|  |
| --- |
| OSTRAVSKÁ UNIVERZITA  PŘÍRODOVĚDECKÁ FAKULTA  KATEDRA INFORMATIKY A POČÍTAČŮ |
| Adaptivní systém pro automatizovaný návrh doporučených produktů v internetovém obchodě  DIPLOMOVÁ PRÁCE |
| Autor práce: Bc. Petr Fajmon  Vedoucí práce: RNDr. Bogdan Walek, Ph.D. |
| 2020 |

|  |
| --- |
| UNIVERSITY OF OSTRAVA  FACULTY NAME  DEPARTMENT NAME |
| Adaptive system for automated proposal of recommended products in the online store  THESIS |
| Author:  Bc. Petr Fajmon  Supervisor:  RNDr. Bogdan Walek, Ph.D. |
| 2020 |

(Zadání vysokoškolské kvalifikační práce)

ABSTRAKT

Tato práce se zabývá návrhem a implementací doporučovacího systému v internetovém obchodě zaměřeného na prodej potravin. Práce se dále zabývá návrhem vlastního řešení pro doporučování produktů pomocí expertního systému na základě vstupních parametrů. Těmi jsou - hodnocení produktu, podobnost produktů a koeficient získaný z historie nákupů. Práce obsahuje analýzu současného stavu možných přístupů, jejich problémů a řešení těchto problémů. Následně pak obsahuje návrh a implementaci vlastního řešení. Systém kombinuje kolaborativní přístup (použití algoritmu SVD na hodnocené produkty) a filtrování podle obsahu (použití algoritmu TF-IDF na prohlížené produkty) spolu s využitím expertního systému. Dále pak práce obsahuje testování navrženého systému na reálných uživatelích a výsledky jednotlivých testování spolu s rozborem těchto dat. Závěr práce se pak věnuje průběhu implementace a zhodnocení navrženého systému. Také jsou zde navrženy možné úpravy a vylepšení stávajícího systému.

*Klíčová slova: doporučovací systém, expertní systém, personalizovaný obsah, PHP, Nette*

**ABSTRACT**

The thesis is focused on design and implementation of recommendation system in the online grocery store. The thesis also focused on design of own solution for product recommendations using and expert system based on input parameters. This thesis includes and an analysis of the current state of possible approaches, their problems and solutions of these problems. This thesis further contains the design and implementation of its own solutions together with the use of an expert system. The system combines a collaborative filtering (SVD algorithm) and content-based filtering (TF-IDF algorithm) together with the use of and expert system. Furthermore, the thesis contains testing of the proposed system on real users and the results of individual testing with the analysis of this data. In the end, the thesis is focused on course of implementation and evaluation of the proposed system. There are also designed possible modifications and improvments of current system.

*Keywords: recommendation systém, expert systém, personalized content, PHP, Nette*

čestné prohlášení

Já, níže podepsaný student, tímto čestně prohlašuji, že text mnou odevzdané závěrečné práce v písemné podobě je totožný s textem závěrečné práce vloženým v databázi DIPL2.

Ostrava dne

………………………………

podpis studenta

|  |
| --- |
| Poděkování  Rád bych poděkoval vedoucímu své diplomové práce panu RNDr. Bogdanu Walkovi, Ph.D. za pomoc při vytváření této diplomové práce, za odbornou pomoc, užitečné rady, pomoc v průběhu testování a konzultace. Rád bych poděkoval rodině a přátelům za podporu při tvorbě této diplomové práce. Nakonec bych rád poděkoval všem, kteří se účastnili jednotlivých testování a především těm, kteří jednotlivá testování poté konzultovali. |
| Prohlašuji, že předložená práce je mým původním autorským dílem, které jsem vypracoval samostatně. Veškerou literaturu a další zdroje, z nichž jsem při zpracování čerpal, v práci řádně cituji a jsou uvedeny v seznamu použité literatury.  V Ostravě dne . . . . . . . . . . . .  . . . . . . . . . . . . . . . . . .  (podpis) |

OBSAH

[ÚVOD 10](#_Toc71526396)

[Cíle práce 11](#_Toc71526397)

[1 Analýza současného stavu 12](#_Toc71526398)

[1.1 Doporučovací systém 12](#_Toc71526399)

[1.2 Kolaborativní filtrování 15](#_Toc71526400)

[1.2.1 Problémy kolaborativního filtrování 16](#_Toc71526401)

[1.2.2 Typy kolaborativního filtrování 17](#_Toc71526402)

[1.3 Filtrování podle obsahu 18](#_Toc71526403)

[1.3.1 Typy filtrování podle obsahu 19](#_Toc71526404)

[1.4 Hybridní přístup 20](#_Toc71526405)

[1.5 Závěr přístupů 23](#_Toc71526406)

[2 Návrh 25](#_Toc71526407)

[2.1 Architektura 25](#_Toc71526408)

[2.2 Použité algoritmy v systému 26](#_Toc71526409)

[2.3 Přístup pro doporučení produktů 34](#_Toc71526410)

[2.3.1 Přístupy využívající části systému 43](#_Toc71526411)

[3 Implementace 57](#_Toc71526412)

[3.1 Programovací jazyk 57](#_Toc71526413)

[3.2 Databáze 58](#_Toc71526414)

[3.3 Webové rozhraní aplikace 59](#_Toc71526415)

[3.3.1 Administrativní část 59](#_Toc71526416)

[3.3.2 Uživatelská část 62](#_Toc71526417)

[4 Testování 66](#_Toc71526418)

[4.1 Popis testování 66](#_Toc71526419)

[4.1.1 První testování 66](#_Toc71526420)

[4.1.2 Druhé testování 66](#_Toc71526421)

[4.1.3 Expertní systém 67](#_Toc71526422)

[4.1.4 Košík.cz 68](#_Toc71526423)

[4.2 Výsledky testování 68](#_Toc71526424)

[4.2.1 První testování 68](#_Toc71526425)

[4.2.2 Druhé testování 77](#_Toc71526426)

[4.2.3 Expertní systém 88](#_Toc71526427)

[4.2.4 Košík.cz 90](#_Toc71526428)

[4.3 Závěr testování 93](#_Toc71526429)

[ZÁVĚR 95](#_Toc71526430)

[RESUMÉ 97](#_Toc71526431)

[SUMMARY 98](#_Toc71526432)

[SEZNAM POUŽITÉ LITERATURY 99](#_Toc71526433)

[SEZNAM POUŽITÝCH SYMBOLŮ 101](#_Toc71526434)

[SEZNAM OBRÁZKŮ 102](#_Toc71526435)

[SEZNAM TABULEK 104](#_Toc71526436)

[SEZNAM PŘÍLOH 105](#_Toc71526437)

ÚVOD

V současné době nám internet velice pomáhá. Používáme jej ke každodenním činnostem jako je vyhledávání informací, komunikace skrz sociální sítě, čtení elektronických knih, nebo nakupování v internetovém obchodě. Také využíváme internet ve volném čase. Například při nakupování přes e-shop, sledováním různých videí na YouTube nebo online filmů na platformě Netflix.

Doručit relevantní obsah aktivnímu uživateli je tedy žádanou schopností současných webových aplikací. Kromě již zmíněných streamovacích platforem a e-shopů se může jednat také o zpravodajské servery nebo sociální sítě. Hlavním cílem doporučování produktů v e-shopech je zajistit aby si uživatel dané produkty zakoupil a obchod tak získal více peněz. Podobně je to na streamovacích aplikacích, které získávají peníze skrz reklamy. Například na platformě YouTube je důležité doporučovat relevantní obsah, aby uživatel shlédl více videí a s nimi i patřičné množství reklam. Naproti tomu doporučování na platformě Netlix chce docílit zakoupení doporučené položky, nebo předplatného k získání výhod.

Všechny tyto platformy se snaží uživateli zjednodušit vyhledávání právě pomocí doporučování a zajistit aby byl zákazník spokojený. Jelikož spokojený zákazník, kterému jsou doporučovány produkty podle jeho preferencí, se pravděpodobně vrátí. Právě toho se snaží platformy, nebo e-sohpy docílit pomocí doporučovacích systémů.

Doporučovací systém je druh filtrování obsahu, který se snaží odhadnout hodnocení nebo preference aktivního uživatele ve webové aplikaci.

V současné době se nakupuje přes e-shopy mnohem častěji. Lidé již nenakupují pouze elektroniku, nábytek a oblečení. Čím dál více nakupují přes internetové obchody také potraviny. A proto je žádoucí aby tyto webové aplikace tuto schopnost měly. Tato práce se zabývá návrhem a implementací adaptivního systému pro doporučování produktů v internetovém obchodě s potravinami.

Cíle práce

Hlavním cílem práce je návrh a implementace adaptivního systému pro automatizovaný návrh doporučených produktů v internetovém obchodě. Součástí systému bude návrh hybridního systému s využitím expertního systému. Finálním výsledným seznamem produktů bude výstup ES.

**Dílčí cíle práce**

* Doporučení relevantních produktů na základě uživatelovy interakce se systémem - prohlížení produktů, hodnocení produktů, nakupování produktů.
* Návrh ES s využitím vstupních proměnných získaných z uživatelovy interakce se systémem.
* Praktické ověření navrženého systému.

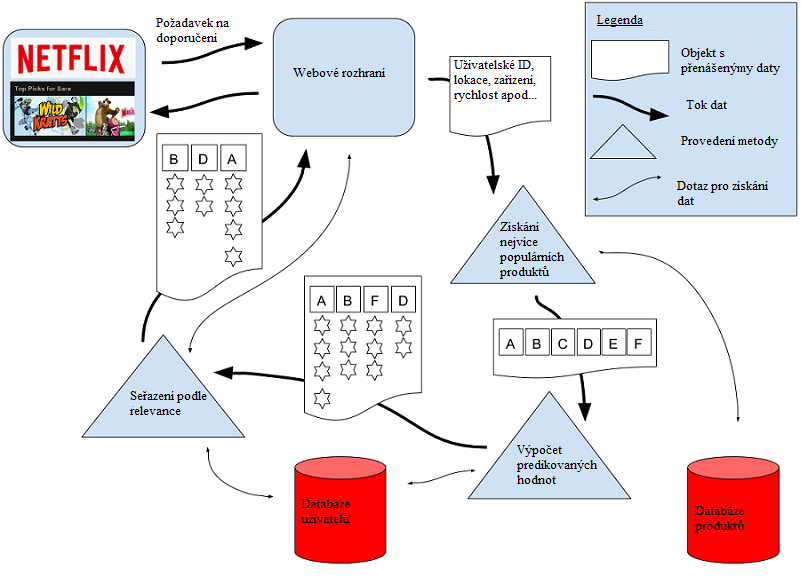
1. Analýza současného stavu
   1. Doporučovací systém

„Doporučovací systém dokáže vykalkulovat a poskytnout relevantní obsah konečnému uživateli, na základě znalosti o uživateli, obsahu a interakce mezi uživatelem a produktem.“ [1]

Doporučovací systém pracuje v několika krocích:

1. Zaslání požadavku na zobrazení stránky s produkty.
2. Server zavolá doporučovací systém, který pomocí doporučovacích metod doručí uživateli relevantní produkty.
3. Získání několika produktů na základě uživatelských preferencí z databáze produktů.
4. Predikce vhodných produktů na základě uživatelských dat z databáze uživatele.
5. Výběr relevantních produktů a jejich seřazení podle uživatelského hodnocení, obsahu a jiných pravidel.
6. Server zobrazí požadovanou stránku s relevantním obsahem.

Následující obrázek graficky znázorňuje výše popsaný proces doporučení na streamovacím webu Netflix.



Obrázek 1: Doporučení produktů na stránce Netflix (zdroj [1])

Před začátkem implementace doporučovacího systému je dobré si ujasnit, jaký by měl náš doporučovací systém být. Je několik okruhů, které popisují náš vytvářený systém:

**Doména**

Typ obsahu, který systém doporučuje uživateli. Může to být doporučování hudby, filmů, produktů a jiné možnosti. Je důležité si uvědomit, že složitost systému závisí na typu domény. Pokud systém bude doporučovat hudební skladby, není velký problém špatně doručená skladba. Zatímco velkou složitost bude mít systém, který rozhoduje v celoživotních situacích jako například doporučení pěstounů. [1]

**Účel**

Jaký je cíl, pro který systém vytváříme. Pro zákazníky to může být lepší vyhledávání vhodných produktů, nebo podobný obsah na jednom systému. Vlastník systému se snaží získat nové zákazníky a stálé, kteří se rádi vracejí a přináší tak vlastníkovi zisk. Například z prémiových účtů s měsíčním paušálem pro neomezené možnosti na stránkách. [1]

**Kontext**

Jakým způsobem bude doporučení uživateli doručeno? Kontext je také o poloze uživatele, jaká je zrovna denní doba, co uživatel právě dělá (například studuje a potřebuje rychlé rozhodnutí), počasí venku poblíž uživatele, ale také o náladě uživatele. [1]

**Úroveň personalizace**

Jakou úroveň personalizace bude systém využívat? Personalizace má několik stupňů, od základních statistických metod po individuální data o uživateli (nepersonalizovaná, částečně personalizovaná, personalizovaná). [1]

**Čí názor**

Jakým způsobem bude doporučení uživateli zadáváno? Bude fungovat s manuálním zadáváním doporučení expertem pro danou oblast, nebo bude systém plně automatizovaný. V dnešní době se manuální expertní systémy moc nevyskytují. [1]

**Soukromí a důvěryhodnost**

Jak bude systém chránit zákazníkovo soukromí? Jak bude nakládáno s informacemi o uživateli? Data umístěna na stránce o uživateli by neměla být poskytovány třetím stranám, které by na základě dat nabízely uživateli další produkty. Pokud například systém doporučuje nejrůznější zdravotní pomůcky, uživatel si pořídí speciální zubní pastu a vzápětí na to mu přijde email od jeho zubaře se speciální nabídkou. [1]

Bude zákazník důvěřovat systému? Důvěryhodnost naznačuje, do jaké míry bude zákazník doporučení důvěřovat. Mnoho zákazníků se na stránkách s doporučovacím systémem mohou cítit manipulováni. Tím ztratí stránka na důvěryhodnosti a zákazník se na stránku nemusí dále vracet. Je důležité, aby produkty byly zákazníkovy doporučovány tak, aby měl pocit, že není manipulován. [1]

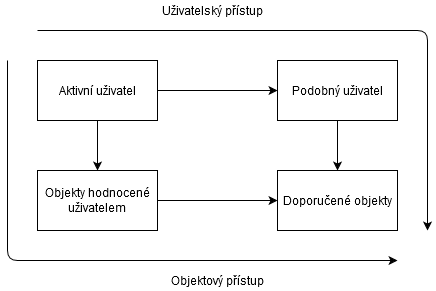
**Rozhraní**

Jaké budou vstupy a výstupy doporučovacího systému? Vstupem může být manuální vkládání dat o uživateli (explicitní) nebo automatické, systém se snaží odhadnout data podle toho, jak uživatel pracuje s doporučovacím systémem (implicitně). Výstupem může být předdefinovaná část stránky (organická část) nebo jasná část stránky, která bude označena „Doporučeno pro Vás“, nebo „Mohlo by Vás zajímat“ (anorganická). [1]

**Algoritmy**

Jaký algoritmus bude systém používat? Algoritmy využívající se pro doporučování se dělí na kolaborativní filtrování, filtrování podle obsahu a hybridní systémy, které kombinují již zmíněné skupiny. Tyto přístupy jsou popsány níže. [1]

* 1. Kolaborativní filtrování

Kolaborativní filtrování je založeno na vypočítávání všech různých možností doručení objektů k uživateli. Je to seznam objektů, určených přímo pro aktivního uživatele. Tento seznam je založen na uživatelích s podobným prohlížením a doručuje aktivnímu uživateli podobné objekty, které mu ještě nebyly doručeny. [1] [13]

Obrázek 2:Kolaborativní filtrování

Kolaborativní filtrování založené na podobnosti uživatelů využívá dvě možnosti doručení objektu. První možnost využívá nalezení uživatele, který má podobné zájmy jako aktivní uživatel a doručení podobného objektu aktivnímu uživateli. Druhá možnost nejprve vybere objekty, které aktivní uživatel viděl a následně doručí podobné objekty aktivnímu uživateli. Předpokládá se, že uživatelé, kteří měli podobné zájmy v minulosti, budou mít podobné zájmy v budoucnosti. [1] [13]

Podobnost objektů je ohodnocována „ratings“ a vkládána do tabulky. Řádky tabulky představují uživatele a sloupce představují objekty. Každé pole tabulky představuje rating objektu daného uživatele. Objekty mohou mít žádný, nízký až vysoký rating. Následně se aktivnímu uživateli najde uživatel s podobným ratingem objektů a doporučí aktivnímu uživateli objekt, který ještě neviděl. [1] [13]

