

Uživatelská dokumentace

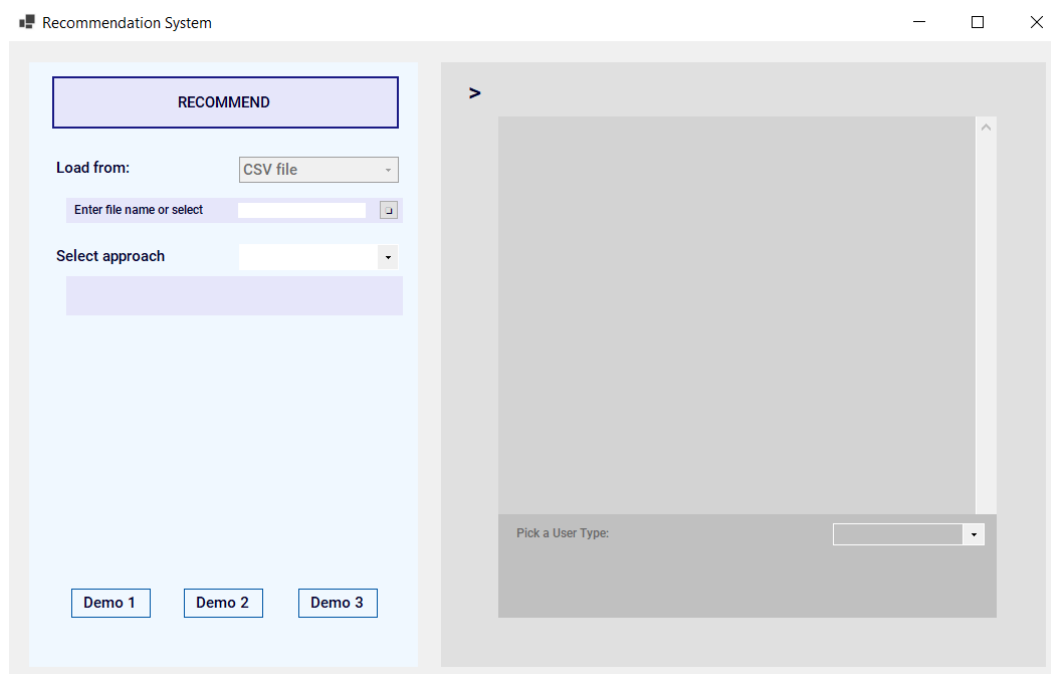
Jedná se o program, který slouží jako základ pro rozšiřitelné rozhraní implementující způsoby doporučení. Program byl zamýšlen spíše jako rozšiřitelná knihovna.

Vyvinut v jazyce C# verze 10.0 s .NET 6.0. Není potřeba žádná externí knihovna.

Návod k užívání programu

Windows Forms projekt:

Po spuštění se zobrazí následující okno:

