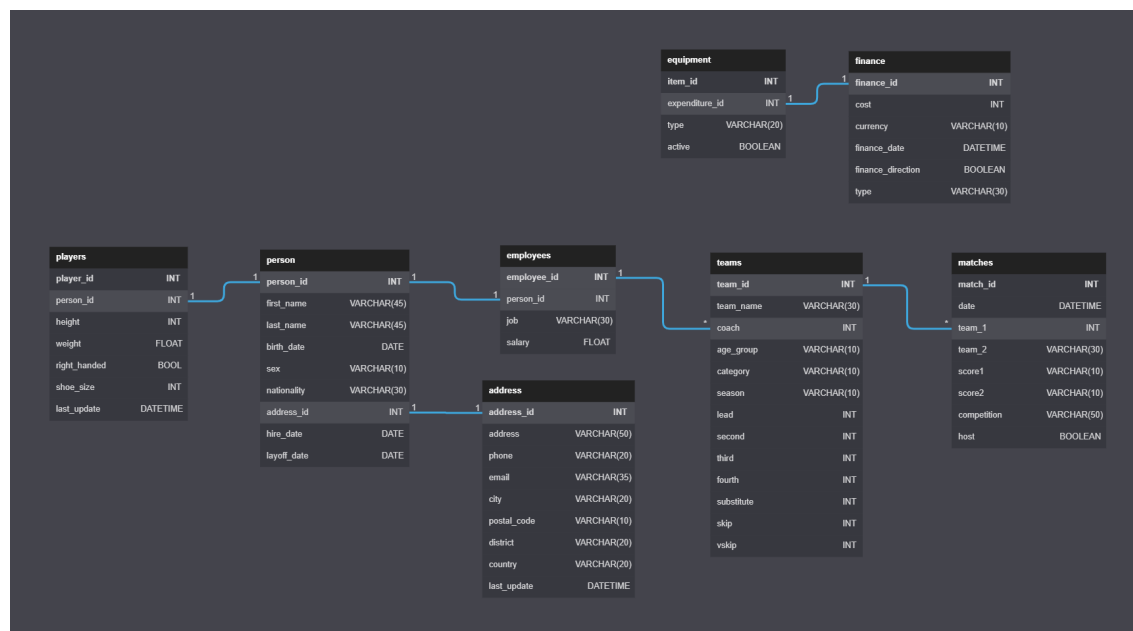
- Players.csv → dane do tabeli player
- Teams.csv → dane do tabeli teams
- kody_ulice_krakow.csv → dane z adresami, plik pomocniczy do wygenrowania tablicy address
- matches_1.csv → plik pomocniczy do wygenrowania tablicy matches
- matches_2.csv → plik pomocniczy do wygenrowania tablicy matches
- Teams.tomatches.xlsx → plik pomocniczy do wygenrowania tablicy matches
- men_last_name.csv → dane z nazwiskami, plik pomocniczy do wygenrowania tablicy matches
- women_last_name.csv → dane z nazwiskami, plik pomocniczy do wygenrowania tablicy matches
- folder 'dokumentacja':
 - dokumentacja.tex
 - dokumentacja.pdf
 - folder select_table → zawiera pliki png przedstawiające przykładowe wywołania dla każdej z tabeli
- folder 'analiza'
 - 'raport.rnw' → plik do wygenrowania analizy
 - 'raport.pdf' → plik .pdf z analizą

1.3 Kolejność i sposób uruchomienia

- uruchomienie pliku 'db_chema.sql' → schemat bazy danych
- uruchomienie pliku 'main.ipynb' (tylko sekcje 'Wczytnie plików csv - dane do tabel' oraz 'Logowanie do bazy i umieszczenie danych na serwerze')
- 'raport.rnw' → analiza

1.4 Schemat projektu bazy danych

- schemat bazy danych (przy pomocy programu diagram.io)

















- wygenerowanie potrzebnych tabel przy użyciu 'create_diagram.sql'