* + 1. Problémy kolaborativního filtrování

**Studený start (Cold start)**

Studený start je pojem, kdy doručovací systém nemá dostatek dat o novém uživateli pro doručení vhodných produktů, tzv. studený návštěvník (Cold visitor). Podobný problém studeného startu je problém šedé ovce (Gray sheep), kdy uživatel má specifické natolik preference, že je těžké vytvořit analýzu na základě podobných uživatelů.[1][13]

**Studený produkt (Cold product)**

Nová položka, přidaná do seznamu na webových stránkách, je tzv. studený produkt. Tento produkt potřebuje manuální propagaci, například zasílání emailů, zavěšení na webu do sekce „novinky“, nebo nabízení uživatelům, kterým se líbí podobné produkty. Možným způsobem je také uměle navýšit popularitu produktu, pokud však o produkt není zájem, pomalu popularitu produktu snížit. [1][13]

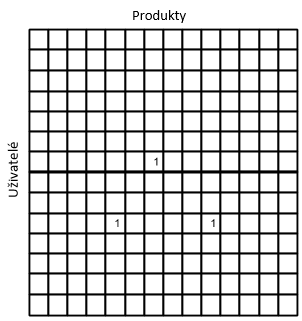
**Studený návštěvník (Cold visitor)**

Podobný problém jako studený start. Uživatel by měl ohodnotit několik produktů, aby doručovací systém mohl zjistit preference uživatele. To ale může chvíli trvat a my potřebujeme doručit uživateli produkty co nejdříve. Jedním z řešení je uživateli při první návštěvě nabídnout k ohodnocení několik produktů, tím však systém nemusí dostat dostatečné množství dat, nebo relevantní data. [1][13]

**Šedá ovce (Gray sheep)**

Je uživatel, který má natolik specifické preference, že nejsou jiní uživatelé, kteří se zajímali o stejné produkty jako šedá ovce. [1][13]

**Řídkost (Sparsity problem)**

Uživatelé a produkty jsou reprezentovány user-item maticí. Řádky matice jsou uživatelé a sloupce produkty, buňky matice znamenají jednotlivé hodnocení uživatele daného produktu. Většina obchodů má mnoho uživatelů a tisíce produktů. Každý uživatel si však zakoupí pouze jeden nebo pár produktů, tudíž user-item matice je spíše prázdná a vypadá jako matice na obrázku. [1][13]

Obrázek 3: Problém prázdnoty (zdroj:[1])

* + 1. Typy kolaborativního filtrování

**Memory-based**

Tento přístup se rozděluje na dvě hlavní sekce, které k doporučení využívají pouze informace z user-item matice. Kterými jsou doporučení produktů na základě podobnosti uživatelů (user-user) a doporučení produktů na základě podobnosti produktů. Filtrování na základě podobnosti uživatelů se nejdříve snaží najít podobné uživatele k aktivnímu uživateli, aby navrhl produkty, jež jsou mezi podobnými uživateli nejoblíbenější, a které jsou pro aktivního uživatele nové. Pro doporučení nehodnoceného produktu aktivnímu uživateli musíme nejprve vytvořit vektory všech uživatelů se všemi produkty. Poté můžeme najít podobnost mezi těmito vektory. Jakmile jsou nalezeny všechny podobnosti s každým uživatelem, můžeme ponechat k-nejbližších sousedů a navrhnout aktivnímu uživateli nejoblíbenější produkty.[10][11]

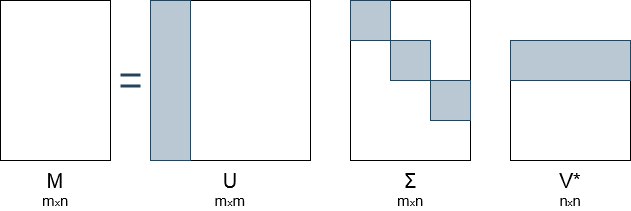
Na druhou stranu filtrování na základě podobnosti mezi produkty najde všechny produkty aktivního uživatele a jeho hodnocení daných produktů. Dva produkty se považují za podobné, pokud většina uživatelů, kteří hodnotili oba produkty, udělili oběma produktům podobné hodnocení. Každý produkt je v matici představován jako vektor. Pro doporučení nehodnoceného produktu aktivnímu uživateli, je vybrán produkt již ohodnoceného produktu a spočítán vektor s každým uživatelem, který daný produkt hodnotil. Následně je vypočítána podobnost mezi těmito vektory a vybráno k-nejbližších sousedů pro doporučení nejlepších produktů aktivnímu uživateli. [10][11]

Pro výpočet takové podobnosti se využívá kosinová podobnost (cosine similarity) nebo Pearsonův korelační koeficient (Pearson correlation coefficient).

**Model-based**

Tento přístup spoléhá pouze na informace o interakcích mezi uživatelem a produkty, a předpokládají model, který má tyto interakce vysvětlit. Hledá skryté položky v datech matice pomocí dekompozice matic. Může se jednat například o algoritmy maticové faktorizace, které rozloží velkou řídkou interakční matici na dvě menší matice, uživatelskou matici faktorů a produktovou matici faktorů, které následně mezi sebou vynásobí. Dalším způsobem maticové dekompozice je algoritmus SVD (Singular-Value Decomposition). [10][11]

Singulární rozklad matice (SVD) je rozklad komplexní nebo reálné matice **M** na maticový součin **UΣV\***, kde **M** je vstupní matice o rozměrech *m* x *n*. **U** je komplexní unitární matice o rozměrech *m* x *m*, **Σ** je diagonální matice s nezápornými reálnými čísly o rozměrech *m* x *n* a **V**je unitární komplexní matice o rozměrech *n* x *n*. [12]



Obrázek 4: Maticový rozklad SVD (zdroj: vlastní)

* 1. Filtrování podle obsahu

Filtrování podle obsahu je lehce komplikovanější než kolaborativní filtrování. Jedná se o získávání informací z objektu, snaha definovat každý objekt jako list hodnot. Filtrování podle obsahu souvisí i s uživatelským profilem. Objekty, které jsou podobné uživatelskému profilu, se doručují uživateli. [1]

Filtrování podle obsahu vybírá objekty podle podobnosti obsahu objektu a preferencí uživatele, na rozdíl od kolaborativního filtrování, které vybírá objekty podle podobnosti uživatelů. [1]

Existují dva hlavní modely získávání informací. Prvním je model založený na preferencích uživatele a druhý je založený na historii interakce uživatele s doporučovacím systémem.

Pro vytvoření fungujícího algoritmu je potřeba vymezit části algoritmu.[1]

* **Analýza obsahu** - vytvoření modelu založeného na obsahu objektů. Vytvoření profilu každého objektu.
* **Uživatelský profil** - vytvoření uživatelského profilu. Seznam objektů, které uživatele zajímají.
* **Dodání objektu** - dodání objektu na podobnosti uživatelského profilu a obsahu objektu.
  + 1. Typy filtrování podle obsahu

**Content only**

Tento přístup doporučí uživateli podobné produkty již dříve hodnocených, nebo zakoupených. Systém nejdříve nalezne všechny možné kombinace podobnosti produktů, následně z nich vybere ty, které uživatel již hodnotil, nebo zakoupil a vygeneruje seznam produktů na doporučení. [2]

**User profile**

Tento přístup doporučí uživateli produkty pouze na základě jeho dřívější interakce se systémem. Systém pracuje s produkty již hodnocenými uživatelem. Vyhne se problému studeného startu.

Algoritmus TF-IDF je metodika hodnocení relevance při vyhledávání v textu. Každý produkt je reprezentován jak TF-IDF vektor. Název je spojením zkratek dvou termínů: TF (term frequency) a IDF (inverse document frequency).

Term frequency znamená četnost daného slova v aktuálním dokumentu. Znamená výskyt slova v dokumentu a dává větší váhu, když je slovo častější. Vypočítá se podle vzorce.

Inverse document frequency nám udává, kolik dokumentů obsahuje dané slovo. Znamená to vzácnost slova, protože slovo vyskytující se v dokumentu je menší, než se zvyšuje IDF. Pomáhá dávat vyšší skóre vzácným výrazům v dokumentech. Vzorec pro IDF je následující.

Konečně TF-IDF je součin složek TF a IDF a slouží jako měřítko k vyhodnocení důležitosti slova pro dokument v těle dokumentů. Význam slova se zvyšuje úměrně s počtem výskytů slova v dokumentu, ale je kompenzován frekvencí slova v těle dokumentů. Vzorec pro výpočet TFIDF je následující.

Protože se metoda při rozlišování jednotlivých položek silně spoléhá na popis, je nutné mít popis co nejpřesněji odpovídající dané položce. Může se tedy do popisu zahrnout také nadpis, slogan, kategorie. Pro specifičtější popis je možné zbavit jej tzv. „stopwords“ - slovy, která se v popise vyskytují, ale nepřidávají mu žádný význam. Takovými slovy mohou být například předložky, spojky a další (např.: aby, který, jsem, zda). [3][4]

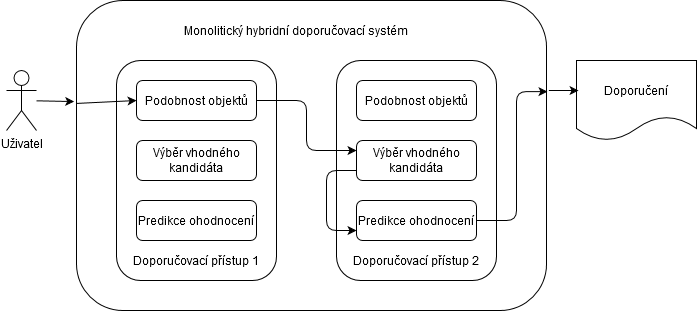
* 1. Hybridní přístup

Hybridní přístup je kombinace několika přístupů do jednoho systému. Hybridní systémy se dělí do 3 hlavních skupin.

* Monolitické - Monolithic hybrid recommenders
* Smíšené - Mixed hybrid recommenders
* Kombinované - Ensemble recommenders

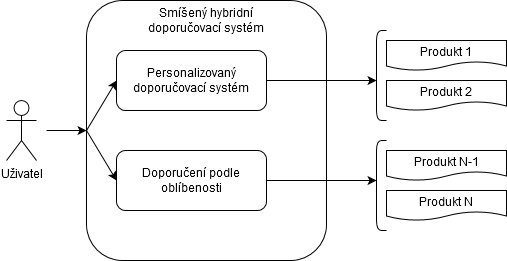
**Monolitické**

Monolitické doporučovací systémy kombinují části různých algoritmů do jednoho systému. Tyto přístupy mohou například využít algoritmy kolaborativního filtrování a doplnit je algoritmy filtrování podle obsahu k doporučení vybraného objektu uživateli. [1]



Obrázek 5: Monolitický hybridní doporučovací systém (zdroj: [1])

**Smíšené**

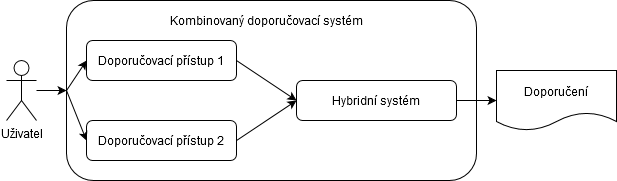
Smíšené doporučovací systémy doplňují výsledky jednoho vybraného algoritmu dalšími výsledky z jiného vybraného algoritmu. Například pokud algoritmus doporučující personalizované produkty uživateli doručí N výsledků, tak je doplněn do M požadovaných výsledků jiným algoritmem doporučujícím například oblíbené produkty všech uživatelů. [1]

Obrázek 6: Smíšený hybridní doporučovací systém (zdroj: [1])

**Kombinované**

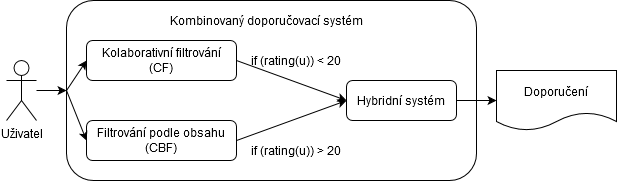
Rozdíl mezi kombinovaným přístupem a smíšeným hybridním přístupem je ten, že smíšený hybridní přístup vždy ukazuje vše z jednoho doporučovacího algoritmu doplněného jiným doporučovacím algoritmem. [1]

V kombinovaném systému může být výsledné doporučení objektů takové, že se udělá průnik výsledků několika doporučovacích algoritmů. Tento způsob není obvyklý a spíše se můžeme setkat s přepínacími kombinovanými systémy a váhovými kombinovanými systémy. [1]



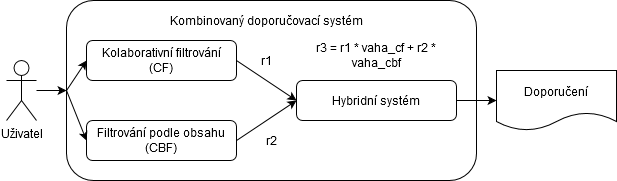
Obrázek 7: Kombinovaný doporučovací systém (zdroj: [1])

**Přepínací kombinovaný systém** přepíná mezi doporučovacími systémy na základě kontextu a dodává všechny výsledky jednoho doporučovacího systému. Například podle země původu, ze které uživatel pochází, nebo podle denní doby. [1]



Obrázek 8: Přepínací kombinovaný doporučovací systém (zdroj: [1])

**Váhový kombinovaný systém** dává jednotlivým doporučovacím systémům váhy. Váhy mohou být stanoveny předem expertem pro danou doménu a platí pro celý systém, nebo se mohou váhy stanovit na jednotlivé objekty. [1]



Obrázek 9: Váhový kombinovaný doporučovací systém (zdroj: [1])

* 1. Závěr přístupů

**Kolaborativní filtrování**

Výhodou kolaborativního filtrování je, že systém potřebuje znát pouze informace získané z user-item matice, tedy kombinace uživatele, produktů a jejich hodnocení. Systém nepotřebuje mít informace o uživatelích nebo o produktech. Další výhodou je využití kolaborativního filtrování na zaběhlých doménách s velkým množstvím dat, například filmy nebo obchody.

Nevýhodou je problém studeného startu (cold-start), kdy systém nemá dostatek dat o uživateli nebo o produktu pro správné doporučení. Pro správné doporučení je vhodné, aby uživatel ohodnotil nějaký počet produktů, například 50, obchody ale potřebují doporučovat produkty mnohem dříve. Některé obchody problém řeší nabídkou ohodnocení několika produktů při první návštěvě.

Navazujícím problémem je problém prázdnoty (sparsity problem). User-item matice je vyplněna velice zřídka a naleznout podobné produkty je pro algoritmus těžší. Obchody mají ve své databázi tisíce produktů, ale uživatelé si zakoupí pouze jeden nebo pár produktů.

Kolaborativní filtrování je spíše vhodné využít na velké, již zaběhlé obchody s velkým množstvím dat a informací o uživatelích. Naproti tomu se nevyplatí využívat na nově vytvářejícím se obchodě, který má minimum dat a málo uživatelů.

**Filtrování podle obsahu**

Výhodou filtrování podle obsahu je, že systém dělá analýzu na základě informací o produktu pro správné doporučení. Systém nepotřebuje velké množství dat jako kolaborativní filtrování a dokáže uživateli doporučit produkt pouze podle jeho historie nebo zvolených kategorií.

Filtrování podle obsahu se lépe využije v obchodech s menším množstvím dat a uživatelů, nebo v začínajících obchodech, kdy obchod má pouze několik produktů a minimum uživatelů. Dobré využití je také na doménách, které nepoužívají hodnocení, například informační portály nebo zpravodajství (www.novinky.cz, [www.dvojka.rozhlas.cz](http://www.dvojka.rozhlas.cz)).

**Hybridní systémy**

Výhodou hybridních systémů je využití výhod výše zmíněných filtrování a předejít tak jejich problémům jako je například problém studeného startu. Tyto systémy dokáží pomocí pravidel doručit aktivnímu uživateli nejrelevantnější výsledky, díky zvolené správné kombinace výše zmíněných filtrování. Sice se hybridní systémy jeví jako nejlepší možnost, ale kvůli jejich komplexnosti a náročnosti na zvolení správné kombinace a implementaci bývají v praxi využity tyto hybridní systémy velice zřídka.