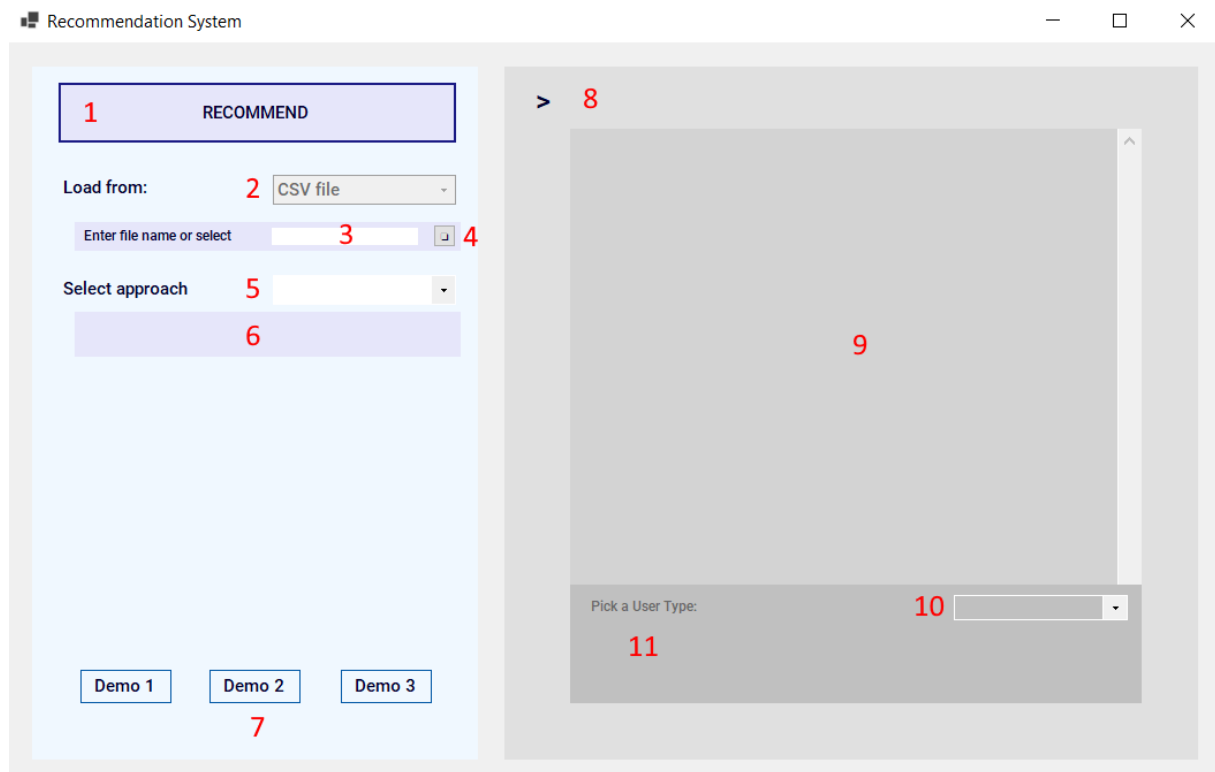


Aby program něco zvládl udělat, je nutné nejdříve zadat specifikace námi zvoleného doporučovacího systému, které budou dávat smysl.

Část vlevo slouží k výběru, odkud čerpáme data a jakým způsobem je chceme zpracovat.

Pravá část okna je téměř jen pro zobrazování výsledků. Nicméně, je nutno v ní definovat našeho příkladového uživatele.

Zde je následující okno pro orientaci popsáno detailněji:



1 – Tlačítko pro finální spuštění doporučovacího systému. Nejprve je nutné zvolit specifikace (konkrétněji v 2, 3 nebo 4, 5, 6, 10, 11). Výsledek se zobrazí v okně 9.

2 – Zde je možno si vybrat, pokud chceme dostávat data z CSV souboru, nebo jinak. Jiná varianta zatím není.

3 – Po napsání cesty k souboru a stisknutí klávesy Enter pro potvrzení se ukáže v okně 9 malá ukázka z načteného souboru.

4 – Viz bod 3, ale otevře se dialog se správcem souborů, kde je možno si soubor zvolit interaktivně.

5 – Zde se zobrazí typy způsobu doporučení.

Pro *Collaborative filtering* varianty je na výběr: *UserUserCfApproach* (pro doporučení na základě vztahů mezi uživateli) a *ItemItemCfApproach* (pro doporučení na základě vztahů mezi předměty). Oba přijímají data v číselném formátu.

Pro *Content based* variantu je tu: *StringSimilarityContentBasedApproach*, který se snaží pouze o výpočet podobností mezi jednotlivými záznamy v poskytnutých datech.

6 – Po výběru varianty v boxu 5 se zobrazí její parametry, které je nutno vyplnit.

7 – Demo tlačítka – pokud nechcete trávit čas s cherry-picking parametrů vašeho doporučovacího způsobu tak, aby dávaly smysl, lze stisknout jedno z nich a vše se vyplní – dokonce i uživatel. Nyní stačí jen stisknout tlačítko 1. Bonus: Pouze s kliknutím jednoho z Demo tlačítek je výsledný soubor převeden na reálné záznamy hodnocených předmětů.

Demo 1: Načte ukázku ze souboru *u.data*, který obsahuje záznamy hodnocení filmů specifickými uživateli v databázi, zvolí uživatele (fandu Sci-fi) a způsob doporučení – *UserUser*, který spočte podobnosti mezi uživateli (pomocí Cosine Similarity) a snaží se předpovědět, jak by uživatelé hodnotili filmy, které neviděli.

Demo 2: Načte ukázku ze souboru *subjects_11310.csv*, která obsahuje záznamy předmětů z Přírodovědecké fakulty UK, zvolí uživatele (fandu pokročilých kurzů) a způsob doporučení – *StringSimilarity*, který použije TF-IDF pro převedení dat na číselnou reprezentaci, poté použije Cosine Similarity pro vypočtení podobnosti. Snaží se najít podobné předměty těm, který si uživatel nadefinoval.