1.5 Przykładowe wyniki dla poszczególnych tabel

Przedstawiamy przykładowe wyniki dla poszczególnych tabel, przy użyciu za-
pytania

```
SELECT * FROM 'nazwa_tabeli' LIMIT 5;
```

- address

| |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| | | address_id int | address varchar(50) | phone varchar(20) | email varchar(35) | city varchar(20) | postal_code varchar(10) | district varchar(20) | country varchar(20) |
|  | 1 | 1 | ul. Kamienna 154/2 | 568620700 | olivia.chrzanowska@op | Kraków | 30-001 | Małopolska | Poland |
|  | 2 | 2 | ul. Prądnicka 57/3 | 511479654 | adam.gujska@icloud.co | Kraków | 30-002 | Małopolska | Poland |
|  | 3 | 3 | ul. Składowa 81/1 | 742835639 | iwona.kaczmarek@iclou | Kraków | 30-002 | Małopolska | Poland |
|  | 4 | 4 | ul. Zbożowa 164 | 620398664 | rysard.kazimierzczak@g | Kraków | 30-002 | Małopolska | Poland |
|  | 5 | 5 | ul. Lubelska 105 | 697877854 | pawel.luks@gmail.com | Kraków | 30-003 | Małopolska | Poland |

- employees

| <input type="checkbox"/> | Q | * employee_id int | * person_id int | * job varchar(30) | salary float |
|--------------------------|---|----------------------|--------------------|----------------------|-----------------|
| <input type="checkbox"/> | 1 | 1 | 1 | SECRETARY | 4735 |
| <input type="checkbox"/> | 2 | 2 | 2 | MARKETING_DIRECTOR | 6611 |
| <input type="checkbox"/> | 3 | 3 | 3 | PRESS_OFFICER | 6329 |
| <input type="checkbox"/> | 4 | 4 | 4 | PRODUCT_MANAGER | 7302 |
| <input type="checkbox"/> | 5 | 5 | 5 | SPORTS_DIRECTOR | 7505 |

- equipment

| <input type="checkbox"/> | Q | * item_id int | * expenditure_id int | * type varchar(20) | * active tinyint |
|--------------------------|---|------------------|-------------------------|-----------------------|---------------------|
| <input type="checkbox"/> | 1 | 1 | 1 | GLOVES | 0 |
| <input type="checkbox"/> | 2 | 2 | 2 | GRIPPER | 0 |
| <input type="checkbox"/> | 3 | 3 | 3 | TAPE | 0 |
| <input type="checkbox"/> | 4 | 4 | 4 | GLOVES | 0 |
| <input type="checkbox"/> | 5 | 5 | 5 | STOPWATCH | 0 |

- finance

| | Q | * finance_id int | * cost int | * currency varchar(10) | * finance_date datetime | * finance_direction tinyint | * type varchar(30) |
|--|---|---------------------|---------------|---------------------------|----------------------------|--------------------------------|-----------------------|
| | 1 | 1 | 36 | \$ | 2019-01-05 07:00:00 | 0 | EQUIPMENT |
| | 2 | 2 | 33 | \$ | 2019-01-05 07:00:00 | 0 | EQUIPMENT |
| | 3 | 3 | 29 | \$ | 2019-01-05 07:01:00 | 0 | EQUIPMENT |
| | 4 | 4 | 45 | \$ | 2019-01-05 07:05:00 | 0 | EQUIPMENT |
| | 5 | 5 | 50 | \$ | 2019-01-05 07:05:00 | 0 | EQUIPMENT |

