

Odporúčanie pomocou grafových štruktúr

Autor: Ondrej Antl

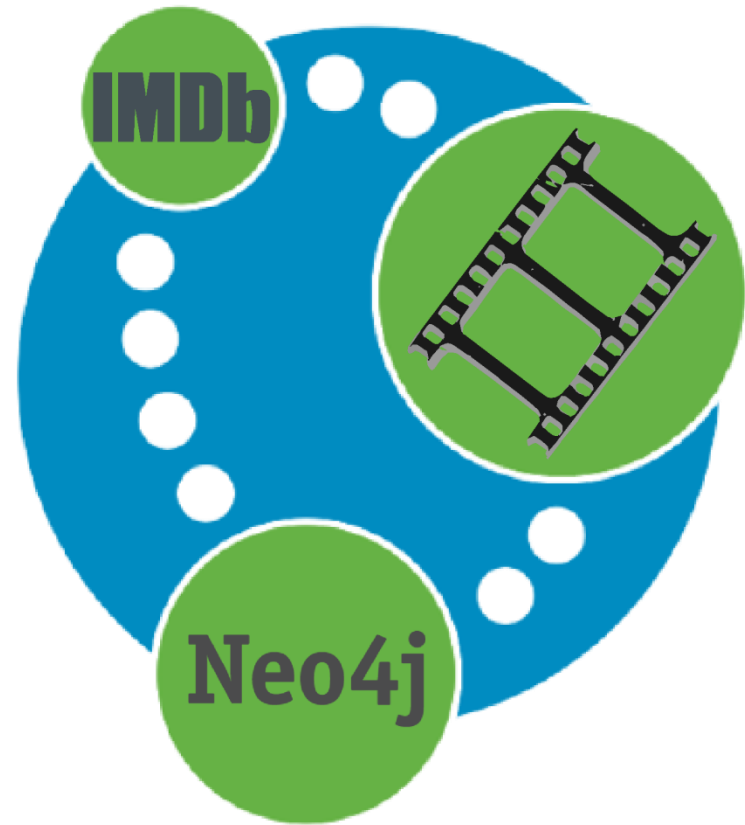
Vedúci práce: Ing. Dušan Zeleník, PhD.

Obsah

- Cieľ práce
- Aktuálne prístupy
- Problémy odporúčania
- Riešenie a aplikácia
- Testovanie metódy
- Možné vylepšenia
- Zhodnotenie

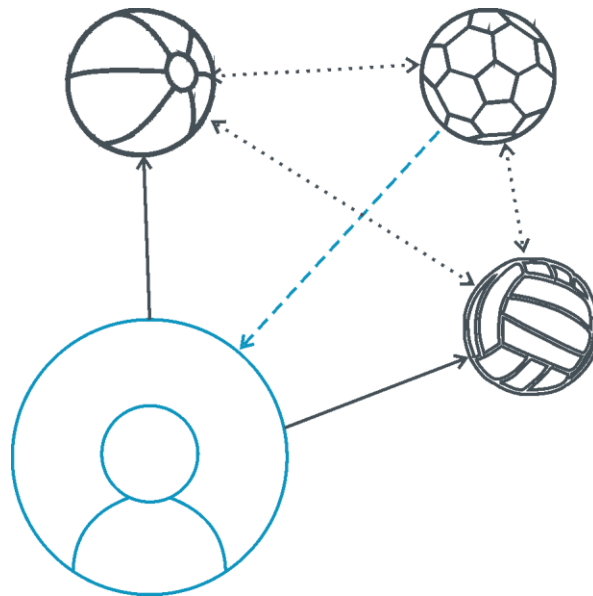
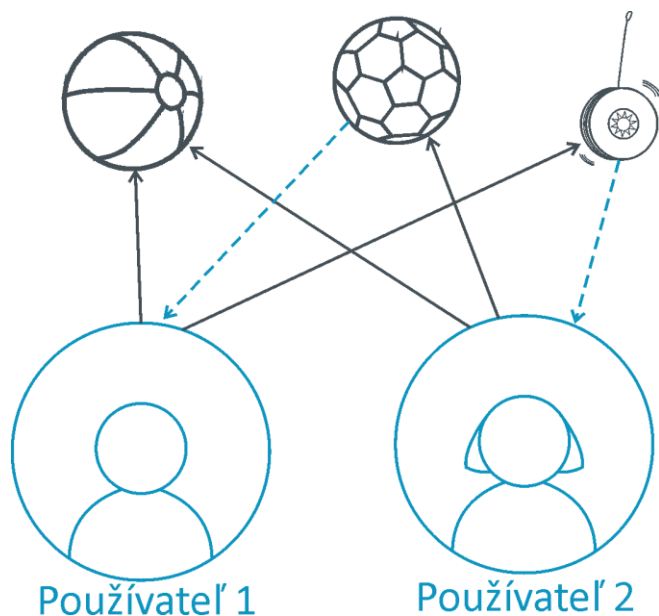
Cieľ práce