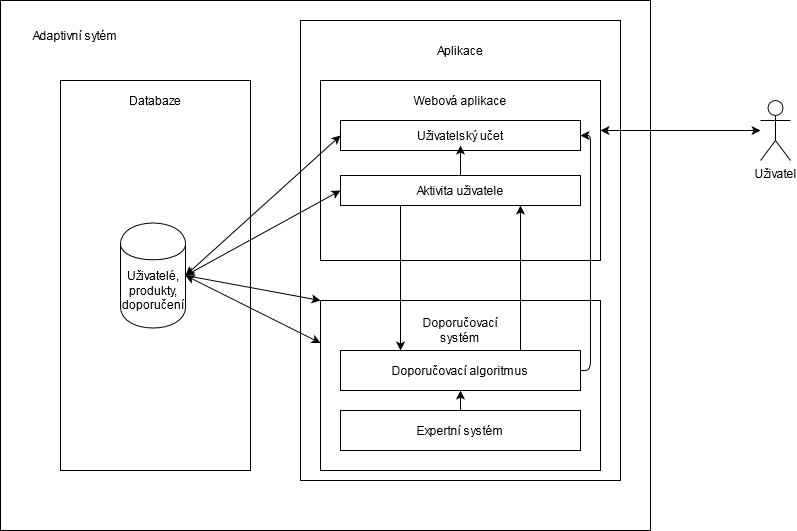
Využití těchto hybridních systémů je vhodné pro středně velké obchody. Hybridní systém aplikován na středně velký internetový obchod může doporučit uživatelům nejrelevantnější produkty a vyhnout se velkým výpočtům kvůli enormním databázím.

1. Návrh

Navržený adaptivní hybridní systém (kombinace algoritmů SVD a TF-IDF) bude zaměřen na doručení obsahu uživateli pomocí dat získaných ze 3 vstupů. Systém bude využívat data získaná od uživatele při vstupním nastavení, při registraci. Nejprve bude uživateli doporučen obsah podle jeho zvolených preferencí. Následně při postupné interakci bude systém získávat data o prohlížení jednotlivých produktů, jejich hodnocení a následné zakoupení daných produktů.

* 1. Architektura

Architektura navrhovaného systému je znázorněna na následujícím diagramu. Je rozdělena na aktivního uživatele a adaptivní systém. Adaptivní systém se skládá z aplikace a databáze, přičemž aktivní uživatel komunikuje s částí webové aplikace.



Obrázek 10: Architektura navrhovaného systému (zdroj: vlastní)

Jak bylo zmíněno výše, aktivní uživatel komunikuje s webovou aplikací, která zaznamenává aktivitu uživatele v aplikaci a ukazuje mu nejrelevantnější produkty. Webová aplikace komunikuje s databází, do které ukládá aktivitu uživatele na webu. Doporučovací systém získává informace o aktivitě uživatele buď přímo, například prohlížení produktů, nebo z databáze, pokud uživatel hodnotil nějaké produkty, či provedl nějaké nákupy. Na základě těchto informací sytém postupně doručuje nejrelevantnější produkty, právě podle uživatelovy interakce se systémem.

* 1. Použité algoritmy v systému

Algoritmus TF-IDF je zástupce ze skupiny - filtrování podle obsahu. Tento algoritmus je aplikován v systému dvakrát, nejprve doporučuje podobné produkty ke skupině prohlížených produktů aktivního uživatele, a později je využit při doporučení podobných produktů k zakoupeným produktům aktivního uživatele. Modifikace algoritmu na jednotlivé přístupy budou popsány níže.

**Registrace**

Uživatel při registraci zvolí oblíbené kategorie, podle kterých systém vyhodnotí prvotní doručený obsah na základě těchto kategorií. Algoritmus nejprve zjistí podobné uživatele, následně získá jejich hodnocené produkty a vrátí uživateli top produkty skupiny podobných uživatelů. Výsledné produkty budou seřazeny podle počtu hodnocení a průměrného hodnocení daného produktu.

1. Uživatel se zaregistruje a zvolí si oblíbené kategorie.
2. Algoritmus aktivního uživatele zařadí do skupiny **U** podobných uživatelů na základě podobných kategorií.
   1. Vytvoří se pole kategorií aktivního uživatele, pole kategorií ostatních uživatelů a je vypočtena podobnost, na základě počtu podobných prvků v poli.
   2. Uživatelé s podobností větší než 50% jsou zařazeni do skupiny **U** spolu s aktivním uživatelem. Pokud neexistuje žádný uživatel s podobností větší než 50%, není vytvořena skupina **U**.
3. Pokud existuje skupina **U**, algoritmus vytvoří skupinu **P,** do které uloží všechny produkty skupiny **U** seřazené podle počtu hodnocení, průměrného hodnocení produktu a vybere top N produktů.

**Historie prohlížení**

Uživateli bude doručován obsah také na historii prohlížení produktů. Je potřeba si uvědomit, jaká data bude potřeba sbírat.

*Zobrazení stránky* může znamenat, že uživatel má o daný produkt zájem, ale také se může uživatel pouze “proklikávat’ napříč produkty. Více kliků nemusí být pro sběr dat pozitivní.

*Rozšířené kliky* jsou další možností, jak získávat informace o zájmu uživatele. Rozšířené kliky mohou být například „o produktu“, nebo „více informací“ a znamenají, že uživatel má o produkt zájem.

*Uložit na později* je pro uživatele velká pomoc, pokud zrovna nemá o produkt zájem, ale později by si jej mohl zakoupit. Pomocí takového tlačítka se dá vytvořit seznam produktů, na základě kterého uživateli mohou být produkty doručovány. Aby si uživatel vytvořil takový seznam, může být nazván „seznam přání“, nebo „seznam oblíbených produktů“.

*Hodnocení produktu* je další možností, jak uchovávat data o produktu. Uživatel hodnocením produktu dá najevo spokojenost s produktem. Pokud uživatelovo ohodnocení produktu je vysoké, bude mít daný produkt větší váhu při doporučování. Naopak při nízkém hodnocení bude mít daný produkt menší váhu při doporučování.

Aktivní uživatel, který se zaregistroval a zvolil si oblíbené kategorie, začne prohlížet produkty. Systém sbírá data o pohybu uživatele na e-shopu. Které kategorie si prohlíží, jaké produkty z daných kategorií, a zda uživatel prohlížený produkt hodnotil. Systém na získaná data aplikuje mírně upravený algoritmus TF-IDF. Produkty, které vstupují do algoritmu TF-IDF mají koeficient na základě toho, zda byly hodnoceny aktivním uživatelem a tento koeficient následně ovlivňuje výslednou hodnotu. Pokud uživatel produkt nehodnotil, je danému produktu přidělen koeficient 0.5 a jeho ovlivnění výsledné skupiny doporučených produktů je tedy menší o polovinu. Pokud uživatel produkt hodnotil, získá daný produkt koeficient 1 a jeho ovlivnění výsledné skupiny je původní. Výsledná skupina doporučených produktů je seřazena podle podobnosti ke všem prohlíženým produktům uživatele. Princip fungování algoritmu TF-IDF na skupinu prohlížených produktů je následující.

1. Uživatel navštíví produkt, tento produkt se uloží do skupiny **NP**.
   1. Pokud uživatel navštívil produkt poprvé, vytvoří se objekt do skupiny **NP** s informacemi - hodnota 1 reprezentující první návštěvu, datum a čas návštěvy produktu a hodnocení produktu uživatelem, pokud produkt není hodnocen, ukládá se hodnota „nehodnoceno“.
   2. Při každé další návštěvě stejného produktu jsou hodnoty aktualizovány. Přičte se hodnota 1 ke stávající hodnotě návštěvy produktu, aktualizuje se poslední návštěva a upraví se hodnocení produktu, pokud uživatel změnil hodnocení.
2. Po naplnění skupiny **NP** v jakémkoliv množství je skupina seřazena podle několika kritérií. Nejdůležitější kritérium značí počet návštěv produktu, následně poslední návštěva a nakonec jsou produkty seřazeny podle hodnocení produktu uživatelem.
3. Aktivní uživatel zvolí možnost „zobrazit podobné produkty k prohlíženým“.
4. Systém přidělí prohlíženým produktům koeficienty a aplikuje algoritmus TF-IDF.
5. Nejpodobnějších top N produktů, je doručeno aktivnímu uživateli.

**Algoritmus prohlížení - princip fungování algoritmu**

Vstup:

S - vytvořená prázdná session

U - aktivní uživatel

Výstup:

T - seznam top produktů

if (neexistuje produkt v S){

přidat produkt do S

}else{

aktualizovat data produktu v S

}

foreach (produkt v S){

if (U hodnotil produkt){

koeficient produktu = 1

} else{

koeficient produktu = 0.5

}

}

seznam top produktů = výpočet algoritmu TF-IDF na produkty v S

seřazení seznamu top produktů podle podobnosti sestupně

vrátit seznam top produktů

**Ohodnocené produkty**

Zástupcem algoritmů kolaborativního filtrování je SVD (singular-value-decomposition). Tento algoritmus je využit v pozdější fázi, kdy má systém více informací o uživateli, především jeho hodnocení různých produktů. Jakmile má uživatel dostatek ohodnocených produktů, je aplikován algoritmus SVD, který vytvoří skupinu doporučených produktů.

Systém získá z databáze podobné uživatele k aktivnímu uživateli a vytvoří pole **U** (uživatelé) těchto uživatelů spolu s aktivním uživatelem. Následně vytvoří pole **P** (produkty), které obsahuje všechny produkty, které uživatelé z pole **U** hodnotili. Systém vytvoří user-item matici, kde řádky představují uživatele z pole **U**, a sloupce představují produkty z pole **P**.Následně na každou pozici uživatel-produkt vloží hodnocení produktu daného uživatele. Pokud uživatel produkt nehodnotil, je na danou pozici doplněna hodnota 0. Algoritmus SVD následně provede dekompozici této matice na matice U, sigma, Vt. Na základě těchto matic vypočítá podobnost mezi jednotlivými produkty pro daného uživatele a vytvoří skupinu doporučených produktů.

1. Uživatel má ohodnocených několik produktů.
2. Systém vytvoří user-item matici doplněnou o hodnocení produktů jednotlivých uživatelů. Na nevyplněná data je doplněna hodnota 0.
3. Systém aplikuje algoritmus SVD a vypočítá podobnost mezi jednotlivými produkty aktivního uživatele.
4. Z nalezených produktů jsou odebrány produkty, již hodnocené aktivním uživatelem.
5. Doručí aktivnímu uživateli podobné produkty.

V následující tabulce je vybráno 16 produktů, které aktivní uživatel hodnotil. Z tabulky můžeme vidět uživatelův zájem o kategorie alkohol, maso, nápoje, ovoce a mléčné výrobky. Naopak aktivní uživatel nejeví zájem o produkty z kategorie koření a zelenina.

Tabulka 1: Hodnocené produkty aktivního uživatele

|  |  |  |  |
| --- | --- | --- | --- |
| **Produkt** | **Hodnocení** | **Produkt** | **Hodnocení** |
| Absinth | 4 | Fanta | 5 |
| Božkov Republica | 5 | Lipánek | 4 |
| Bio hovězí | 4 | Mléko trvalivé | 3 |
| Hovězí mleté | 4 | Paprika červená | 2 |
| Kari | 1 | Řepa | 3 |
| Paprika sladká | 1 | Jahody | 4 |
| Kofola | 3 | Švestky | 5 |

Výsledné hodnoty po aplikaci algoritmu SVD a odebrání produktů, které již uživatel hodnotil, jsou znázorněny v následující tabulce.

Tabulka 2: Podobné produkty k hodnoceným produktům aktivního uživatele

|  |  |  |  |
| --- | --- | --- | --- |
| **Produkt** | **Podobnost** | **Produkt** | **Podobnost** |
| Mandarinka | 0,8247 | Mango | 0,6428 |
| Pepsi cola | 0,7657 | Vepřová panenka | 0,6314 |
| Mléko plnotučné | 0,7483 | Kiwi | 0,6137 |
| Sýr gouda | 0,7095 | Kuřecí čtvrtky | 0,5827 |
| Energy drink Tiger | 0,6952 | Jack Daniels Honey | 0,5798 |
| Beefeter Gin Pink | 0,6655 | Koláč tvarohový | 0,5678 |
| Hovězí burger | 0,6573 | Mléko čerstvé | 0,5669 |

V tabulce výše je 14 nejpodobnějších produktů k hodnoceným produktům aktivního uživatele. Objevují se zde produkty z kategorií ovoce, nápoje, alkohol, maso a mléčné produkty. Produkty z těchto kategorií hodnotil aktivní uživatel vyšším hodnocením.

**Algoritmus hodnocení - sestavení user-item matice**

Vstup:

U - seznam uživatelů z databáze

P - seznam produktů z databáze

H - seznam hodnocení uživatel/produkt

Výstup:

UPH - uživatel-produkt matice s hodnocením

//vytvoření user-item matice a doplnění dat

foreach (Ui) ze seznamu U{

foreach (Pj) ze seznamu P{

if (existuje hodnocení H[i][j]){

matice[i][j] = H[i][j] //uloží hodnocení uživatel/produkt do matice

} else{

matice[i][j] = 0 //uloží hodnotu 0 do matice pro vyplnění dat

}

}

}

return matice uživatel-produkt s hodnocením

**Historie nákupů**

Při doručování produktů bude brána v potaz historie nákupů. Uživatel, který si zakoupil produkty s jistými parametry (kategorie), by mohl mít zájem o produkty s podobnými parametry k daným produktům. Uživateli bude doručen podobný produkt na základě content-based algoritmu. Podle výpočtu podobnosti produktů pomocí TF-IDF vektoru.

*Produkty* budou nejdůležitější složkou pro výpočet doručení podobných produktů. Bude důležité, jaké produkty si uživatel zakoupil a také jak často, nebo v jakém množství. Pokud například uživatel nakupuje mléčné výrobky, především jogurty, budou mu doručeny podobné produkty s těmito parametry.

*Uskutečnění nákupu* bude určovat váhu zakoupených produktů pro výpočet. Pokud si uživatel daný produkt nezakoupil za posledních pár nákupů, může to znamenat, že o daný produkt nemá zájem. Takový produkt s danými parametry tedy nebude brán do hodnocení pro podobnost produktů.

Historie nákupu využívá pro doporučení modifikovaný TF-IDF algoritmus. Aktivní uživatel provede nákup několika produktů v daný okamžik. Systém zaznamenává, zdali uživatel nakoupené produkty již hodnotil, či nikoliv. Na základě hodnocení těchto produktů je každému ze zakoupených produktů přidělen koeficient, který ovlivní výpočet algoritmu TF-IDF. Pokud uživatel produkt nehodnotil, je mu automaticky přidělen koeficient 0,5. Ostatní produkty získají koeficient podle jejich hodnocení uživatelem.

* Hodnocení produktu 5\* - získaný koeficient 1.
* Hodnocení produktu 4\* - získaný koeficient 0,8.
* Hodnocení produktu 3\* - získaný koeficient 0,6.
* Hodnocení produktu 2\* - získaný koeficient 0,4.
* Hodnocení produktu 1\* - získaný koeficient 0,2.
* Nehodnocený produkt - získaný koeficient 0,5.

**Algoritmus historie nákupů - princip fungování algoritmu**

Vstup:

P - seznam zakoupených produktů aktivního uživatele

U - aktivní uživatel

H - hodnocení produktu (Pi)

Výstup:

T - seznam top produktů, seřazených podle podobnosti

foreach (Pi){

if (U hodnotil Pi){

switch (H){

case 5:

koeficient produktu = 1

case 4:

koeficient produktu = 0.8

case 3:

koeficient produktu = 0.6

case 2:

koeficient produktu = 0.4

case 1:

koeficient produktu = 0.2

}

} else{

koeficient produktu = 0.5

}

}

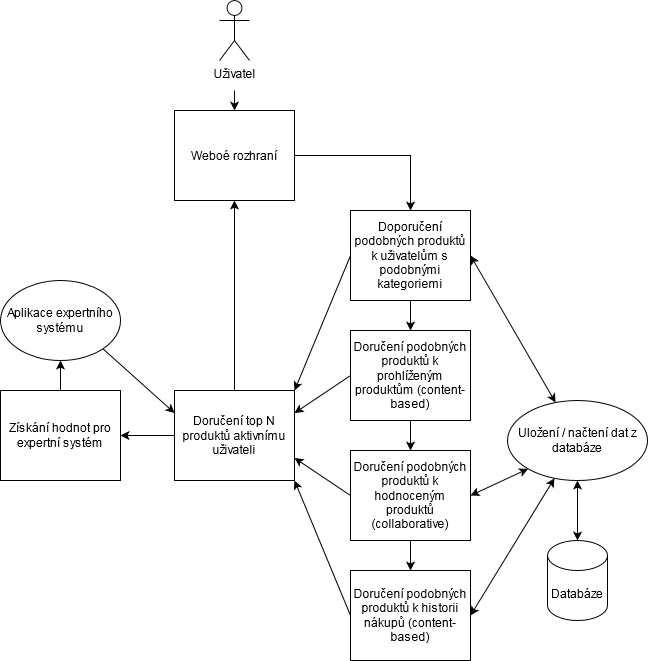
seznam top produktů = výpočet algoritmu TF-IDF na produkty s koeficienty

seřazení seznamu top produktů podle podobnosti sestupně

return seznam top produktů

* 1. Přístup pro doporučení produktů

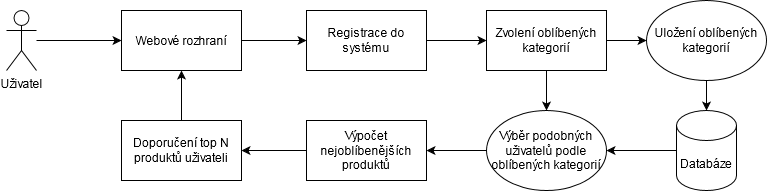
Na základě hodnot získaných z výše uvedených přístupů, bude uživateli doručován ideální obsah produktů. Na následujícím obrázku je zobrazen návrh fungování systému.