- matches

| | Q | * match_id int | * date datetime | * team_1 int | team_2 varchar(30) | * score1 varchar(10) | * score2 varchar(10) | competition varchar(50) | * host tinyint |
|--|---|-------------------|---------------------|-----------------|-----------------------|-------------------------|-------------------------|----------------------------|-------------------|
| | 1 | 1 | 2019-08-12 00:00:00 | 1 | POC SZABŁOWSKI | 13.0 | 4.0 | FRIENDLY MATCH | 0 |
| | 2 | 2 | 2020-07-03 00:00:00 | 1 | AZS PRYBACKI | 3.0 | 22.0 | TRAINING MATCH | 1 |
| | 3 | 3 | 2020-12-04 00:00:00 | 1 | AZS SZARY | 5.0 | 29.0 | TRAINING MATCH | 1 |
| | 4 | 4 | 2020-01-19 00:00:00 | 1 | AZS SARNA | 0.0 | 3.0 | TRAINING MATCH | 1 |
| | 5 | 5 | 2020-01-02 00:00:00 | 1 | POC RUDNIK | 7.0 | 3.0 | NATIONAL LEAGUE | 0 |

- person

| | Q | * person_id int | * first_name varchar(45) | * last_name varchar(45) | * birth_date date | * sex varchar(10) | * nationality varchar(30) | * address_id int | * hire_date date | layoff_date date |
|--|---|--------------------|-----------------------------|----------------------------|----------------------|----------------------|------------------------------|---------------------|---------------------|---------------------|
| | 1 | 1 | OLIwia | CHRZANOWSKA | 1998-03-13 | W | PL | 1 | 2019-01-05 | (NULL) |
| | 2 | 2 | ADAM | GUJSKA | 1999-03-26 | W | PL | 2 | 2019-01-05 | 2020-10-24 |
| | 3 | 3 | IWONA | KACZMAREK | 1991-02-13 | W | PL | 3 | 2019-01-05 | 2019-10-18 |
| | 4 | 4 | RYSZARD | KAZIMIERCZAK | 1962-06-13 | M | PL | 4 | 2019-01-05 | (NULL) |
| | 5 | 5 | PAWEŁ | LUKS | 1996-11-04 | M | PL | 5 | 2019-01-05 | (NULL) |

- player

| | Q | * player_id int | * person_id int | * height int | * weight float | * right_handed tinyint | shoe_size int | last_update datetime |
|--|---|--------------------|--------------------|-----------------|-------------------|---------------------------|------------------|-------------------------|
| | 1 | 1 | 25 | 160 | 62.29 | 1 | 36 | 2021-02-08 00:00:00 |
| | 2 | 2 | 26 | 164 | 68.24 | 1 | 38 | 2021-02-08 00:00:00 |
| | 3 | 3 | 27 | 170 | 71.53 | 0 | 46 | 2020-03-08 00:00:00 |
| | 4 | 4 | 28 | 161 | 59.31 | 0 | 36 | 2021-02-08 00:00:00 |
| | 5 | 5 | 29 | 174 | 61.73 | 1 | 45 | 2021-02-08 00:00:00 |

- teams

| | * team_id int | * team_name varchar(30) | * coach int | * age_group varchar(10) | * category varchar(10) | season varchar(10) | * lead int | * second int | * third int | * fourth int | substitute int | * skip int | vskip int |
|--|------------------|----------------------------|----------------|----------------------------|---------------------------|-----------------------|---------------|-----------------|----------------|-----------------|-------------------|---------------|--------------|
| | 1 | KRK GRABOWSKI | 12 | SENIOR | M | 2019/2020 | 30 | 28 | 27 | 26 | 40 | 30 | 28 |
| | 2 | KRK MACKOWIAK | 12 | SENIOR | M | 2019/2020 | 21 | 3 | 42 | 33 | 20 | 21 | 3 |
| | 3 | KRK GUT | 12 | SENIOR | M | 2019/2020 | 41 | 18 | 15 | 22 | 29 | 18 | 41 |
| | 4 | KRK LENG | 13 | SENIOR | K | 2019/2020 | 10 | 31 | 25 | 34 | 4 | 10 | 31 |
| | 5 | KRK ZAK | 13 | SENIOR | K | 2019/2020 | 17 | 9 | 24 | 6 | 2 | 17 | 9 |

