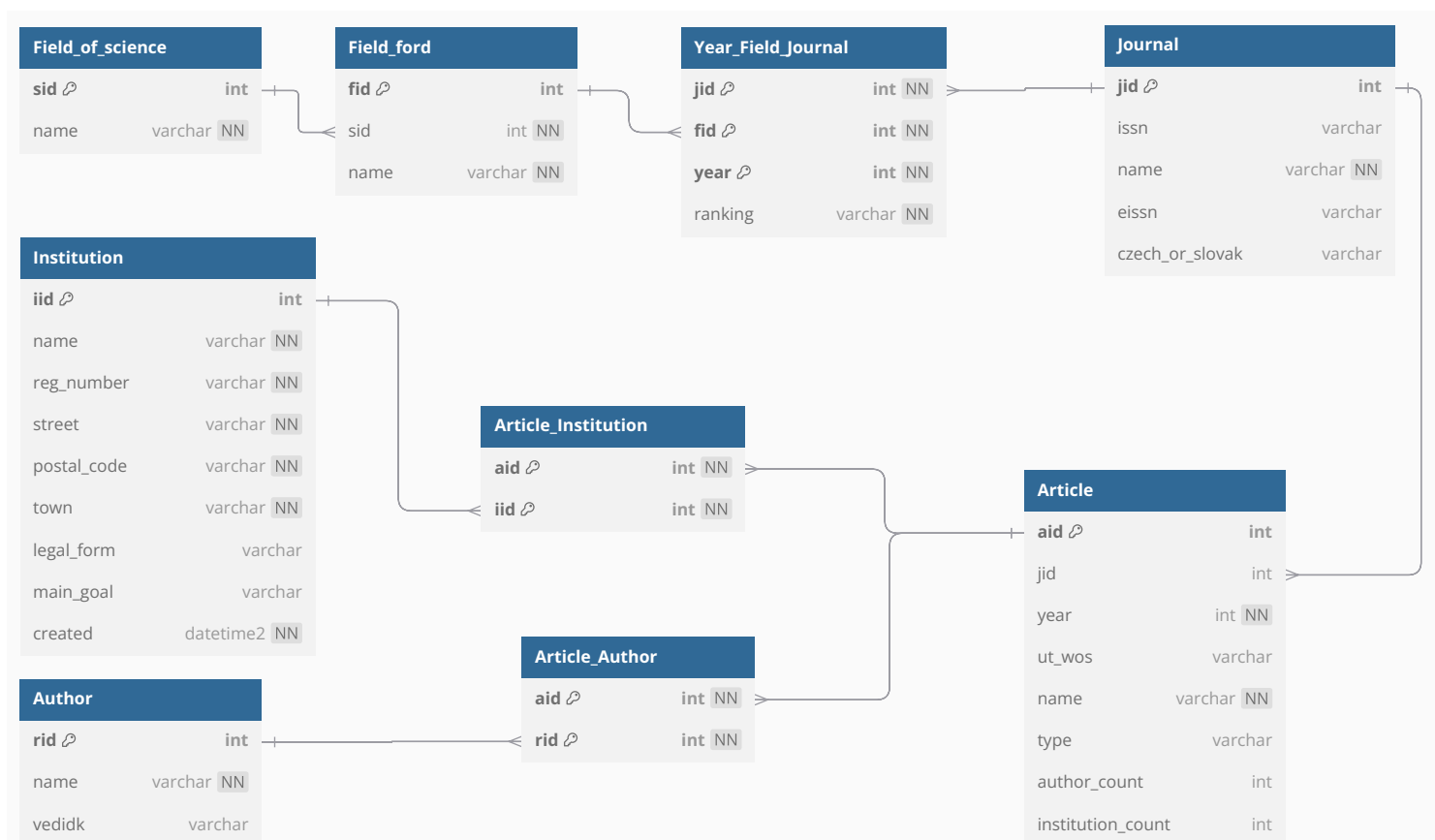


Popis relačního datového modelu databáze pro první test (SQL SELECT) 2024/25

Radim Bača

10. října 2024



1 Popis

Databáze je vytvořena na základě veřejně dostupných dat¹. Dále budeme tuto databázi nazývat RVVI. Zde je kompletní seznam tabulek, které jsme v RVVI vytvořili spolu s krátkým popisem toho jaký je význam jednoho řádku (t.j. jednoho záznamu) v této tabulce:

- **Author** - Osoba s příslušným jménem.
- **Article** - Jeden článek publikovaný v určitém časopise. Toto je ústředním tématem celé databáze.
- **Article_Author** - Vazební tabulka, kde jeden záznam vyjadřuje skutečnost, že autor s daným `rid` je spoluautorem článku s `aid`. Takovéto vazební tabulky jsou v databázích běžné a vyjadřují vlastně vztah M:N mezi entitami v datech. V našem případě vyjadřují vztah M:N mezi autorem a článkem. Takže článek může mít mnoho autorů a autor může mít mnoho článků.
- **Institution** - Jedna instituce spolu s navazujícími atributy. Zde se vyskytují pouze české instituce. Na článku mohou spolupracovat osoby ze zahraničí, nicméně jejich instituce se zde neobjevuje.
- **Article_Institution** - Opět vazební tabulka propojující článek s institucí.
- **Journal** - Časopis ve kterém daný článek vyšel. Článek vždy náleží právě jednomu časopisu.
- **Field_of_science** - Jedná se vědní oblast dle odborných panelů uvedených na stránkách RVVI. Jedná se o číselník s šesti záznamy.
- **Field_ford** - Jedná se o další rozřazení vědních oborů dle metodiky FORD. Opět se jedná o číselník s relativně malým množstvím záznamů.
- **Year_field_journal** - Vazební tabulka, kde jeden záznam vyjadřuje skutečnost, že časopis s daným `jid` byl v daném roce přiřazen do určitého oboru FORD v určitém ratingu. Hodnota rating pak může nabývat pouze hodnot Decil, Q1, Q2, Q3, Q4. Časopis může být zařazen v jednom roce do více oborů.

Poznámky:

- Data nejsou pročištěna a některé záznamy mohou v databázi chybět. Nicméně v podstatě se podařilo naimportovat 99% článků a souvisejících dat.
- Ve vstupních datech není žádný identifikátor osoby, tzn. že v databázi nejsou žádné osoby se stejným jménem. Je jasné, že to neodpovídá realitě.
- Pokud se v zadání píše článek v prvním decilu, pak se myslí článek, jehož rating v tabulce `Year_field_journal` je 'Decil'.
- **Pozor!** Tabulku `Year_field_journal` a `Article` je potřeba spojovat na základě dvou atributů (`jid`, `year`).
- **Pozor!** Jednomu záznamu v `Article` může odpovídat více záznamů v tabulce `Year_field_journal`. Časopis totiž může být v daném roce ve více oborech a s různým rating!

¹<https://hodnoceni.rvvi.cz/hodnoceni2022/biblio-obory>