

## Posudek oponenta bakalářské práce

**Student:** Šorm Jan  
**Téma:** Spojování záznamů v genealogických datech (id 22057)  
**Oponent:** Rozman Jaroslav, Ing., Ph.D., UITS FIT VUT

1. **Náročnost zadání** **obtížnější zadání**  
Zadání je spíše obtížnější, neboť na rozdíl od jiných typů zadání tady není zaručeno, že výsledek bude fungovat a také to není příliš rozšířená oblast zkoumání.
2. **Splnění požadavků zadání** **zadání splněno**  
Zadání práce bylo splněno a nemám k němu výhrady.
3. **Rozsah technické zprávy** **je v obvyklém rozmezí**  
Zadání práce je v obvyklém rozmezí. Postrádám tam ale lepší popis druhé části zadání, tzn. spojování záznamů, např. v teoretické části jsou kapitoly o spojování záznamů a o pravděpodobnosti, v implementaci ale toto není zmíněno.
4. **Prezentační úroveň předložené práce** **75 b. (C)**  
Prezentační část práce je na dobré úrovni a práce je pochopitelná pro čtenáře.
5. **Formální úprava technické zprávy** **75 b. (C)**  
Po typografické stránce je práce na dobré úrovni a po jazykové je na obvyklé úrovni našich studentů.
6. **Práce s literaturou** **55 b. (E)**  
Literatura obsahuje pouze 8 položek, což se domnívám, že je pro práci tohoto typu celkem málo.
7. **Realizační výstup** **65 b. (D)**  
Kód práce je ve dvou souborech \*.cpp, které obsahují pouze main, zbylé funkce jsou v hlavičkových souborech, což asi není úplně standardní způsob psaní C++ programů.
8. **Využitelnost výsledků**  
Předpokládáme, že výsledky práce použijeme v projektu DEMoS.
9. **Otázky k obhajobě**
  - Podle čeho určujete tzv. ground truth a jestli je to součástí vstupního souboru, jak tam poznáte, které záznamy mají být spojeny se kterými záznamy?
  - Mohl byste úspěšnost vašeho spojování záznamů vyjádřit pomocí true positive, true negative, false positive a false negative, případně pomocí F-measure?
10. **Souhrnné hodnocení** **75 b. dobře (C)**  
V textu práce chybí lepší vysvětlení toho, jak dochází ke spojování záznamů, např. jak je tam použita pravděpodobnost, nebo v úvodu uvedený Fellegi-Sunter model. Dále kód práce není příliš dobře strukturovaný a výsledky testování by bylo lepší uvádět pomocí true/false positive/negative. Z těchto důvodů práci hodnotím stupněm C, i když měla potenciál na mnohem lepší výsledek.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 2. června 2019

Rozman Jaroslav, Ing., Ph.D.  
oponent