Demo 3: Načte ukázku ze souboru *u.data*, zvolí uživatele (fandu Sci-fi) a způsob doporučení – *ItemItem*, který spočte podobnosti mezi filmy (pomocí Cosine Similarity) a snaží se předpovědět, jak by uživatelé hodnotili filmy, které neviděli. Hodnoty jsou bohužel velice malé, díky velkému počtu filmů a málo hodnocení. Více lidí se dívá na známější filmy – u *UserUser* způsobu nebyl problém, protože spíše bude hodnotit každý uživatel než, že je každý film v databázi hodnocen.

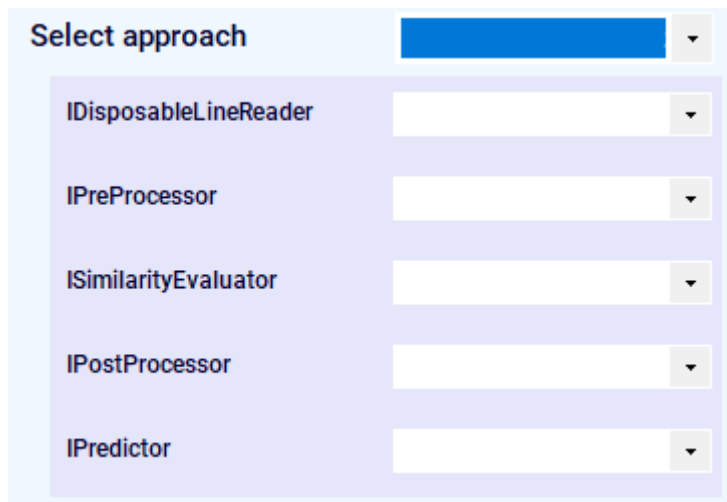
8 – Zde budou po stisknutí tlačítka 1 vidět tečky, které značí, že program je v běhu.

9 – Okno pro zobrazení výstupů programu.

10 – Zde je nutno si zvolit typ svého uživatele. Na výběr bude z: *MovieDbUser* (uživatel filmové databáze) a *SisUser* (uživatel databáze s předměty).

11 – V tomto oknu se po vybraní Dema nebo typu uživatele (10) zobrazí jeho příklad – jak by měl v okně být, při spuštění hlavní části programu (1), správně definován.

K bodu 6 – zobrazí se tyto boxy:



Select approach	
IDisposableLineReader	<input type="text"/>
IPreProcessor	<input type="text"/>
ISimilarityEvaluator	<input type="text"/>
IPostProcessor	<input type="text"/>
IPredictor	<input type="text"/>

IDisposableLineReader bude obsahovat *Readery*, se kterými se bude číst vstupní soubor – základní je **FileStreamLineReader**, který čte ze souboru řádek po řádku.

IPreProcessor bude obsahovat způsoby, kterými si původní data zpracují do matice. Na výběr bude **TF-IDF** (ten slouží hlavně pro zpracování slov, ne číselných záznamů) způsob, nebo převedení databázových záznamů do matice (**userItemMatrix** – ten používat pro číselné záznamy), kde řádky jsou uživatelé a sloupce jsou předměty, které hodnotí.

ISimilarityEvaluator bude obsahovat způsoby jakými lze mezi dvěma vektory vypočítat podobnost – Kosinova podobnost (**CosineSimilarity**), nebo Euklidovskou podobnost (**EuclideanSimilarity**). Hodnoty podobnosti jsou mezi -1 až 1.

IPostProcessor, kde na výběr bude **UserItemMatrixPP** (pro user-item matice) nebo **StringVectorSimilarityPP** (pro ContentBased Approach, kde je výsledek pouze jeden vektor), které vezmou dosud vytvořenou matici a zapíší ji do výsledného souboru. Každý uživatel bude mít vlastní řádek a jeho hodnocení budou seřazená (od nejlepšího po nejhorší).

IPredictor, který obsahuje **UserSimilarityPredictor** a **ItemSimilarityPredictor**. Jeden předpovídá hodnocení uživatele podle podobností uživatelů a druhý podle podobností předmětů.

Běh po stisknutí 1 by neměl být delší než 3 minuty. Pokud je ale ve snaze použít textovou podobnost (s TfIdf) pro soubor *u.data*, program poběží výrazně déle než 3 min (kolem 10). Při vybírání Approach je, bohužel, nutné brát velký zřetel na logiku programu.