Obrázek 11: Návrh systému (zdroj: vlastní)

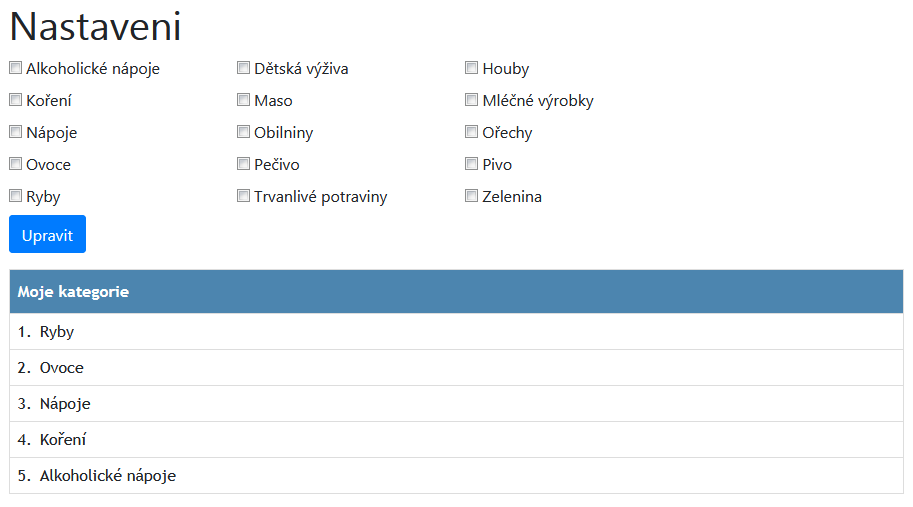
Navržený systém je automatizovaný tak, aby byly uživateli doručovány produkty postupně podle jeho interakce se systémem. Každá část navrhovaného systému bude dále popsána detailněji. Aby uživateli byly doručeny produkty, není nutné, aby prošel všechny části. Systém zaznamenává chování uživatele a doručuje systémy postupně. Čím více uživatel interaguje se systémem, tím více jsou personalizované produkty aktivnímu uživateli doručovány.

**První část systému - oblíbené kategorie**



Obrázek 12: Diagram doporučení podle oblíbených kategorií (zdroj: vlastní)

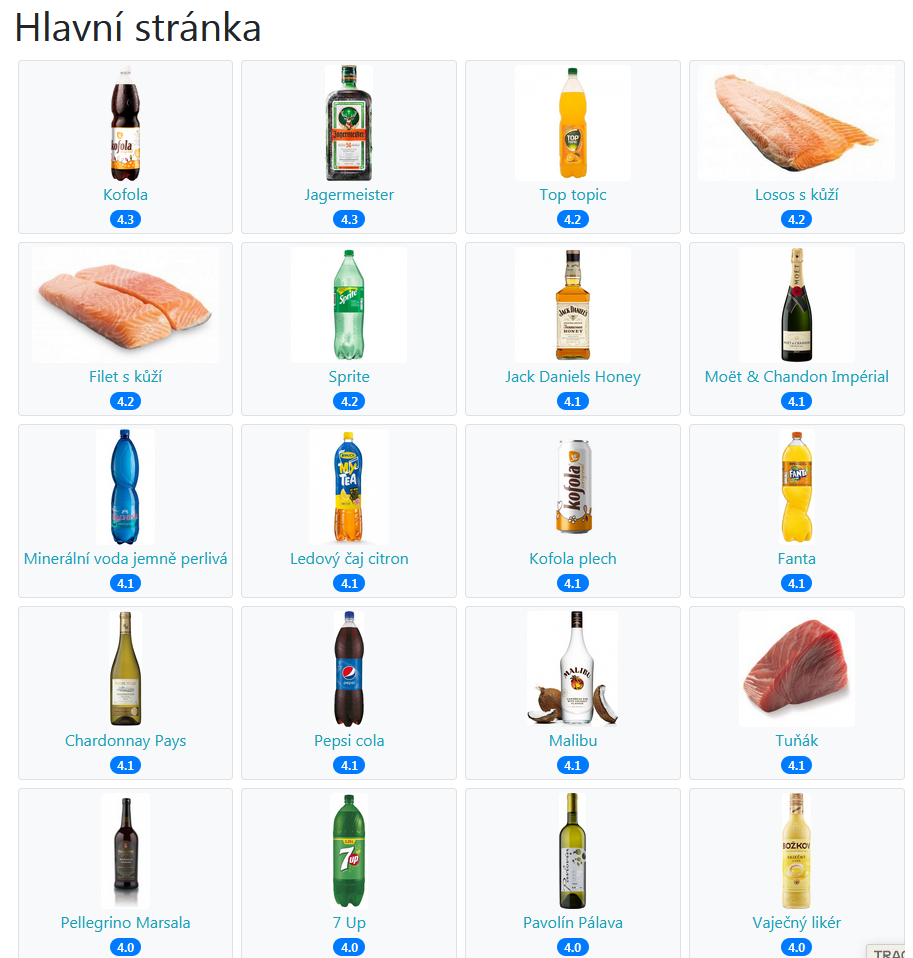
Systém po registraci vyzve aktivního uživatele ke zvolení oblíbených kategorií, na základě kterých vybere podobné uživatele a všechny jejich hodnocené produkty. Systém vybrané produkty seřadí podle počtu hodnocení, průměrného hodnocení a doručí uživateli první skupinu navržených produktů, které se zobrazí na hlavní stránce aplikace. Níže je uveden příklad prvního kontaktu se systémem.



Obrázek 13: Nastavení oblíbených kategorií (zdroj: testovací e-shop)

Na obrázku výše můžeme vidět úvodní nastavení oblíbených kategorií. Jsou zde všechny kategorie s tlačítkem „Upravit“ pro uložení zvolených kategorií a sekci „Moje kategorie“, která zobrazuje zvolené oblíbené kategorie uživatele. Aktivní uživatel si zde zvolil kategorie Ryby, Ovoce, Nápoje, Koření a Alkoholické nápoje. První doporučené produkty budou obsahovat nejvíce produkty z těchto kategorií, které hodnotili podobní uživatelé k aktivnímu uživateli.

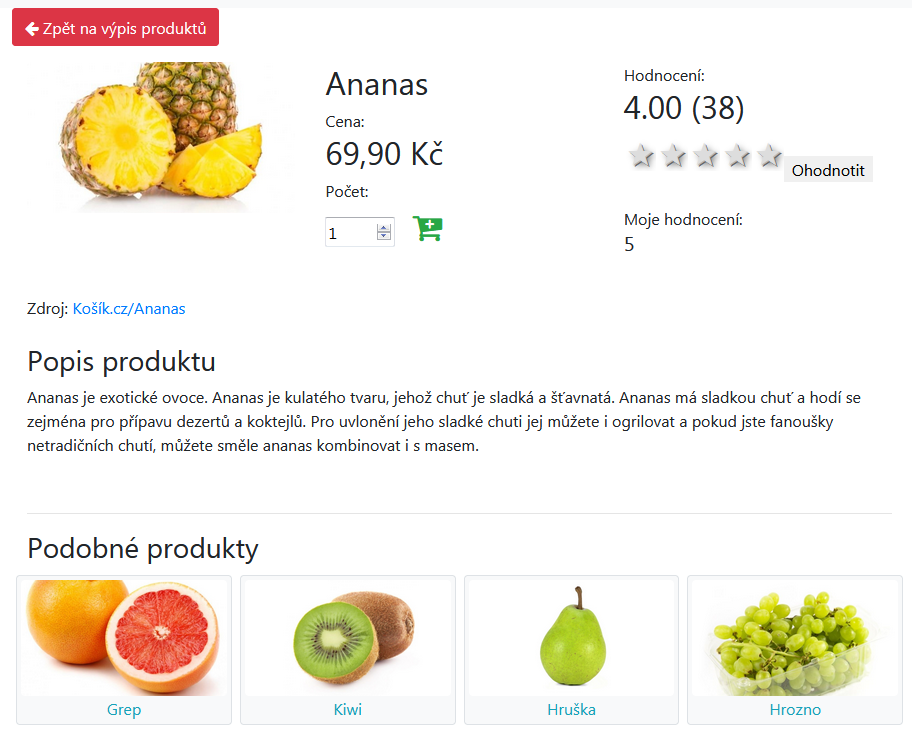
Na následujícím obrázku je doručeno 20 nejlépe hodnocených produktů, hodnocených podobnými uživateli k aktivnímu uživateli. Můžeme vidět, že těchto top produktů zahrnuje oblíbené kategorie aktivního uživatele.



Obrázek 14: Hlavní stránka s doporučenými produkty (zdroj: testovací e-shop)

**Druhá část systému - prohlížení**

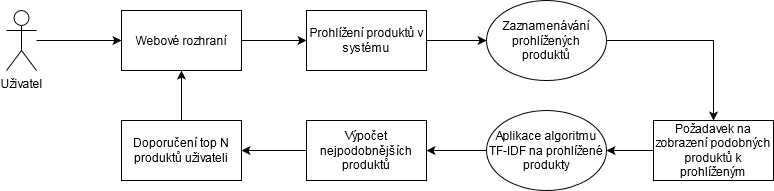
Dalším aspektem automatizovaného doručování je chování uživatele na webu. Jak je zvykem na různých e-shopech, při prohlížení produktů jsou doporučovány podobné produkty. Systém doporučuje uživateli nejpodobnější produkty k prohlíženému produktu. Také však zaznamenává historii prohlížených produktů, a také to, zdali dané produkty uživatel hodnotil. Detail produktu s podobnými produkty můžeme vidět na následujícím obrázku.



Obrázek 15: Detail produktu (zdroj: testovací e-shop)

V detailu produktu můžeme vidět základní informace o produktu (obrázek, název, cena), možnost zakoupení produktu, dále průměrné hodnocení zobrazeného produktu, celkový počet hodnocení produktu (je znázorněn v závorce), a možnost produkt ohodnotit s zobrazením hodnocení aktivního uživatele. Následuje zdroj informací o daném produktu (všechny produkty jsou získány z reálného e-shopu Košík.cz) a nakonec popis produktu. Ten je důležitý pro doporučování podobných produktů pomocí algoritmu TF-IDF. Nejpodobnější produkty k zobrazenému produktu jsou znázorněny v sekci „Podobné produkty“ pod zobrazeným produktem.

Jak bylo zmíněno výše, systém zaznamenává chování uživatele na e-shopu a ukládá historii prohlížení produktů. Aktivní uživatel podá požadavek pro zobrazení podobných produktů k prohlíženým, systém zpracuje požadavek, na prohlížené produkty aplikuje upravený algoritmus TF-IDF, vypočítá nejpodobnější produkty a doručí top N produktů aktivnímu uživateli. Modifikace je popsána v kapitole 3.2 Algoritmy použité v systému v části s podtitulem **Historie prohlížení**.



Obrázek 16: Diagram doporučení na základě prohlížení

Prohlížené produkty a informace o nich jsou znázorněny v následující tabulce.



Obrázek 17: Příklad prohlížených produktů (zdroj: testovací e-shop)

V tabulce výše můžeme vidět produkty, které si uživatel prohlédl. Skupina prohlížených produktů obsahuje 8 produktů z kategorií „Ovoce“ a „Zelenina“, dále tabulka obsahuje informace o počtu návštěv daného produktu, kdy naposledy uživatel produkt navštívil a zda produkt hodnotil.

Výsledná tabulka podobných produktů k prohlíženým produktům vypadá následovně.

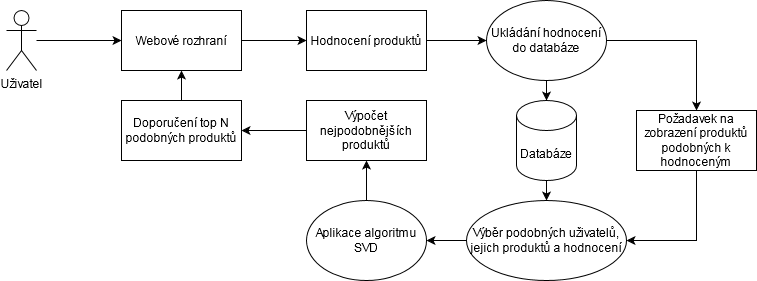


Obrázek 18: Výsledky doporučení na základě prohlížení (zdroj: testovací e-shop)

V tabulce výše můžeme vidět prvních 20 produktů podobných k prohlíženým produktům uživatele. Jak bylo zmíněno výše, produkty, které uživatel nehodnotil, mají o polovinu menší vliv na výsledná data. Uživatel z prohlížených produktů ohodnotil pouze dva z jedné skupiny. „Jahody“ a „Ananas“ - produkty které uživatel jako jediné hodnotil, nemají snížený vliv na doporučení produktů. Z toho důvodu jsou na prvních příčkách produkty z kategorie Ovoce. Výsledná tabulka výše je jedna ze dvou částí, která bude použita pro následné přístupy doporučování produktů.

**Třetí část systému - hodnocení produktů**

Další částí navrhovaného systému je doporučení podle podobnosti již hodnocených produktů aktivního uživatele. Uživatel, který si prohlíží produkty, může tyto produkty také hodnotit v detailu produktu, jak bylo zmíněno výše. Hodnocení jednotlivých produktů je důležité pro další část doporučení, která je znázorněna na následujícím diagramu.



Obrázek 19: Diagram doporučení podle hodnocení produktů (zdroj: vlastní)

Jednotlivá hodnocení se ukládají do databáze. Pokud uživatel provede změnu hodnocení, tak nové hodnocení přepíše původní hodnotu v databázi. Uživatel následně podá požadavek na zobrazení podobných produktů. Systém vybere všechna data z databáze, vytvoří user-item matici a provede algoritmus SVD, který je popsán v kapitole 3.2 Algoritmy použité v systému v části s podtitulem **Hodnocení produktů.** Následně systém vybere top N nejvhodnějších produktů, které doručí aktivnímu uživateli. Na následujícím obrázku je hodnocení aktivního uživatele.



Obrázek 20: Příklad hodnocených produktů (zdroj: testovací e-shop)

Aktivní uživatel zde hodnotil 14 produktů z různých kategorií. Vysokým hodnocením uživatel ohodnotil kategorie Obilniny, Alkoholické nápoje a Pečivo. Na druhou stranu nízkým hodnocením uživatel ohodnotil kategorie Koření a Ryby. Systém po provedení algoritmu SVD by měl doporučit produkty podobné těm vysoce hodnoceným. Naopak produkty s nízkým hodnocením by systém doporučit neměl. Na následujícím obrázku jsou doporučené produkty pomocí algoritmu SVD, aplikovaného na již hodnocené produkty aktivního uživatele.

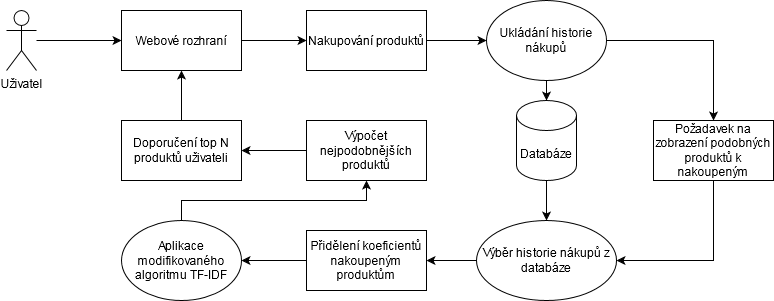


Obrázek 21: Výsledky doporučení podle hodnocených produktů (zdroj: testovací e-shop)

Na obrázku je zobrazeno prvních 20 nejpodobnějších produktů k produktům aktivního uživatele. Můžeme vidět, že všechny produkty mají vysoké hodnocení a výrazná většina produktů je z kategorií, které uživatel hodnotil vysokým hodnocením. Produkty z kategorie Ryby a Koření, které uživatel hodnotil nízkým hodnocením, ve výsledném seznamu nejsou. Výsledná tabulka znázorněna výše je druhým vstupním parametrem pro následné přístupy doporučování produktů.

**Čtvrtá část systému - provedení nákupů**

Poslední částí navrhovaného systému je doporučení produktů podle historie nákupů aktivního uživatele. Systém zaznamenává a ukládá historii nákupů, podle které následně uživateli doporučuje vhodné produkty.



