

Inżynieria Danych - opis pierwszej fazy projektu **PawnToDB4**

Skład grupy projektowej: Michał Hoffmann, Adam Szwaja, Piotr Kubicki

Tematyka:

Projekt przedstawia bazę danych systemu szachowego, która zapewnia informacje o szachistach, partiach szachowych i turniejach, podczas których zawodnicy rozgrywają swoje partie. Szachiści przynależą do grup i posiadają swój ranking ELO, który uaktualniany jest po każdej rozegranej przez nich partii. Każda partia przechowywana jest jako parowanie dwóch graczy mające miejsce w danym turnieju i określonym czasie. Dodatkowo dostępny jest jej techniczny opis wliczając w to formalny zapis posunięć, nazwę otwarcia, czy sposób zakończenia. Partia rozgrywana w turnieju umieszczona jest na pewnym jego szczęblu świadczącego o jej znaczeniu. Każdy turniej posiada swoją nazwę i odbywa się w pewnych ramach czasowych i w miejscu o określonym adresie. Sprecyzowany jest również typ turnieju świadczący o sposobie rozgrywania partii i liczbie zawodników biorących w nim udział

Założone cele:

Głównym założeniem bazy jest gromadzenie informacji o partiach i turniejach, które modelują rzeczywisty postęp rozgrywek, bieżące zmiany w rankingu graczy i przebieg kolejnych turniejów zapewniając szeroki obraz tego pięknego i emocjonującego sportu. Najważniejsze cele to:

- Przechowywanie szczegółowych informacji o partiach, pozwalających prześledzić ich przebieg, określić rangę, umieścić w odpowiednich kategoriach.
- Możliwość wykorzystania bazy do prowadzenia regularnych zapisów rozgrywek w czasie rzeczywistym, tak aby możliwe było obliczanie aktualnych rankingów ELO dla graczy i wyświetlanie rankingu zbiorczego.
- Przeglądanie informacji w kontekście turniejów o różnych typach (śledzenie przebiegu wybranego turnieju, zestawianie ze sobą historycznych danych z poprzednich edycji, prowadzenie statystyk dotyczących popularności otwarć czy sukcesów danej grupy szachistów)
- Dostęp do wszystkich partii gracza, pozwalający na śledzenie jego rozwoju, określenie specyfikacji gry, czy największych sukcesów

Problemy i rozwiązania:

- Problem ze sposobem przechowywania rankingu ELO.
Najbardziej efektywne jest uaktualnianie rankingu po każdej partii. Najbardziej interesuje nas bieżący ranking, który jest stale potrzebny przy obliczaniu jego własnych zmian. To rozwiązanie wymaga wyzwalacza dla nowo dodanej partii, który uruchomi funkcję obliczającą i zmieniającą rankingi obu graczy. Dodatkowo, przechowywany będzie maksymalny ranking gracza jako ważna informacja. Kolejny wyzwalacz będzie sprawdzał, czy nowy ranking jest rekordowy
- Problem z liczeniem rankingu ELO opierającego się na ilości punktów obu graczy w momencie rozgrywania partii. Koniecznością jest dodawanie partii chronologicznie (w przeciwnym wypadku dodanie partii zaburzyłoby rankingi graczy, które wpłynęły na wszystkie późniejsze zmiany rankingów).
Rozwiązanie symuluje bieżące aktualizowanie stanu rozgrywek i wymaga wyzwalacza sprawdzającego czy nowo dodana partia jest najmłodsza.
- Problem z synchronizacją turniejów z dodawanymi do nich partiami.
Rozwiązaniem jest ustalenie konwencji tworzenia turnieju w pierwszej kolejności, a następnie dodawania do niego kolejnych partii. To również wymaga wprowadzenia pewnych ograniczeń.
- Redundancja wyniku partii w tabeli parowań i tabeli zapisu gry.
Tabela parowań odwołuje się do graczy i turnieju w którym się odbyła i wymaga kluczowej informacji o wyniku partii. Z drugiej strony, tabela zapisu gry ma za zadanie zawierać wszystkie dane techniczne przebiegu partii, nie uwzględniając zawodników i turniejów, więc również powinna mieścić w sobie wynik partii.
- Przechowywanie danych historycznych: kłopot z obliczaniem rankingu gracza dla danego momentu w przeszłości. Potencjalnym rozwiązaniem byłoby dodanie rankingów obu graczy do krotki w tabeli pairings. Wtedy będzie się dało wyliczyć bilanse rankingu dla ciągu partii w przeszłości gracza i dodać je do aktualnego ELO.