Console projekt:

Po spuštění se zobrazí krátký úvod: „*This is a small scale recommendation system interface. Choose Approach, User as you need and get your recommendations! To view what you can do type 'help'.*“

Při použití příkazu 'help' se zobrazí dostupné příkazy (podobné funkcionalitám WinForms projektu):

loadcsv – Ve formátu *,loadcsv [file-name/path]'* je načten soubor. Po načtení je ukázka.

selectapproach – Uvede do dialogu s načtením způsobu doporučení a vyplnění parametrů zvoleného způsobu.

recommend – Jako tlačítko 1 ve WinForms. Je nutné mít vyplněné ostatní věci.

usertype – Dá na výběr typy uživatele.

useradd – Ve formátu *,useradd [index-vlastnosti-do-které-chceme-přidat] [co-přidat]'*.

userdemo – Pokud je zvolený typ uživatele, doplní jeho vlastnosti z dema.

userclear – Vyresetuje vlastnosti uživatele, nikoliv jeho typ.

user – Ukáže, jak má uživatel doposud vyplněné vlastnosti (jak vypadá).

summary – Ukáže souhrn informací, které budou použity k doporučení – uživatel, způsob doporučení, soubor.

demo – Vybere, stejným způsobem jako ve WinForms projektu, demo a ukáže, co se načetlo.

Uživatelé:

MovieDbUser: Udrží si seznam filmů/předmětů s jejich hodnocením.

SisUser: Má dva seznamy – jeden pro předměty, které se mu líbily, druhý pro předměty, které ho zajímají.

Interpretace výsledků:

Výsledky jsou uloženy do CSV souboru s názvem [způsob_doporučení]Results.csv a jsou zobrazeny v samotném programu. Ve výsledcích doporučení jsou předměty, které uživatel hodnotil.

Každý záznam je ve formátu [*identifikace_předmětu*] [*hodnocení*]. Každá řádka je jeden uživatel. První řádka je uživatel, kterého jsme definovali.

```
50 5000,174 4000,210 4000,222 4000,181 3237,100 2525,172
2460,127 2311,98 2280,56 2191,258 2137,121 2117,79 2008,7
1973,204 1969,117 1923,168 1881,195 1841,22 1808,96 1808,69
1766,183 1766,237 1758,318 1730,89 1726,64 1698,176 1685,257
1672,423 1654,151 1648,294 1627,405 1607,28 1574,288 1568,12
153...
1 5000,6 5000,9 5000,12 5000,13 5000,14 5000,15 5000,16
5000,19 5000,32 5000,42 5000,44 5000,45 5000,48 5000,50
5000,55 5000,57 5000,59 5000,60 5000,64 5000,81 5000,82
5000,86 5000,87 5000,89 5000,91 5000,93 5000,96 5000,100
5000,108 5000,109 5000,111 5000,113 5000,114 5000,115
5000,119 5000,124 500...
50 5000,100 5000,127 5000,242 5000,251 5000,272 5000,275
5000,283 5000,285 5000,302 5000,311 5000,313 5000,316 5000,1
4000,13 4000,14 4000,25 4000,111 4000,237 4000,255 4000,257
4000,269 4000,273 4000,276 4000,277 4000,279 4000,282
4000,284 4000,286 4000,292 4000,293 4000,295 4000,297
4000,299 4000,...
```

```
recommend
This might take a while. (max 3 minutes)
11 5000,8 4000,50 3532,174 3087,100 3001,10 3000,98 2945,56 2938,181 2808,172 2749,127 2600,79 2578,1 2483,7 2437,22 242
8,173 2428,64 2398,168 2384,12 2376,318 2363,204 2341,96 2328,69 2238,195 2209,210 2208,191 2126,258 2122,176 2106,183 2
105,117 2072,423 2067,121 2058,89 2056,20 2000,28 1995,216 1...
1 5000,6 5000,9 5000,12 5000,13 5000,14 5000,15 5000,16 5000,19 5000,32 5000,42 5000,44 5000,45 5000,48 5000,50 5000,55
5000,57 5000,59 5000,60 5000,64 5000,81 5000,82 5000,86 5000,87 5000,89 5000,91 5000,93 5000,96 5000,100 5000,108 5000,1
09 5000,111 5000,113 5000,114 5000,115 5000,119 5000,124 500...
50 5000,100 5000,127 5000,242 5000,251 5000,272 5000,275 5000,283 5000,285 5000,302 5000,311 5000,313 5000,316 5000,1 40
00,13 4000,14 4000,25 4000,111 4000,237 4000,255 4000,257 4000,269 4000,273 4000,276 4000,277 4000,279 4000,282 4000,284
4000,286 4000,292 4000,293 4000,295 4000,297 4000,299 4000,...
```