Obrázek 22: Diagram doporučení podle historie nákupů (zdroj: vlastní)

Uživatel provede nákup několika produktů, systém tento nákup zaznamená do databáze. Jakmile je podán požadavek na zobrazení podobných produktů, systém získá data o provedených nákupech daného uživatele a přidělí jim koeficienty podle data zakoupení a hodnocení uživatele. Tento postup je popsán výše. Na vybrané produkty s koeficienty je aplikován modifikovaný algoritmus TF-IDF, jež provede výpočet nejpodobnějších produktů a doručí top N produktů aktivnímu uživateli. Tato část systému je využita pro přístup 3 a 4.

* + 1. Přístupy využívající části systému

Tato část práce se bude zabývat jednotlivými přístupy, které doporučují produkty aktivnímu uživateli podle jeho chování v systému. Jednotlivé přístupy jsou kombinovaný hybridní systém, propojující výsledky dvou algoritmů do jednoho výstupu. Každý z přístupů disponuje jinými pravidly pro doručení nejvhodnějších produktů a je rozdělen do několika skupin, které využívají tato pravidla odlišně.

**Přístup 1 - procentuální podíl**

První z přístupů využívá k vytvoření top doporučených produktů procentuální podíl z druhé a třetí části navrhovaného systému. Tedy kombinace výstupů algoritmů SVD (podobné k hodnoceným) a TF-IDF (podobné k prohlíženým). Výsledný seznam je tvořen maximálně 40 produkty a jsou vybrány vždy top produkty daných výstupů seřazené podle podobnosti. Každá skupina představuje jiný poměr podílu na jednotlivé výstupy a z nich vždy odpovídající počet produktů. Pokud vezmeme poměr 30% / 70%, výsledný seznam bude obsahovat 30% nejvhodnějších produktů z výsledků podobných prohlížení, tedy aplikování algoritmu TF-IDF na prohlížené produkty, a 70% nejvhodnějších produktů z výsledků podobných již hodnocených produktů, tedy aplikování algoritmu SVD na již hodnocené produkty aktivním uživatelem. Podíl jednotlivých skupin je následující.

* Skupina A: Podobné prohlížení 30% / 70% podobné k hodnoceným.
* Skupina B: Podobné prohlížení 50% / 50% podobné k hodnoceným.
* Skupina C: Podobné prohlížení 70% / 30% podobné k hodnoceným.

**Algoritmus přístupu 1 a Přístupu 2**

Vstup:

A - výstup podobných produktů k prohlíženým (algoritmus TF-IDF)

B - výstup podobných produktů k hodnoceným (algoritmus SVD)

P - získána procenta podle skupiny

Výstup:

VS - výsledný seznam top produktů seřazených podle podobnosti sestupně

//počet produktů pro výslednou skupinu

početA = 40 \* (P/100)

početB = 40 \* (1-P/100)

vybranéA = vybrání prvních početA produktů z A

vybranéB = vybrání prvních početB produktů z B

//VS = výsledný seznam

VS = vybranéA + vybranéB

seřazení VS podle podobnosti sestupně

return VS

**Přístup 2 - procentuální podíl hodnocení produktů**

Druhý přístup používá k vytvoření top doporučených produktů podíl na základě hodnocených produktů. Podle počtu již hodnocených produktů je vytvořena skupina doporučených produktů. Pokud uživatel hodnotil málo produktů, tak systém neposbíral dostatečné množství dat a uživateli doporučí malé množství produktů z výsledků podobných k hodnoceným produktům. Doporučí uživateli větší množství produktů z výsledků podobných k prohlíženým. Například pokud uživatel hodnotil alespoň 5 produktů, ale ne více než 11, je doporučeno pouze 10% produktů ze skupiny podobných k hodnoceným produktům. Podíl jednotlivých skupin je následující.

* Skupina A: Alespoň 5 produktů ohodnocených - podobné prohlíženým 90% / 10% podobné k hodnoceným.
* Skupina B: 11 - 20 produktů ohodnocených - podobné prohlíženým 70% / 30% podobné k hodnoceným.
* Skupina C: 21 - 30 produktů ohodnocených - podobné prohlíženým 50% / 50% podobné k hodnoceným.
* Skupina D: 31 a více produktů ohodnocených - podobné prohlíženým 30% / 70% podobné k hodnoceným.

**Přístup 3 - podobnost k historii nákupů**

Třetí přístup používá k vytvoření top doporučených produktů čtvrtou část navrhovaného systému, a to historii nákupů aktivního uživatele. Tento přístup přiděluje produktům koeficient podle toho, kdy byly zakoupeny. Například produkty zakoupeny aktuální den získávají nejvyšší koeficient, naopak produkty zakoupené před 3 měsícem získávají nejnižší koeficient.

* Produkty zakoupené dnes - koeficient 1.
* Produkty zakoupené do 7 dní - koeficient 0,7.
* Produkty zakoupené do 30 dní - koeficient 0,5.
* Produkty zakoupené do 90 dní - koeficient 0,2.
* Produkty zakoupené déle jak 90 dní - nemají vliv na výsledný výpočet.

**Algoritmus přístupu 3**

Vstup:

P - seznam zakoupených produktů zakoupených uživatelem

Výstup:

T - seznam top produktů

foreach (Pi) ze skupiny P{

if (datum zakoupení (Pi) <= 24h){

koeficient produktu = 1

} elseif (datum zakoupení (Pi) <= 7 dní){

koeficient produktu = 0.7

} elseif (datum zakoupení (Pi) <= 30 dní){

koeficient produktu = 0.5

} elseif (datum zakoupení (Pi) <= 90 dní){

koeficient produktu = 0.2

} elseif (datum zakoupení (Pi) > 90 dní){

koeficient produktu = 0

}

}

seznam top produktů = výpočet algoritmu TF-IDF na produkty s koeficienty

seřazení seznamu top produktů podle podobnosti sestupně

return seznam top produktů

**Přístup 4 - upravená podobnost k historii nákupů**

Čtvrtý přístup je založen na přístupu třetím. Využívá přidělování koeficientu podle času zakoupení, nicméně tento přístup je rozšířen o další koeficient a zvýšení hranice zobrazování relevantních produktů. Přístup má dvě skupiny, kde každá přičítá koeficient podle jiných parametrů. Tato modifikace má zajistit zobrazování menšího počtu nerelevantních produktů.

* Skupina A - produkty získávají navíc koeficient na základě počtu produktů v historii.
* Skupina B - produkty získávají koeficient podle kategorií.

Ve skupině A získávají produkty koeficient podle počtu zakoupených produktů. Pokud při prvním nákupu uživatel zakoupí 5 produktů, je těmto produktům přidělen koeficient 0,125 (výpočet: počet produktů v historii / maximálním počtem / 2, tedy 5 / 20 = 0,250 / 2 = 0,125).

Ve skupině B získávají produkty navíc koeficient, pokud jsou ze stejné kategorie, jako produkty, které uživatel zakoupil. Pokud uživatel zakoupil několik produktů z kategorie Pečivo, tak doporučené produkty z kategorie Pečivo získávají navíc koeficient 0,5.

Například aktivní uživatel provede dva nákupy v rozmezí jednoho týdne. První nákup proveden 13. 4. 2021 se skládá z 8 produktů ve dvou kategoriích.

1. Ovoce: ananas, grep, bluma, banán
2. Zelenina: rajče, paprika červená, řepa, paprika žlutá

Druhý nákup provede 19. 4. 2021, který obsahuje 7 produktů opět ve dvou kategoriích.

1. Pečivo: kaiserka cereální, makový koláč, pletenec, kobliha
2. Trvanlivé potraviny: červená čočka, rýže basmati, hrách

Výstup přístupu 4 ve skupině B ke dni 19. 4. 2021 pro aktivního uživatele bude seznam doporučených produktů vypadat tak, jak je znázorněn v tabulce č. 3.

V tabulce můžeme vidět, že na prvních příčkách jsou produkty podobné k nákupu provedenému tentýž den. Koeficient těchto produktů je 1. K této hodnotě je připočten koeficient skupiny B, tedy hodnota 0,5 pokud kategorie produktu odpovídá kategoriím zakoupeným. Výsledný koeficient je pak součet těchto dvou hodnot.

Tabulka 3: Příklad výstupu přístupu 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Produkt** | **Podobnost** | **Koeficient** | **Koeficient skupiny** | **Celkový koeficient** |
| Kaiserka natural | 1 | 1 | 0,5 | 1,5 |
| Rýže dlouhozrnná | 0,963 | 1 | 0,5 | 1,5 |
| Hrách žlutý | 0,909 | 1 | 0,5 | 1,5 |
| Koláč tvarohový | 0,892 | 1 | 0,5 | 1,5 |
| Bulka slunečnicová | 0,890 | 1 | 0,5 | 1,5 |
| Preclík | 0,880 | 1 | 0,5 | 1,5 |
| Rýže parboiled | 0,828 | 1 | 0,5 | 1,5 |
| Houska | 0,792 | 1 | 0,5 | 1,5 |
| Paprika bílá | 0,790 | 0,7 | 0,5 | 1,2 |
| Hruška | 0,775 | 0,7 | 0,5 | 1,2 |

**Algoritmus přístupu 4**

Skupina A:

Vstup:

V - výstup algoritmu Přístupu 3

P - počet produktů v historii

M - maximální počet produktů = 20

Výstup

T - seznam top produktů s upravenými koeficienty

if (P > M){

P = M

koeficient = P / M / 2

} else{

koeficient = P / M / 2

}

foreach (Vi) ze seznamu V proveď{

podobnost (Vi) += koeficient

}

return seznam V

Skupina B:

Vstup:

V - výstup algoritmu Přístupu 3

K - oblíbené kategorie aktivního uživatele

Výstup

T - seznam top produktů s upravenými koeficienty

foreach (Vi) ze seznamu V proveď{

if (kategorie (Vi) je v seznamu K){

podobnost (Vi) += 0.5

}

}

T = seřazení seznamu V podle podobnosti

return T

**Expertní systém**

Poslední částí systému, která finálně seřadí seznam produktů zobrazený pro uživatele, bude expertní systém. Data získaná z výše popsaných částí systému budou vložena do expertního systému a z nich budou vybrány nejvhodnější produkty. Expertní systém obsahuje 3 vstupní proměnné a 1 výstupní proměnnou. Vstupní proměnné zahrnují výsledky částí navrhovaného systému. První vstupní parametr je získaná podobnost daného produktu pomocí algoritmu SVD. Dalším vstupním parametrem je koeficient získán z výsledků historie nákupů. Poslední parametr zahrnuje průměrné hodnocení produktu. Následně, pokud kategorie produktu odpovídá oblíbené kategorii aktivního uživatele, je k výslednému ranku připočten koeficient 0,2. Pomocí IF-THEN pravidel bude vypočítaná výstupní proměnná. Výstupní proměnná bude nejvhodnější doručení uživateli.

Expertní systém je počítačový program, který simuluje rozhodovací činnost od expertů při řešení složitých úloh. Jedná se o systém, který nevyužívá znalostí vlastně nabytých, ale využívá znalosti expertů v dané oblasti. Expertní systém umožňuje využívat znalosti expertů, bez lidských omylů. [5]

Expertní systém využívá pro svou reprezentaci jazykové proměnné spolu s bází takzvaných IF-THEN pravidel. Tyto pravidla, která zadá do systému expert, pomáhají systému určit výslednou hodnotu. Jazykové proměnné využívají pro interpretaci fuzzy množiny, které jsou dané v určitém intervalu, například 0 až 100 <0, 100> a jednotlivé prvky představují určitou hodnotu fuzzy množiny.[5]

Jak bylo zmíněno výše, expertní systém využívá tři vstupní proměnné, které následně vytvářejí bázi IF-THEN pravidel. Tyto pravidla jsou znázorněny v následující tabulce.

Tabulka 4: Vstupní proměnné expertního systému

|  |  |
| --- | --- |
| **Název** | **Interval** |
| Podobnost | <0, 1> |
| Koeficient | <0, 1.5> |
| Průměrné hodnocení | <0, 5> |

Jednotlivé vstupní proměnné představují vstupy z různých částí doporučovacího systému.

**Podobnost** určuje jak moc je zvolený produkt podobný ostatním produktům již ohodnocených aktivním uživatelem. Tato podobnost je vypočítána podle algoritmu SVD, který je popsán v kapitole 3.2 v podtitulu **hodnocené produkty**.

**Koeficient** je hodnota, která je získána z doporučených produktů podle nákupů aktivního uživatele. Koeficient určuje význam produktu v doporučeném seznamu produktů podle historie nákupů uživatele. Získání koeficientu je popsáno výše v této podkapitole v části zabývající se **Přístupu 3 - podobnost k historii nákupů**.

**Průměrné hodnocení** produktu je získáno jako podíl součtu všech hodnocení daného produktů a počtu hodnocení uživateli.

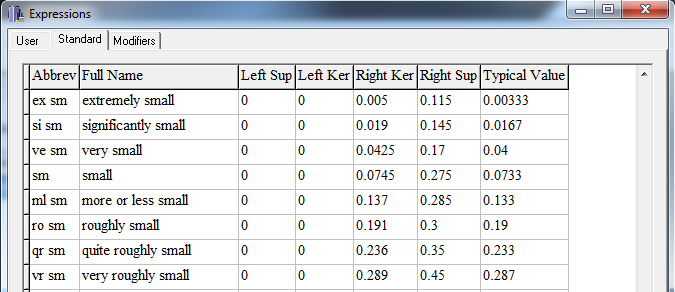
V následující tabulce je znázorněno rozložení fuzzy množin jednotlivých vstupních proměnných.

Tabulka 5: Rozložení fuzzy množin vstupních proměnných expertního systému

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Název** | **Hodnota** | **Left Sup** | **Left Ker** | **Right Ker** | **Right Sup** |
| Podobnost | Ve\_low\_sim | 0 | 0 | - | 0,25 |
| Low\_sim | 0 | 0,25 | - | 0,5 |
| Med\_sim | 0,25 | 0,5 | - | 0,75 |
| High\_sim | 0,5 | 0,75 | - | 1 |
| Ve\_high\_sim | 0,75 | 1 | - | 1 |
| Koeficient | Low | 0 | 0 | 0,225 | 0,445 |
| Med | 0,225 | 0,45 | 0,675 | 0,9 |
| High | 0,675 | 0,9 | 1,05 | 1,27 |
| Ve\_high | 1,05 | 1,27 | 1,5 | 1,5 |
| Průměrné hodnocení | Low | 0 | 0 | 1 | 2 |
| Med | 1 | 2 | 3 | 4 |
| High | 3 | 4 | 5 | 5 |

V tabulce výše jsou vidět všechny vstupní proměnné a rozdělení jejich fuzzy množin. Každá hodnota vstupní proměnné je dána levým (Left Sup) a pravým (Right Sup) ohraničením nosiče a levým (Left Ker) a pravým (Right Ker) ohraničením jádra. Levé a pravé ohraničení udává začátek a konec každého nosiče fuzzy množiny, prvky v tomto ohraničení jsou nenulové. Levé a pravé ohraničení jádra označuje začátek a konec maximálního stupně příslušnosti, prvky v tomto ohraničení do příslušné hodnoty patří. Vstupní proměnné, jež nemají znázorněn pravý konec jádra, jsou trojúhelníkového tvaru a jejich maximální stupeň příslušnosti je dán právě v bodě levého ohraničení jádra (Left Ker).

Výstupní proměnnou expertního systému je standartní proměnná s využitím modifikátorů. Využité hodnoty mají rozsah extremely small až extremely big. Všechny využité hodnoty jsou small, medium a big na tyto hodnoty jsou aplikovány modifikátory extremely, significantly, very, more or less, roughly, quite roughly a very roughly. Výstupní proměnná udává míru důležitosti daného produktu pro výsledné doručení uživateli. Rozsah výstupní proměnné je v intervalu <0, 1>. Ukázka výstupní proměnné a její hodnoty jsou znázorněny na následujícím obrázku.



Obrázek 23: Výstupní proměnná expertního systému

Báze znalostí v navrženém expertním systému tvoří 60 pravidel. Tato pravidla jsou tvořena kombinací vstupních proměnných, kde každé kombinaci je přidělena výstupní proměnná. Obecné znázornění pravidel vypadá následovně.

Výběr několika pravidel expertního systému je znázorněno v následující tabulce.

Tabulka 6: Seznam několika pravidel expertního systému

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Číslo pravidla** | **Podobnost** | **Koeficient** | **Průměrné hodnocení** | **Rank** |
| 1 | ve\_low\_sim | low | low | ex sm |
| 7 | low\_sim | high | low | sm |
| 26 | low\_sim | med | med | qr sm |
| 32 | med\_sim | ve\_high | med | qr bi |
| 48 | high\_sim | med | high | vr bi |
| 60 | ve\_high\_sim | ve\_hig | high | ex bi |

Každý produkt získá z expertního systému výsledný rank, tato hodnota bude připočtena k výsledkům navrženého doporučovacího systému.