- Metóda odporúčania
- Grafová databáza - Neo4j
- Aplikácia v doméne - filmy



Aktuálne prístupy

- Rôzne druhy odporúčania - kolaboratívne, na základe obsahu, hybridné, a i.



Problémy odporúčania

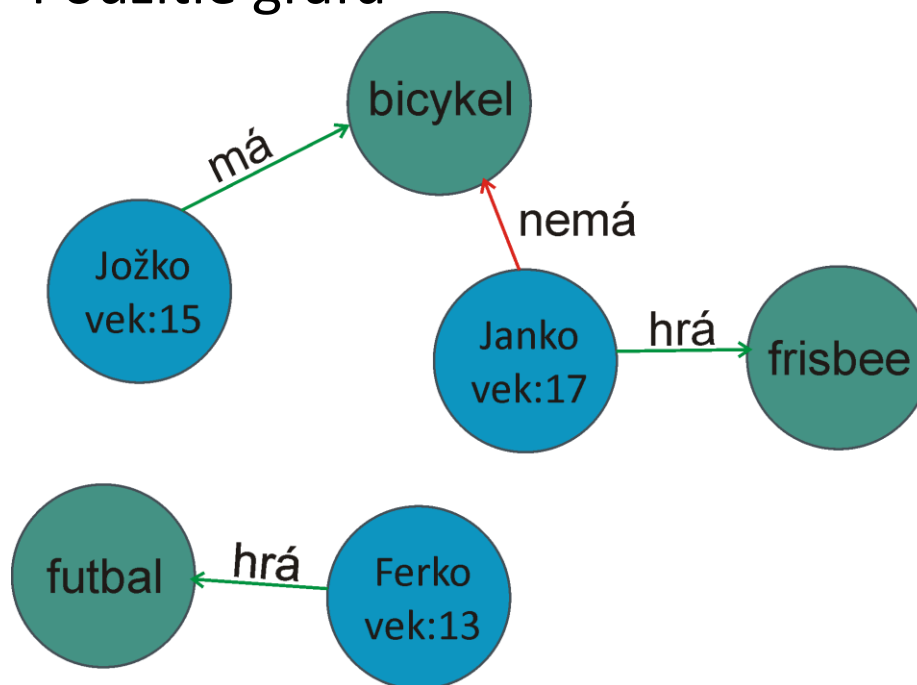
- Reprezentácia dát, množstvo
- Riedkosť matice
- Pomalý prístup k dátam - nutnosť prehľadávania
- Cold-start