Opis struktury:

Baza składa się z 11 tabel. Najważniejsze z nich opisują parowania/partie (pairings), graczy (players) oraz turnieje (tournaments) i są wzajemnie powiązane z użyciem tabeli pośrednich (partie są rozgrywane przez graczy i mogą się odbywać podczas turniejów, gracze posiadają swoje partie i biorą udział w turniejach, do każdego turnieju przynależy zbiór graczy i partii). Pozostałe tabele wprowadzają szczegółowe informacje o grupach szachistów, zapisie gry, miejscach i typach turniejów.

Krótkie opisy wszystkich tabel:

- **pairings** - tabela parowań, która przechowuje informacje o danej partii pod kątem relacji z tabelami players i tournaments. Zawiera id partii (klucz), id obu graczy (białych i czarnych bierok), id turnieju, datę, wynik partii oraz id zapisu partii (odnoszące się do tabeli game_record).
- **game_record** - prezentuje techniczne szczegóły partii. Jej kluczem jest id zapisu partii. Zawiera id zapisu posunięć, id otwarcia, rodzaj zakończenia i wynik partii (redundancja, aby mieć dostęp do informacji o wyniku z poziomu tej tabeli, nieuwzględniającej graczy i turniejów).
- **move_record** - zawiera klucz - id zapisu posunięć i zapis posunięć w partii w notacji szachowej.
- **openings** - zawiera klucz - id otwarcia, nazwę otwarcia oraz oznaczenia pierwszych dwóch posunięć
- **pairing_tournament** - tabela pośrednia. Zawiera id parowania, id turnieju oraz dodatkowo informacje o szczeblu w turnieju na którym została rozegrana partia.
- **tournaments** - przedstawia najważniejsze informacje dotyczące turnieju i mieści w sobie klucz - id turnieju, nazwę turnieju, id typu turnieju, id miejsca oraz daty rozpoczęcia i zakończenia turnieju.
- **types** - tabela typów turniejów. Zawiera klucz - id typu, nazwę typu, oraz ilość graczy.
- **places** - tabela miejsc organizacji turniejów. Zawiera klucz id miejsca, nazwę miasta, nazwę i numer ulicy.
- **player_tournament** - tabela pośrednia łącząca graczy z turniejami w których brali udział. Zawiera id gracza i id turnieju.
- **players** - tabela graczy, zawiera klucz - id gracza, imię, nazwisko, id grupy, do której należy, aktualny i rekordowy ranking elo.
- **groups** - zawiera klucz - id grupy oraz nazwę grupy.

Ograniczenia i struktury potrzebne w dalszej części projektu:

- Dodawanie partii chronologicznie potrzebuje ograniczeń w postaci wyzwalacza, sprawdzającego, czy nowo dodana partia rozegrała się nie wcześniej niż pozostałe.
- Kolejne wyzwalacze będą potrzebne do uruchomienia funkcji obliczającej ranking ELO po rozegranej partii oraz aktualizacji wyniku maksymalnego jeśli nowy ranking jest dla gracza rekordowy.
- Również tworzenie nowego turnieju i dodawanie do niego partii wymaga zestawu funkcji i wyzwalaczy.
- Indeksy będą zastosowane dla największych i najważniejszych tabel: graczy i parowań.