Jako inferenční metoda pro navrhovaný expertní systém byla zvolena metoda CNF (conjuctive normal form) neboli fuzzy aproximace s konjunkcemi. Tato metoda byla zvolena z toho důvodu nejpoužívanější metody v expertních systémech. Také v rámci testování různých inferenčních metod s deffuzifikačními metodami měla tato inferenční metoda nejmenší rozdíly ve výsledných hodnotách.

Jako deffuzifikační metoda byla zvolena metoda MCOG (modified center of gravity) neboli metoda upraveného těžiště. Při testování byly brány v potaz metody COG a MCOG, zvolená metoda však lépe odpovídala očekávaným výsledkům.

Několik testovacích výstupů expertního systému je znázorněno v následující tabulce. Byly testovány různé kombinace pro daný produkt. Výsledky testu produktu 1 a produktu 5 vyšly stejně, přičemž produkt 1 má průměrné hodnoty podobnosti a průměrného hodnocení a koeficient produktu je 1,3, tedy vysoký. Produkt 5 má nízkou podobnost k ostatním produktům, ale vysoké průměrné hodnocení a vysoký koeficient produktu, stejně tak jako produkt 1. Naproti tomu produkty 6 a 7 získaly velmi nízkou hodnotu proměnné rank, i když mají oba produkty průměrnou podobnost, tak koeficient produktů je velice nízký a průměrné hodnocení produktu je nižší až nízký.

Tabulka 7: Testování expertního systému

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Podobnost** | **Koeficient** | **Průměrné hodnocení** | **Rank** |
| 1 | 0,5 | 1,3 | 3,1 | 0,75 |
| 2 | 0,7 | 0,2 | 4 | 0,51 |
| 3 | 0 | 1,5 | 4,2 | 0,5 |
| 4 | 0,8 | 1 | 4,5 | 0,9 |
| 5 | 0,2 | 1,2 | 4,7 | 0,75 |
| 6 | 0,5 | 0 | 3 | 0,25 |
| 7 | 0,5 | 0,2 | 1,6 | 0,23 |

Průměrnou výslednou proměnnou rank disponují testované produkty 2 a 3, oba mají relativně vysoké průměrné hodnocení, produkt 2 má vysokou podobnost k jiným produktům, ale nízký koeficient a přesně naopak to má produkt 3, kdy jeho koeficient je nejvyšší, ale podobnost k jiným produktům je nulová. Nejlepší výsledek proměnné rank má testovaný produkt 4, který má vysokou podobnost k ostatním produktům, vysoké průměrné hodnocení a vyšší koeficient.

Hodnota doporučení se získá jako součin výsledné proměnné rank každého produktu a podobnosti získané doporučovacím systémem, ke které je přičtena hodnota 1. Produkty se podle hodnoty doporučení seřadí a doručí se uživateli top N nejvhodnějších produktů.

Následující tabulka představuje prvních 10 doporučených produktů, na které byl aplikován expertní systém. Pokud kategorie produktu odpovídá oblíbeným kategoriím aktivního uživatele, je k tomuto produktu přičtena hodnota 0,2.

Tabulka 8: Výpočet hodnoty doporučení

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Produkt** | **Podobnost produktu** | **Koeficient produktu** | **Hodnocení produktu** | **Oblíbená kategorie** | **Rank** | **Hodnota doporučení** |
| Šnek sýrový | 0,968 | 1,5 | 4 | 0,2 | 0,960 | 1,889 |
| Quinoa | 0,971 | 0 | 4 | 0,2 | 0,870 | 1,714 |
| Makový koláč | 0,793 | 1,5 | 3,9 | 0,2 | 0,940 | 1,685 |
| Pletenec | 0,669 | 1,5 | 3,7 | 0,2 | 0,930 | 1,552 |
| Kaiserka natural | 0,540 | 1,5 | 3,8 | 0,2 | 0,890 | 1,370 |
| Bageta | 0,468 | 1,5 | 4,2 | 0,2 | 0,900 | 1,321 |
| Chleba | 0,504 | 1,5 | 3,6 | 0,2 | 0,880 | 1,323 |
| Gyros | 0,248 | 1,2 | 3,8 | 0,2 | 0,790 | 0,985 |
| Čokoládový Croissant | 0,228 | 1,5 | 3,9 | 0,2 | 0,800 | 0,984 |
| [Moët](http://localhost/dipl/product/show?name=Mo%C3%ABt+%26+Chandon+Imp%C3%A9rial+) | 0,790 | 0 | 4,1 | 0,2 | 0,500 | 0,895 |

Jak lze vidět v tabulce výše, výstup expertního systému mírně změnil seřazení doporučených produktů oproti výstupu SVD na obrázku č. 21. Můžeme si všimnout produktů z kategorie „Pečivo“, které mají menší míru podobnosti k hodnoceným produktům uživatele. Na druhou stranu získávají vysoký koeficient produktu, což znamená zakoupení podobného produktu tentýž den. Díky těmto faktorům jsou produkty z této kategorie posunuty ve výsledném seznamu výše. Naopak produkt [Moët](http://localhost/dipl/product/show?name=Mo%C3%ABt+%26+Chandon+Imp%C3%A9rial+) se umístil v seznamu níže, jelikož uživatel zatím nekoupil žádný podobný produkt z kategorie „Alkohol“.

1. Implementace

V této kapitole jsou popsány technologie využité pro vývoj adaptivního systému. Nejprve je popsán programovací jazyk PHP a Framework Nette. Dále databázový systém MySQL a jeho formát MyISAM, které jsou využity pro ukládání dat aplikace a struktura databáze vyvíjené aplikace.

* 1. Programovací jazyk

Jako programovací jazyk byl vybrán PHP. Tento programovací jazyk byl vybrán z důvodu jeho všestrannosti, umožnění vytvářet webové aplikace, dobré komunikace s většinou běžně používaných databází (v expertním systém MySQL databáze).

PHP (zkratka pro PHP Hypertext Preprocessor) je skriptovací programovací jazyk zabudovaný na straně serveru, všechny úkony jsou prováděny na hostingu, tudíž není zatíženo uživatelovo zařízení. Je určen k vývoji webových aplikací a dynamických webových stránek většinou v kombinaci s operačním systémem, databázovým systémem (MySQL) a webovým serverem Apache. Podporuje přístup k databázovému systému (např.: MySQL, PostgeSQL, Oracle). Díky své jednoduchosti a bohaté zásobě funkcí se stal jazyk oblíbeným. V PHP jsou napsány například Wikipedie nebo Facebook. [6][8]

**Framework Nette**

Pro vývoj byl zvolen Framework Nette. Framework podporuje vývoj v programovacím jazyce PHP. Je zaměřen na objektové programování a s tím spojenou znovu použitelnost kódu. Eliminuje bezpečnostní rizika. Využívá událostmi řízené programování a z velké části je založen na použití komponent. Framework nabízí jednoduchou, úspornou a srozumitelnou syntaxi. Nabízí intuitivní ladící nástroje, má aktivní komunitu v České republice a obsáhlou dokumentaci v českém jazyce. Autorem a hlavním programátorem je David Grudl.[9]

**Knihovny a externí zdrojové kódy**

Jak bylo zmíněno výše, expertní systém využívá pro doporučování produktů algoritmy TF-IDF a SVD. Tyto algoritmy mají velkou základnu v programovacím jazyce Python, na internetu jsou k dostání zdrojové kódy nebo přímo knihovny pro tyto algoritmy. Avšak navrhovaný doporuočvací systém je implementován v jazyce PHP a bylo nutné tyto algoritmy do systému přidat. Implementace algoritmu TF-IDF není složitá, tudíž pro tento algoritmus nebyla přidána žádná knihovna a algoritmus byl naimplementován do prostředí PHP pomocí tutoriálů popisující algoritmus v programovacím jazyce Python.

Oproti tomu větší složitost algoritmu SVD vyžadovala využití externího zdrojového kódu. Jedná se o zdrojový kód algoritmu SVD v programovacím jazyce PHP naimplementovaného programátorem Yehia Abedem v roce 2010. Zdrojový kód obsahuje funkce a operace k provedení algoritmu SVD[[1]](#footnote-1).

Doporučení pomocí algoritmu SVD vyžaduje složité maticové operace. Z toho důvodu je do projektu přidána knihovna NumPhp[[2]](#footnote-2), jedná se o známou knihovnu NumPy programovacího jazyka Python, naimplementovanou jako balíček pro programovací jazyk PHP. Tato knihovna umožnuje práci s vektory, maticemi a obecně vícerozměrnými poli.

* 1. Databáze

Aplikace využívá MyISAM formát úložiště dat databázového systému MySQL. MySQL je relační databáze typu DBMS (database management system) a vychází z deklarativního programovacího jazyka SQL. V poslední době je tento relační systém velice oblíbený díky své jednoduchosti a rychlosti. MySQL dosahuje vyšší tychlosti díky omezením, které ostatní databázové systémy nemají. Zkratka znamená „My Structured Query Language“ v překladu systém pro řízení databází. Do databáze lze ukládat různá data (text, obrázky, hodnoty atd.) s nimiž se dá jednoduše pracovat (třídit, řadit, filtrovat atd.). Systém se nejčastěji používá v kombinaci s programovacím jazykem PHP, který umožňuje přístup k uloženým datům.

Pro jednoduchou správu databáze se využívá nástroj PhpMyAdmin. Nástroj je napsán v jazyce PHP, který umožňuje kompletní správu a práci s databází (vytváření tabulek, mazání, editování, vytváření databází apod.) přes webové rozhraní. [7]

MyISAM formát je využíván především pro velké množství čtení dat a menším nebo žádným množstvím ukládání dat. Typickým využitím jsou datové sklady, protože obsahuje velké množství tabulek a dat a aktualizace se provádí v době, když se databáze nepoužívá. Rychlost čtení způsobuje struktura indexů, kde každá položka ukazuje na záznam a je řazena od začátku souboru. Vkládání dat je také jednoduché, jelikož se záznamy pouze zařadí na konec datového souboru. Operace mazání a aktualizace jsou však problematičtější, odstranění musí zanechat prázdné místo, jinak by se změnily posuny řádků. Hlavní nevýhodou MyISAM je absence transakcí a nejsou podporovány cizí klíče. MyISAM podporuje FULLTEXT indexování.[7]

* 1. Webové rozhraní aplikace

Tato část se bude věnovat webovému rozhraní aplikace. Administrativní části, ve které jsou vidět statistiky uživatelů a uživatelské části, která provádí uživatele e-shopem od registrace až po odeslání výsledků.

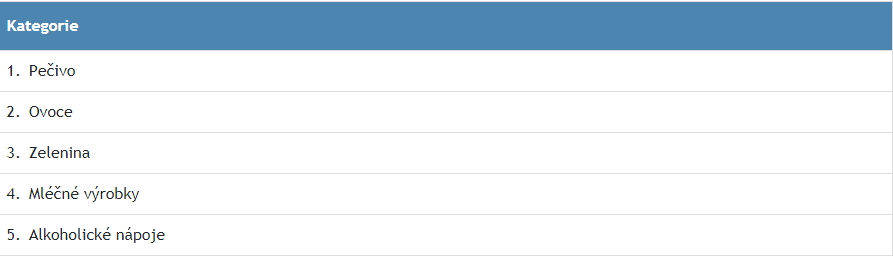
* + 1. Administrativní část

Administrativní část webové aplikace vidí pouze administrátoři aplikace. Tato část má hlavní stránku a ta obsahuje seznam všech uživatelů, kteří provedli testování. Z této stránky je možné otevřít si detaily o uživatelích a detaily hodnocení jednotlivých přístupů daného uživatele.

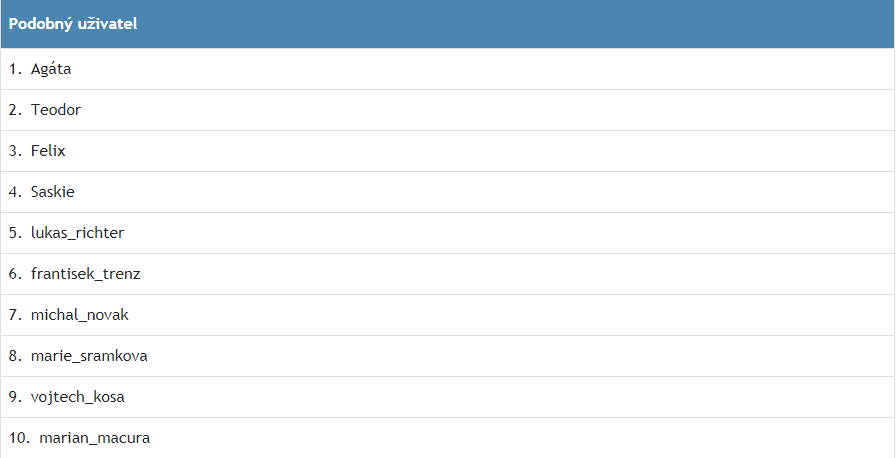


Obrázek 24: Administrativní část - seznam uživatelů (zdroj: testovací e-shop)

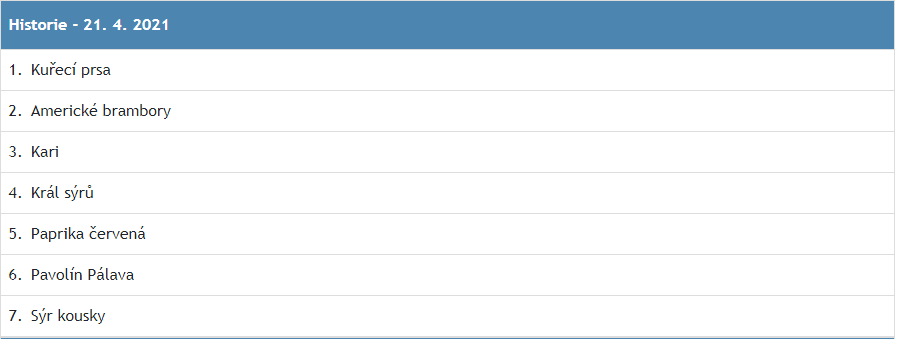
V detailu uživatele se nachází informace o oblíbených kategoriích, kolik uživatelů mají zvoleny podobné kategorie a jsou aktivnímu uživateli podobní, jaké nákupy uživatel provedl a jaké produkty prohlížel. Na následujících obrázcích jsou znázorněny informace o vybraném uživateli.



Obrázek 25: Oblíbené kategorie vybraného uživatele



Obrázek 26: Podobní uživatelé k vybranému uživateli (zdroj: testovací e-shop)



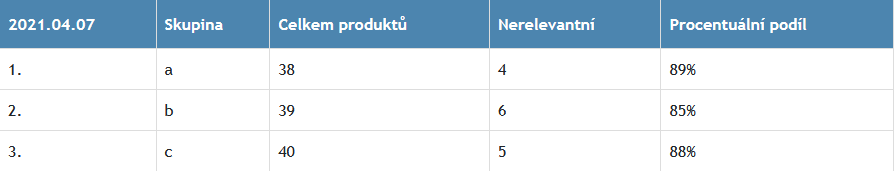
Obrázek 27: Historie nákupu vybraného uživatele (zdroj: testovací e-shop)



Obrázek 28: Prohlížené produkty vybraného uživatele (zdroj: testovací e-shop)

V tabulkách výše získáme základní informace o uživateli, jaké jsou jeho preference, do které skupiny podobných uživatelů patří a jaké úkony provádí na e-shopu.

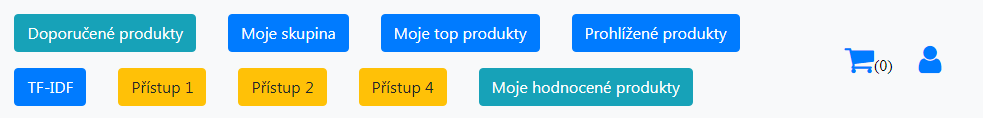
Druhou, důležitější částí v administrativě, je kontrola přístupů jednotlivých uživatelů. U každého uživatele je možnost zobrazit výsledky přístupů. Detail výsledků je rozdělen do tabulek podle data testování daným uživatelem. Tyto tabulky jsou rozděleny do čtyř sloupců - *skupina*, *celkem produktů*, *nerelevantní* a *procentuální podíl*. Sloupec skupina udává, pro kterou skupinu v daném přístupu výsledky v daném řádku platí. Další sloupec ukazuje, kolik bylo celkem doručeno produktů v dané skupině pro daného uživatele. Předposlední sloupec zobrazuje počet produktů, které daný uživatel označil jako nerelevantní a nakonec poslední sloupec reprezentuje procentuální podíl relevantních produktů z celkového počtu produktů. Detail výsledků přístupu 1 pro vybraného uživatele je na následujícím obrázku.