Problémy odporúčania – riedka matica

Použitie matice

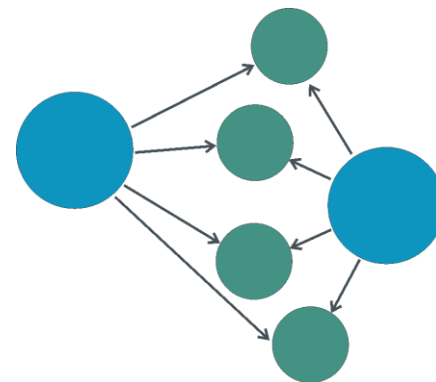
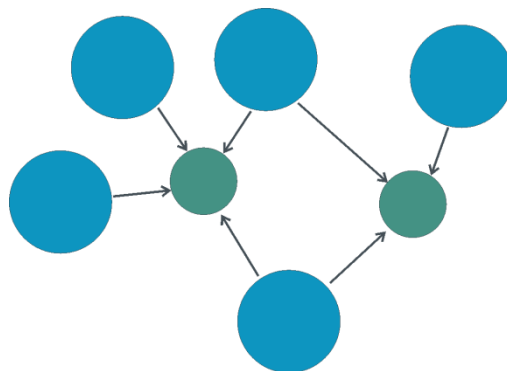
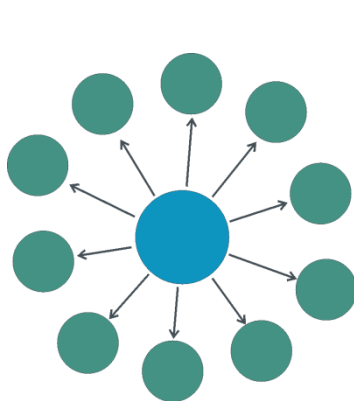
Používateľ	bicykel	futbal	frisbee	vek	...
Jožko	má			15	...
Ferko		hrá		13	...
Janko	nemá		hrá	17	...
...					

Použitie grafu



Riešenie a aplikácia

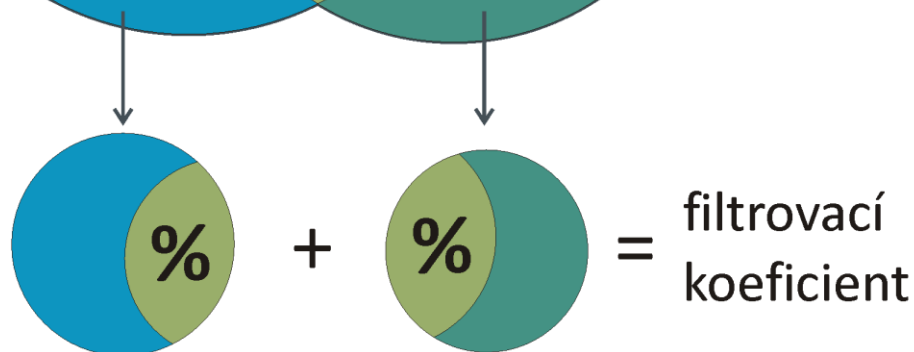
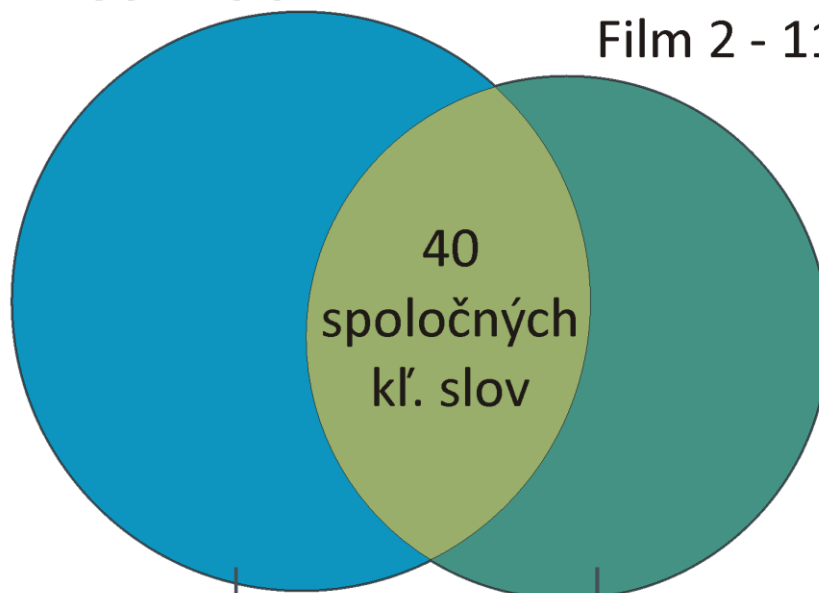
- Grafová databáza Neo4j
- Dátová sada IMDb – 49 000 filmov, 132 000 kľúčových slov
- Podobnosť medzi filmami opísanými kľúčovými slovami
- Na základe obsahu