2 Lista zależności funkcyjnych

- employees
 - klucz główny: employee_id
 - klucze kandydujące: employee_id, person_id
 - zależności funkcyjne: employee_id → pozostałe pozdbiory atrybutów relacji, person_id → pozostałe pozdbiory atrybutów relacji
 - komentarz: Person_id jest kluczem obcym do tabeli person zawierającej dane osobowe pracowników. Zakładamy, że jedna osoba może być zatrudniona tylko na jednym stanowisku, zatem person_id jest kluczem kandydującym.
- players
 - klucz główny: player_id
 - klucze kandydujące: player_id, person_id
 - zależności funkcyjne: player_id → pozostałe pozdbiory atrybutów relacji, person_id → pozostałe pozdbiory atrybutów relacji
 - komentarz: Zakładamy, że dany gracz może występować w systemie tylko raz, zatem person_id jest kluczem kandydującym.

- address
 - klucz główny: adres_id
 - klucze kandydujące: adres_id
 - zależności funkcyjne: adres_id \rightarrow pozostałe pozdbiory atrybutów relacji
 - komentarz: Ponieważ system nie przewiduje weryfikacji adresu email oraz numeru telefonu, zakładamy, że dane podane przez użytkowników mogą być niewiarygodne, a więc zawierać duplikaty. Z tego powodu zarówno email, jak i numer telefonu nie są kluczami kandydującymi.
- finance
 - klucz główny: finance_id
 - klucze kandydujące: finance_id
 - zależności funkcyjne: finance_id \rightarrow pozostałe pozdbiory atrybutów relacji
 - komentarz: finance_id jest unikalnym kluczem głównym oraz nie zawiera wartości pustych.
- equipment
 - klucz główny: item_id
 - klucze kandydujące: item_id, expenditure_id
 - zależności funkcyjne: item_id \rightarrow pozostałe pozdbiory atrybutów relacji, expenditure_id \rightarrow pozostałe pozdbiory atrybutów relacji
 - komentarz: Tabela equipment zawiera przedmioty dopasowane do zakupu, więc typy przedmiotów występujących w tabeli equipment nie są unikalne. Z tego powodu type nie jest kluczem głównym, ponieważ zawiera duplikaty oraz expenditure_id jest kluczem kandydującym.
- matches
 - klucz główny: match_id
 - klucze kandydujące: match_id, (date, team_1)
 - zależności funkcyjne: match_id \rightarrow pozostałe pozdbiory atrybutów relacji, (date, team_1) \rightarrow pozostałe pozdbiory atrybutów relacji
 - komentarz: W jednym dniu dana drużyna może odbyć tylko jeden mecz, zatem para (date, team_1) jest kluczem kandydującym.
- person
 - klucz główny: person_id
 - klucze kandydujące: person_id, (first_name, last_name)

- zależności funkcyjne: `person_id` \rightarrow pozostałe pozdbiory atrybutów relacji, `(first_name, last_name)` \rightarrow pozostałe pozdbiory atrybutów relacji
- komentarz: W systemie imię i nazwisko jednoznacznie identyfikują osobę, więc kombinacja tych atrybutów stanowi klucz kandydujący.
- teams
 - klucz główny: `team_id`
 - klucze kandydujące: `team_id`
 - zależności funkcyjne: `team_id` \rightarrow pozostałe pozdbiory atrybutów relacji
 - komentarz: `team_id` jest unikalnym kluczem głównym oraz nie zawiera wartości pustych.

2.1 Uzasadnienie, że baza jest w EKNF

Jak wykazaliśmy w poprzednim podpunkcie, każda nietrywialna zależność funkcyjna albo zaczyna się od nadklucza albo kończy się atrybutem elementarnym. Oznacza to, że baza jest w EKNF.

2.2 Napotkane trudności

- Trudność z dostępem do danych na temat curlingu, np. problemy ze zdobyciem danych odnośnie średniego wzrostu osób nieletnich ich wag i rozmiarów butów
- Zaprojektowanie tego jak zmienia się zespół na przestrzeni lat
- Trudności związane z normalizacją
- Trudności związane z wyliczeniem realnych kosztów