Obrázek 29: Detail výsledků vybraného přístupu (zdroj: testovací e-shop)

* + 1. Uživatelská část

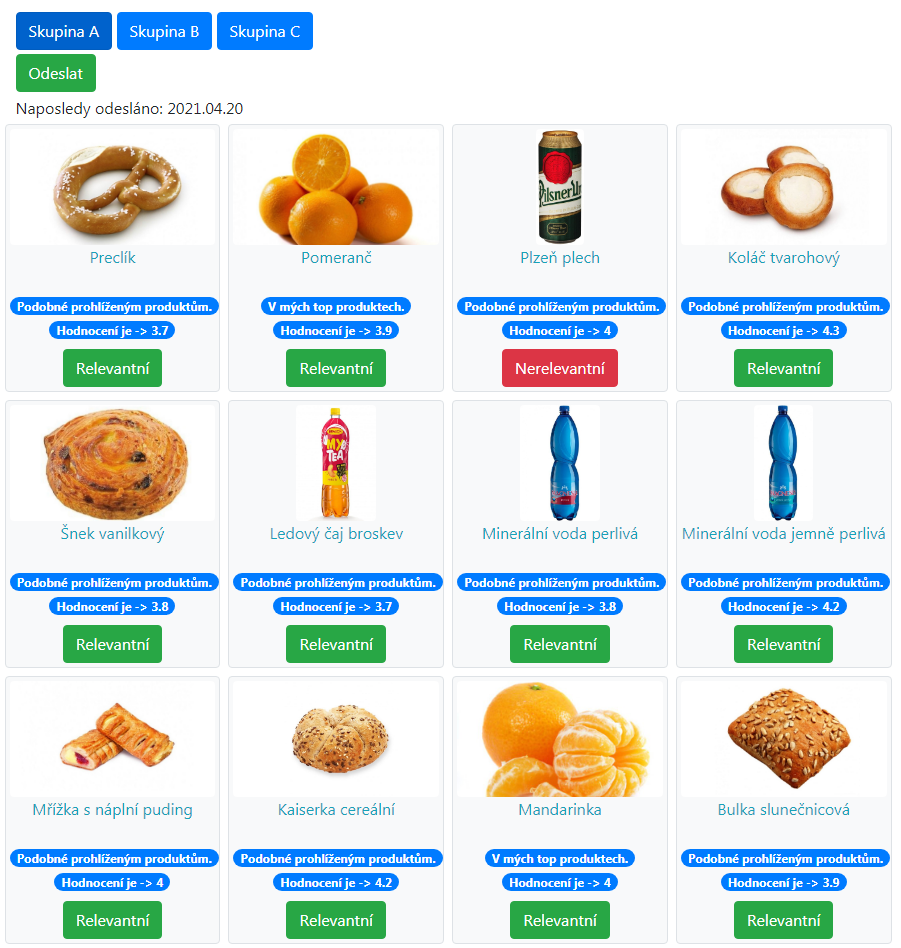
Tato část je pro všechny uživatele stejná. Hlavním prvkem této části je uživatelský panel, který umožňuje uživatelům lepší orientaci v systému. Panel obsahuje ikonu panáčka, pomocí kterého se uživatelé mohou odhlásit nebo si mohou zobrazit nastavení obsahující výběr oblíbených kategorií. Druhou ikonou v uživatelském panelu je košík, díky kterému uživatelé mohou dokončovat nákupy či sledovat historii svých nákupů.



Obrázek 30: Uživatelský panel (zdroj: testovací e-shop)

Uživatelský panel obsahuje několik barevně odlišných tlačítek. Modrá tlačítka po kliknutí zobrazí stránku s výsledky některého z algoritmů. *Moje skupina* zobrazuje podobné uživatele aktivnímu uživateli, tlačítko *Moje top produkty* zobrazuje výstup algoritmu SVD k hodnoceným produktům. Tlačítko *Prohlížené produkty* nabízí uživateli seznam prohlížených produktů a nakonec tlačítko *TF-IDF* dodává uživateli seznam podobných produktů k prohlíženým produktům. Tyto algoritmy jsou popsány v kapitole 3.3 Přístup pro doporučování i s obrázky jednotlivých výstupů.

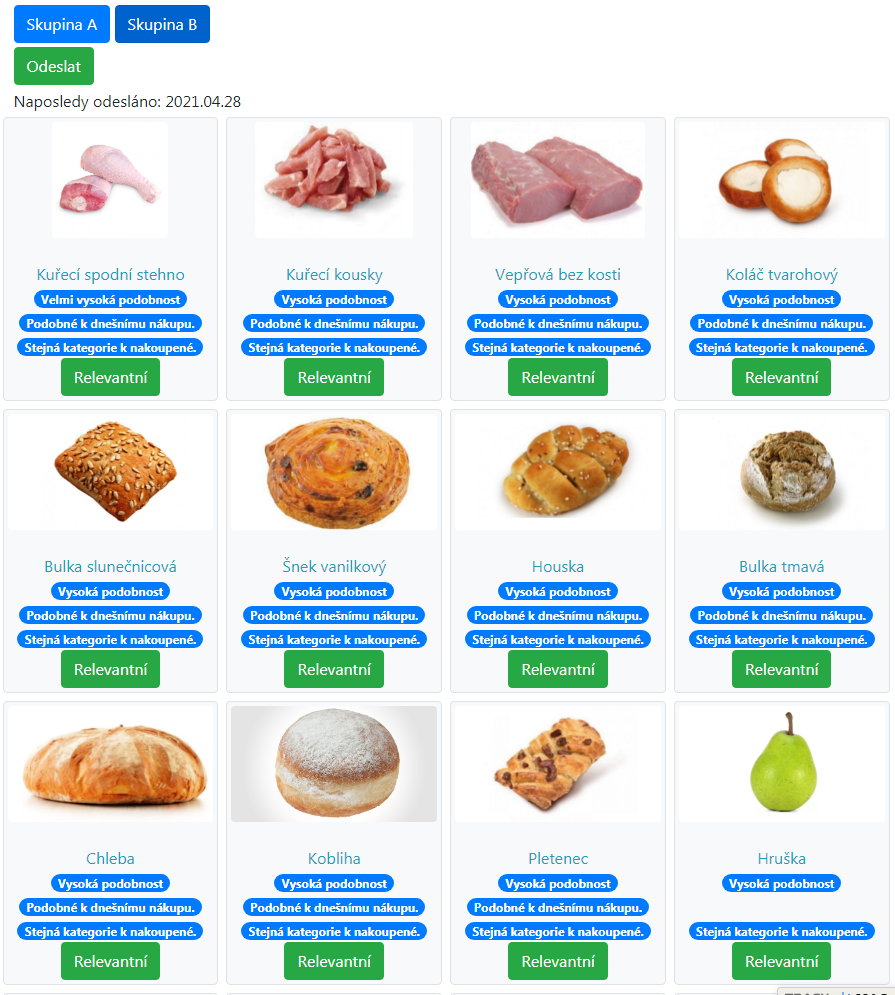
Žlutá tlačítka reprezentují jednotlivé přístupy. Tyto přístupy doporučují produkty na základě kombinace jednotlivých algoritmů, každý přístup má na své stránce popis, aby uživatel věděl, proč se mu dané produkty doporučují. Nahoře na stránce přístupu jsou tlačítka záložek jednotlivých skupin (Skupina A, Skupina B, Skupina C), které uživatel postupně prochází a kontroluje relevantnost výsledků. Následuje tlačítko „Odeslat“ pro odeslání výsledků do databáze a datum posledního odeslání výsledků aktivním uživatelem, tedy poslední testování vybrané skupiny v daném přístupu.



Obrázek 31:Doručení produktů u přístupu 1 a přístupu 2 (zdroj: testovací e-shop)

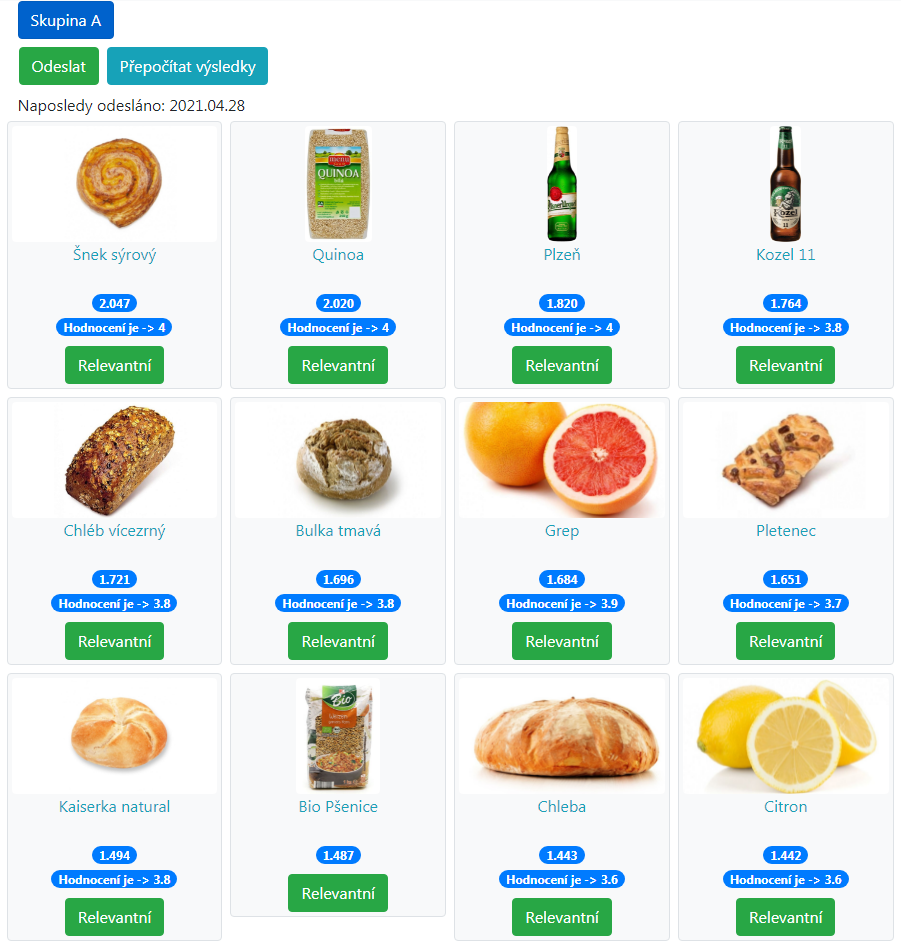
Hlavní částí každého přístupu je seznam doporučených produktů seřazených podle podobnosti. Na obrázku můžeme vidět nejlepších 12 produktů skupiny A v přístupu 1. Každá položka se skládá ze 3 částí - obrázek s názvem produktu, informace proč se produkt doporučuje (Podobné k prohlíženým, V mých top produktech, Hodnocení produktu) a tlačítko pro zvolení produktu za relevantní či nerelevantní.

Poslední přístup 4 je velice podobný přístupům 1 a 2, liší se pouze v části, proč se daný produkt doporučuje, jelikož využívá jiný algoritmus pro doporučení relevantních produktů. Místo popisků proč se produkt doporučuje na obrázku č. 31, obsahuje popisky týkající se historie zakoupení (Velmi vysoká podobnost, Podobné k dnešnímu nákupu, Stejná kategorie k nakoupené).



Obrázek 32: Doručení produktů u přístupu 4 (zdroj: testovací e-shop)

Poslední část systému, se kterým se uživatelé setkají, se nachází pod tlačítkem *Doporučené produkty*. Zde je seznam nejrelevantnějších produktů z přístupu 1 skupiny B a přístupu 4 skupiny B, tyto doporučené produkty jsou výstup expertního systému a doručují uživateli produkty na základě vstupních parametrů.



Obrázek 33: Doporučené produkty pomocí expertního systému (zdroj: testovací e-shop)

Můžeme si všimnout, že zde přibylo tlačítko *Přepočítat výsledky*. Toto tlačítko umožňuje přepočítat výsledky, pokud aktivní uživatel provedl nějakou změnu v přístupech, které vstupují do expertního systému.

1. Testování

Tato kapitola se věnuje testování a jejím výsledkům. Testování provádělo několik respondentů. Testování bylo rozděleno na 4 fáze testování, které jsou popsány níže v této kapitole

* 1. Popis testování

V této části budou popsány všechny testování a jejich postup s instrukcemi pro uživatele.

* + 1. První testování

První testování mělo zjistit, zdali navržený systém doporučuje relevantní produkty aktivnímu uživateli po provedení posloupnosti několika kroků. Prvního testování se účastnilo 30 respondentů z řad studentů, rodiny a přátel. Každý z respondentů obdržel instrukce s posloupností jednotlivých kroků. Instrukce k prvnímu testování jsou jako příloha **První testování - instrukce**.

Instrukce jsou navrženy tak, aby uživatele provedl systémem a uživatel prošel všemi nutnými kroky až k odesláním výsledků. Při registraci obdrží každý uživatel jedinečné ID, které umožňuje systému sbírat informace o chování uživatele v systému. Po ukončení a odeslání všech přístupů uživatelé vyplnili anonymní dotazník, kde odpovídali na otázky týkající se převážně relevantnosti doporučených produktů.

Dotazník na první testování obsahuje 5 otázek, z nichž 4 jsou povinné a jedna nepovinná. Nepovinná otázka se týká Přístupu 4, který byl přidán během prvního testování, a tedy někteří uživatelé Přístup 4 netestovali.

* + 1. Druhé testování

Druhé testování sytému je velice podobné prvnímu, avšak trochu se liší. Cílem bylo, aby všem respondentům byly doručeny nejpodobnější výsledky, a aby se zjistila odchylka v doporučovacím systému. Proto byli ze skupiny testujících respondentů odebráni ti, jejichž výsledky testování dopadly velice dobře a stejně tak ti, jejichž výsledky testování dopadly velice špatně. Jejich výsledky by mohly zkreslit celkové výsledky, jelikož by tito uživatelé mohli být vždy spokojeni, nebo naopak vždy nespokojeni. Byli tedy označeni jako „vždy spokojený“ a „vždy nespokojený“ zákazník a byli odebráni.

Abychom doručili vybraným respondentům stejné produkty, je potřeba toho docílit nastavením stejných vstupních parametrů. Od vybraných uživatelů byly získány kategorie a vybráno 5 nejčetnějších. Následně bylo zvoleno 40 produktů pro ohodnocení. Z oblíbených kategorií bylo vybráno 25 produktů a 15 produktů z jiných kategorií. Následně bylo vybráno 40 produktů pro prohlédnutí. Opět bylo vybráno 25 produktů z oblíbených kategorií a 15 produktů z jiných kategorií. Vybrané produkty pro prohlédnutí nebyly hodnoceny uživatelem. Takto vybrané a ohodnocené produkty byly naimportovány všem testerům a jejich úkolem bylo pouze prohlédnout vybrané produkty a odeslat výsledky relevantnosti.

Postup pro druhou fázi testování byl velice podobný první fázi testování. Druhého testování se účastnilo 20 vybraných respondentů a každý z respondentů prošel první fází testování. Jednotlivé kroky pro druhou fázi testování jsou v příloze jako **Druhé testování - instrukce**.

Instrukce v druhé fázi testování jsou navrženy tak, aby všichni uživatelé měli stejné výsledky. Uživatelé jsou požádáni o kontrolu vygenerovaných dat a postupným provedením několika specifických kroků. Po splnění několika málo kroků vedoucím k odeslání výsledků jsou uživatelé požádáni o vyplnění dotazníku týkajícího se právě doručeným produktům.

Dotazník pro druhé testování obsahuje 7 otázek, z nichž všechny jsou povinné. První 4 otázky jsou převzaty z prvního dotazníku. Další otázky se týkají doporučování podle historie nákupů a popisů u doporučených produktů.

* + 1. Expertní systém

Testování výstupu expertního systému mělo za cíl zjistit, zda došlo ke zlepšení doporučení produktů aktivnímu uživateli. Po provedení předchozích testování, mají uživatelé vytvořený vlastní profil. Na profilu je zaznamenána veškerá aktivita, kterou využívá expertní systém pro doporučení nejrelevantnějších produktů. Instrukce k této fázi testování jsou k nalezení v příloze jako **Třetí testování (ES) - instrukce.**

Tato fáze testování již není nijak náročná, jelikož uživatelé mají vytvořené své profily a expertní systém pouze získá vstupní parametry z přístupu 1 skupiny B a přístupu 4 skupiny B. Uživatelé zvolí nerelevantní produkty a potvrdí odesláním do databáze. Následně vyplní dotazník, který se týká testování expertního systému a také porovnání testovacího e-shopu s internetovým obchodem košík.cz, jehož průběh testování je popsán v následující podkapitole.

* + 1. Košík.cz

Pro porovnání doporučovacího systému s jiným e-shopem byl zvolen Košík.cz. Tento e-shop byl inspirací pro vytvoření vlastního s doporučovacím systémem. Veškeré produkty na testovacím e-shopu jsou převzaty právě z internetového obchodu Košík.cz nabízející potraviny.