Riešenie a aplikácia – filtrovací koeficient

Film 1 - 150 kľ. slov

Film 2 - 110 kľ. slov



$$\frac{40 \cdot 100}{150} + \frac{40 \cdot 100}{110} = 26,\overline{66} + 36,\overline{36} = 63,\overline{03}$$

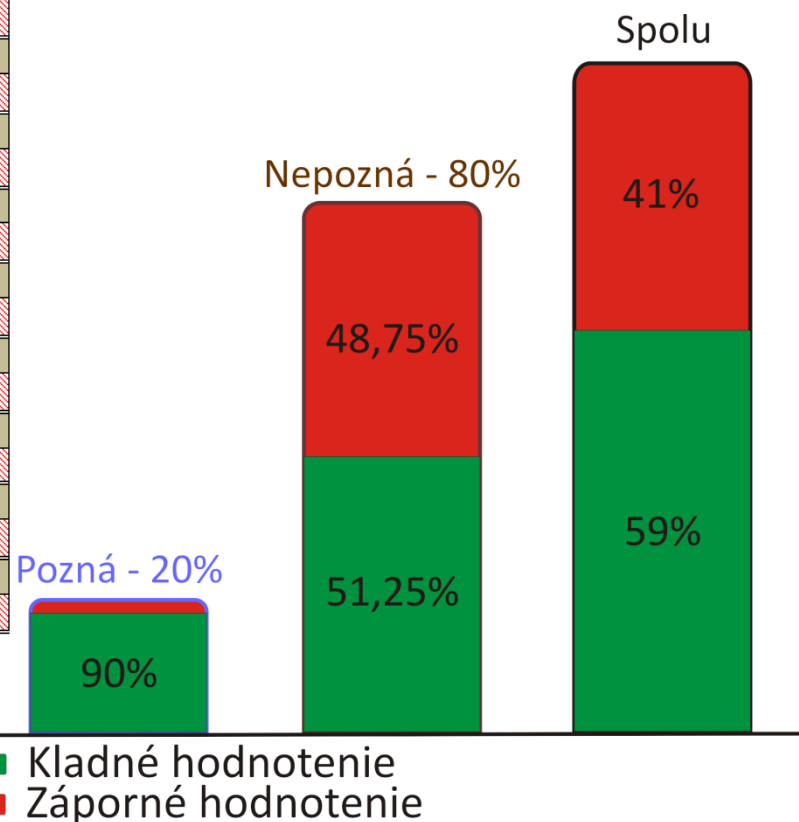
Testovanie metódy

- 10 ľudí
- 3 obľúbené filmy
- 10 nájdenných podobných filmov
- 2 otázky

Testovanie metódy - výsledky

Používateľ	Pozná		Nepozná	
	Hodnotí kladne	Hodnotí záporne	Hodnotí kladne	Hodnotí záporne
1	1	0	5	4
2	0	0	4	6
3	3	0	3	4
4	1	0	4	5
5	2	1	6	2
6	2	0	4	4
7	5	1	2	3
8	0	0	8	2
9	4	0	2	4
10	2	0	3	5

Používatelia poznali	20%	
	90%	10%
Používatelia nepoznali	80%	
	51,25%	48,75%
Spolu	Hodnotili kladne	Hodnotili záporne
	59%	41%



Možné vylepšenia

- Odporúčanie či podobnosť?
- Viac opisných atribútov
 - váhovanie kľúčových slov
 - žáner
 - hodnotenie používateľmi
- Vytvorenie a publikovanie funkčnej, dostupnej aplikácie
- Sledovanie rýchlosti a efektivity odporúčania a následne vylepšenie

Zhodnotenie

- ✓ Metóda využívajúca graf na reprezentáciu a prístup k dátam
- ✓ Aplikácia v doméne filmov
- ✓ Možnosť budúceho rozšírenia databázy
- ✗ Skôr hľadanie podobnosti ako odporúčanie
 - ✗ Vylepšenie odporúčania zohľadnením ďalších atribútov
- ✗ Vytvorenie použiteľnej služby
- ✗ Zameranie na efektivitu odporúčania