

Zimny report

1. Úvod a ciele semestra Hlavným cieľom tohto semestra bolo polozenie pevných technologických základov pre vývoj inteligentného systému na odporúčanie videohier (Game Recommender System). Kľúčovou úlohou nebolo len "napísanie kódu", ale predovšetkým návrh udržateľnej a škálovateľnej architektúry, ktorá umožní v ďalších fázach efektívne trénovanie modelov strojového učenia (LLM).

Ciele stanovené na tento semester zahŕňali:

- Inicializáciu vývojového prostredia a nastavenie verziovania kódu.
- Výber a technickú analýzu zdroja dát (RAWG API).
- Návrh a implementáciu databázovej schémy.
- Vývoj automatizovaného ETL procesu (Extract, Transform, Load) pre zber a čistenie dát.

2. Hĺbková analýza projektu Pred začiatkom implementácie bola vykonaná **dôkladná a hrubá analýza** požiadaviek projektu s cieľom zabezpečiť vysokú kvalitu realizácie.

- **Analýza dát:** Rozhodli sme sa pre stratégiu "Detail Fetching" (dvojfázové sťahovanie), aby sme maximalizovali kvalitu dát pre budúce LLM, pričom sme optimalizovali využitie API limitov (Rate Limiting).
- **Architektúra databázy:** Po zvážení možností normalizácie vs. denormalizácie sme pre analytické potreby zvolili hybridný prístup v SQLite, kde sú komplexné metadáta (žánre, platformy) spracované do formátu vhodného pre Data Science (textová reprezentácia a JSON serializácia).

3. Dosiahnuté výsledky a realizácia V priebehu semestra boli úspešne realizované nasledujúce kľúčové komponenty:

- **Vývojové prostredie a CI/CD základy:**
 - Konfigurácia izolovaného prostredia Python (`venv`) pre zamedzenie konfliktov závislostí.
 - Inicializácia Git repozitára so správne nastaveným `.gitignore` pre vylúčenie systémových súborov a citlivých údajov (API kľúče).
- **Databázový backend:**
 - Návrh schémy pre SQLite databázu, ktorá obsahuje 19 atribútov vrátane detailných popisov, hodnotení a hardvérových požiadaviek.
 - Implementácia DDL skriptov pre automatické vytvorenie štruktúry databázy.
- **ETL Pipeline (Extrakcia a Transformácia):**
 - Implementácia robustného skriptu v jazyku Python, ktorý komunikuje s RAWG API.
 - Vytvorenie pokročilej logiky na čistenie dát (`Regex`) pre odstránenie HTML značiek a formátovacích znakov z popisov hier.
 - Implementácia mechanizmu `executemany` pre optimalizovaný hromadný zápis dát do databázy.

- Ošetrenie chybových stavov (Error Handling) pre HTTP requesty a integritu databázy.

4. Použitá literatúra a zdroje Pri návrhu a implementácii sme vychádzali z nasledujúcej technickej dokumentácie:

1. *RAWG Video Games Database API Documentation* – Špecifikácia endpointov a dátových modelov.
2. *Python 3 Documentation (sqlite3, json, re)* – Oficiálna dokumentácia pre prácu s databázami a regulárnymi výrazmi.
3. *Pandas Documentation* – Preštudovanie možností manipulácie s dátami pre budúce analytické úlohy.
4. *Git SCM Documentation* – Best practices pre správu verzií.