Internetový obchod košík.cz nabízí podobné produkty k prohlíženým a také před zaplacením objednávky doporučuje podobné produkty, které by si uživatel mohl ještě zakoupit. Můžeme tedy porovnat výsledky doporučení podobných produktů k prohlíženým a také porovnat podobné produkty k zakoupeným, nebo připravené k zaplacení v případě košíku.cz. Jednotlivé kroky pro tuto fázi testování jsou v příloze jako **Čtvrté testování (www.kosik.cz) - instrukce.**

Instrukce v poslední fázi jsou nastaveny tak, aby uživatel porovnal doporučování produktů na testovacím e-shopu a reálným e-shopem košík.cz. Uživatelé jsou požádáni o provedení nákupu v testovacím e-shopu a vložení stejných produktů do košíku v internetovém obchodě košík.cz. Následně jsou požádáni o porovnání doporučených výsledků. Nakonec uživatelé vyplní otázky v dotazníku z předchozí fáze testování.

* 1. Výsledky testování

Tato podkapitola se věnuje výsledkům jednotlivých testování.

* + 1. První testování

Výsledky prvního testování byly zaznamenávány a analyzovány v tabulkách Microsoft Excel. Soubor obsahuje záložky pro každý přístup. V každém přístupu jsou sloupce pro danou skupinu a každý sloupec skupiny je rozdělen na celkový počet doručených produktů. Kolik těchto označil tester jako nerelevantních a celkové procento úspěšnosti. Každá záložka také obsahuje sloupec s testery a jejich celkové procento pro všechny skupiny daného přístupu. Tyto výsledky jsou přidány jako příloha k diplomové práci.

První testování bylo rozděleno na dvě fáze. Důvodem byla neschopnost účastnit se všech testerů prvního testování v jeden den. Díky první fázi testování, které se zúčastnila většina testerů, bylo možné zjistit nedostatky a tyto nedostatky upravit pro zbylé testery. Tedy druhou fázi testování. Díky první fázi testování byl upraven Přístup 3 do Přístupu 4.

V následující tabulce je výběr několika testerů pro Přístup 1, Skupina A. Testerům bylo doručeno maximálně 40 produktů kombinací podobných produktů k prohlíženým a podobných produktů k již hodnoceným produktům.

Tabulka 9: První testování - Přístup 1, Skupina A

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 38 | 0 | 100% |
| User 2 | 38 | 1 | 97% |
| User 3 | 31 | 3 | 90% |
| User 4 | 37 | 5 | 86% |
| User 5 | 39 | 15 | 62% |
| User 6 | 39 | 17 | 56% |

Můžeme vidět, že počet doporučených produktů se blíží k maximálnímu počtu 40 doporučených produktů. Byli vybráni testeři z předních příček s vysokou úspěšností, ale také testeři ze spodních příček s nízkou úspěšností, tedy nejnižší úspěšnost se pohybuje nad 50%. Testeři *User 3* a *User 4*, kteří jsou vybráni ze středních příček, mají úspěšnost doporučených produktů nad 85%, tedy relativně vysokou. Průměrný procentuální podíl Přístupu 1, Skupiny A je 86%.

V následující tabulce jsou vybrání stejní uživatelé, avšak pro Přístup 2, také skupiny A.

Tabulka 10: První testování - Přístup 2, Skupina A

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 39 | 0 | 100% |
| User 2 | 38 | 0 | 100% |
| User 3 | 21 | 4 | 90% |
| User 4 | 37 | 5 | 81% |
| User 5 | 38 | 10 | 74% |
| User 6 | 39 | 21 | 46% |

S porovnáním s Přístupem 1 Skupiny A se pořadí testerů nezměnilo, můžeme si však povšimnout, že tester *User 6* označil větší polovinu produktů jako nerelevantní, nicméně tester *User 5* označil o 10% méně produkty jako nerelevantní. Průměrný procentuální podíl Přístupu 2, Skupiny A je 84%, pouze o 2% horší než přístup 1, Skupina A.

Jak bylo napsáno výše, Přístup 3 byl nahrazen Přístupem 4, který modifikuje vytvořený Přístup 3. V následující tabulce jsou výsledky Přístupu 3 stejných testerů, jako byli zvoleni v Přístupu 1 a Přístupu 2.

Tabulka 11: První testování - Přístup 3

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 79 | 12 | 85% |
| User 2 | 99 | 9 | 91% |
| User 3 | 65 | 11 | 83% |
| User 4 | 39 | 5 | 87% |
| User 5 | 70 | 17 | 76% |
| User 6 | 86 | 51 | 41% |

Maximální počet doporučených produktů byl nastaven na 100 v Přístupu 3, počet produktů odpovídal velikosti nákupu jednotlivých testerů. Můžeme vidět, že u všech testerů s velkým procentuálním podílem se tento podíl velice snížil, výjimkou je tester *User 4*, kterému systém doručil menší počet produktů a jeho procentuální podíl nerelevantních výsledků se snížil. Nejnižší procentuální podíl tohoto přístupu je 41%. Průměrný procentuální podíl Přístupu 3 je 77%. Tento procentuální podíl je velice nízký pro doporučovací systém a modifikací tohoto přístupu chceme dosáhnout zvýšení výsledného podílu.

Výsledky Přístupu 3 vedly k modifikaci tohoto přístupu, aby systém doručoval méně produktů a více relevantních. Jednotlivé modifikace pro Přístup 4 jsou popsány v kapitole 3.3.1 v části věnované **Přístupu** **4 a jeho modifikace**.

Jelikož byl Přístup 4 vytvořen po první fázi prvního testování, provedli testování Přístupu 4 pouze ti testeři, kteří se neúčastnili první fáze prvního testování. Pro lepší porovnání zlepšení Přístupu 4 jsou v následující tabulce vybrání ti uživatelé, kteří testovali Přístup 3 i Přístup 4.

Tabulka 12: První testování - Přístup 3, noví uživatelé

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 100 | 5 | 95% |
| User 2 | 65 | 11 | 83% |
| User 3 | 33 | 7 | 79% |
| User 4 | 100 | 32 | 68% |
| User 5 | 83 | 41 | 51% |
| User 6 | 87 | 50 | 43% |

V následujících dvou tabulkách jsou znázorněny výsledky skupin Přístupu 4, každá skupina má jiný výpočet koeficientu, tudíž půjde vidět zlepšení jednotlivých skupin oproti Přístupu 3. V tabulce níže jsou výsledky Skupiny A.

Tabulka 13: První testování - Přístup 4, Skupina A

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 48 | 0 | 100% |
| User 2 | 40 | 10 | 75% |
| User 3 | 8 | 0 | 100% |
| User 4 | 100 | 32 | 68% |
| User 5 | 22 | 7 | 68% |
| User 6 | 29 | 14 | 52% |

Počet doporučených produktů je oproti přístupu 3 menší, procentuální podíl relevantních výsledků však vzrostl. Zvýšení procentuálního podílu nerelevantních produktů se objevilo u testera *User 2*, kterému byl doručen stejný počet nerelevantních produktů, avšak celkový počet doručených produktů je nižší. Průměrný procentuální podíl této skupiny je 86%.

Tabulka 14: První testování - Přístup 4, Skupina B

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 74 | 4 | 95% |
| User 2 | 46 | 10 | 78% |
| User 3 | 21 | 0 | 100% |
| User 4 | 52 | 2 | 96% |
| User 5 | 30 | 3 | 90% |
| User 6 | 50 | 28 | 44% |

Doporučené produkty ve skupině B přístupu 4 se liší jak od přístupu 3, tak i od skupiny A. Důvodem je větší důraz na produkty ze stejné kategorie, jako jsou produkty nakoupené. Nejrazantnější rozdíl je u uživatelů *User 4* a *User 5*. Uživateli *User 4* systém doporučil o polovinu méně produktů s velice nízkým počtem nerelevantních produktů oproti přístupu 3 a skupině A v přístupu 4. Další velký rozdíl lze vidět u uživatele *User 5*, kterému systém doporučil o větší polovinu méně produktů než v přístupu 3 a o 1/3 více produktů než ve skupině A. V obou případech systém doporučil menší množství relevantních produktů.

V následující tabulce je souhrn všech přístupů prvního testování s celkovým počtem testujících osob a průměrným procentuálním podílem relevantních produktů na skupinu.

Tabulka 15: Souhrn výsledků prvního testování

|  |  |  |  |
| --- | --- | --- | --- |
| **Skupina přístupu** | **Počet uživatelů** | **Průměrný počet doporučených produktů** | **Průměrný procentuální podíl** |
| **Přístup 1** | | | |
| Skupina A | 30 | 34 | 86% |
| Skupina B | 30 | 36 | 85% |
| Skupina C | 30 | 37 | 85% |
| **Přístup 2** | | | |
| Skupina A | 30 | 32 | 84% |
| Skupina B | 30 | 31 | 86% |
| Skupina C | 30 | 32 | 86% |
| Skupina D | 30 | 33 | 85% |
| **Přístup 3** | | | |
| Skupina A | 25 | 62 | 77% |
| **Přístup 4** | | | |
| Skupina A | 20 | 20 | 86% |
| Skupina B | 20 | 30 | 90% |

Z tabulky výše lze vidět, že přístup 1 a 2 mají v průměru 85% relevantních doporučených produktů. Přístup 3 doručil velké množství produktů, nicméně relevantnost těchto produktů je relativně nízká. Přístup 4, který je modifikace přístupu 3 doporučil sice menší množství produktů, za to více relevatní.

Po skončení testování byli uživatelé požádáni o vyplnění dotazníku, jak bylo zmíněno výše. Dotazník se skládal z 5 otázek týkajících se testování. Otázky vypadaly následovně.

První otázka se týkala systému jako celku.

Obrázek 34: Dotazník - První testování, první otázka

Tato otázka byla pro všechny testery povinná. Pro většinu testerů byl systém intuitivní. Jedna třetina testerů označila systém jako spíše intuitivní a dva testeři označili systém jako neintuitivní.

Následovala otázka týkající se samotného doporučování. **„Odpovídaly navržené produkty v jednotlivých přístupech Vašim preferencím?“**

Obrázek 35: Dotazník - První testování, druhá otázka

Na tuto otázku odpověděli všichni testeři kladně, žádný z testerů nebyl s výsledky doporučených produktů nespokojen.

Následující otázky se týkaly konkrétních přístupů a jejich skupin. Cílem bylo zjistit, které kombinace v různých přístupech dodávají nejrelevantnější výsledky. Otázka k přístupu 3 nebyla zadaná, jelikož výstupem byla pouze jedna skupina, která byla následně modifikována jako vstup pro přístup 4.

Obrázek 36: Dotazník - První testování, třetí otázka

Pro polovinu testerů byla zvolena skupina A. Tato skupina kombinuje výsledné doručené produkty v poměru 30% podobné k prohlíženým produktům a 70% podobné k hodnoceným produktům. Podle analýzy ze souboru Microsoft Excel s výslednými daty má tato skupina nejvyšší průměrný procentuální podíl.

Obrázek 37: Dotazník - První testování, čtvrtá otázka

Podle analýzy se lišil průměrný procentuální podíl všech skupin maximálně o 1%. Jako nejlepší však zvolila třetina testerů skupinu D, která kombinuje 30% prohlížených a 70% podobných k hodnoceným.

Obrázek 38: Dotazník - První testování, pátá otázka

Tři čtvrtiny testerů zvolilo skupinu B. Z tabulek výše je to patrné, jelikož vybraní testeři měli větší procentuální podíl relevantních výsledků ve skupině B oproti skupině A v daném přístupu 4.

* + 1. Druhé testování

Výsledky druhého testování byly analyzovány a zapsány v souboru Microsoft Excel v přehledných tabulkách. Soubor je příloha k diplomové práci. Obsahuje sloupce pro každou skupinu přístupu. Sloupec vybraných 20 testerů a celkový procentuální podíl. Jednotlivé přístupy jsou rozděleny v přehledných záložkách (příklad: přístup1-fáze2).

Druhé testování probíhalo v jedné fázi a díky zvolenému postupu byly uživatelům doručeny stejné produkty. V následujících tabulkách je vybráno několik uživatelů z druhé fáze testování a porovnání jejich výsledků.

Počet doporučených produktů je vždy stejný, jelikož všichni uživatelé mají stejné doporučené produkty. V následující tabulce jsou znázorněny výsledky skupiny A prvního přístupu.

Tabulka 16: Druhé testování - Přístup 1, Skupina A

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 37 | 0 | 100% |
| User 2 | 37 | 0 | 100% |
| User 3 | 37 | 0 | 100% |
| User 4 | 37 | 2 | 95% |
| User 5 | 37 | 2 | 95% |
| User 6 | 37 | 3 | 92% |

V následující tabulce jsou výsledky skupiny B. Můžeme vidět mírnou změnu v relevantnosti doporučených produktů. To způsobuje změna v procentuální kombinaci doporučených produktů.

Tabulka 17: Druhé testování - Přístup 1, skupina B

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 37 | 0 | 100% |
| User 2 | 37 | 0 | 100% |
| User 3 | 37 | 0 | 100% |
| User 4 | 37 | 0 | 100% |
| User 5 | 37 | 1 | 97% |
| User 6 | 37 | 3 | 92% |

V následující tabulce je zobrazena skupina C. Tato kombinace doporučených produktů vyšla nejlépe. Skupina C využívá 70% produktů z top produktů podobných k prohlíženým.

Tabulka 18: Druhé testování - Přístup 1, Skupina C

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 37 | 0 | 100% |
| User 2 | 37 | 0 | 100% |
| User 3 | 37 | 0 | 100% |
| User 4 | 37 | 0 | 100% |
| User 5 | 37 | 0 | 100% |
| User 6 | 37 | 3 | 92% |

Následující tabulky obsahují výsledky přístupu 2. Všimněme si postupného zlepšování v doporučování produktů. Systém doporučuje méně nerelevantních produktů, pokud výsledný seznam produktů obsahuje víc z top produktů podobných k prohlíženým.

Tabulka 19: Druhé testování - Přístup 2, Skupina A

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 40 | 0 | 100% |
| User 2 | 40 | 0 | 100% |
| User 3 | 40 | 1 | 98% |
| User 4 | 40 | 2 | 95% |
| User 5 | 40 | 2 | 95% |
| User 6 | 40 | 4 | 90% |

Tabulka 20: Druhé testování - Přístup 2, Skupina B

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 38 | 0 | 100% |
| User 2 | 38 | 0 | 100% |
| User 3 | 38 | 1 | 98% |
| User 4 | 38 | 2 | 95% |
| User 5 | 38 | 2 | 95% |
| User 6 | 38 | 3 | 92% |

Tabulka 21: Druhé testování - Přístup 2, Skupina C

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 36 | 0 | 100% |
| User 2 | 36 | 0 | 100% |
| User 3 | 36 | 1 | 98% |
| User 4 | 36 | 2 | 95% |
| User 5 | 36 | 2 | 95% |
| User 6 | 36 | 2 | 95% |

Tabulka 22: Druhé testování - Přístup 2, Skupina D

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 36 | 0 | 100% |
| User 2 | 36 | 0 | 100% |
| User 3 | 36 | 0 | 100% |
| User 4 | 36 | 2 | 95% |
| User 5 | 36 | 1 | 98% |
| User 6 | 36 | 1 | 98% |

Výsledky přístupu 4 jsou znázorněny v následujících tabulkách. Ve skupině A byly dole v seznamu doporučených produktů doručeny produkty z kategorie pečiva. Tyto produkty se však ve skupině B neobrazovaly.

Tabulka 23: Druhé testování - Přístup 4, Skupina A

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 82 | 0 | 100% |
| User 2 | 82 | 0 | 100% |
| User 3 | 82 | 5 | 94% |
| User 4 | 82 | 11 | 87% |
| User 5 | 82 | 1 | 99% |
| User 6 | 82 | 7 | 91% |

Někteří uživatelé vyhodnotili doporučené produkty z kategorie pečivo za relevantní z důvodu zvolené této kategorie jako oblíbené.

Tabulka 24: Druhé testování - Přístup 4, Skupina B

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 61 | 0 | 100% |
| User 2 | 61 | 0 | 100% |
| User 3 | 61 | 5 | 100% |
| User 4 | 61 | 3 | 95% |
| User 5 | 61 | 0 | 98% |
| User 6 | 61 | 6 | 98% |

Následující tabulka obsahuje souhrn výsledků druhého testování.

|  |  |  |  |
| --- | --- | --- | --- |
| **Skupina přístupu** | **Počet uživatelů** | **Průměrný počet doporučených produktů** | **Průměrný procentuální podíl** |
| **Přístup 1** | | | |
| Skupina A | 20 | 37 | 99% |
| Skupina B | 20 | 37 | 99% |
| Skupina C | 20 | 37 | 99% |
| **Přístup 2** | | | |
| Skupina A | 20 | 40 | 97% |
| Skupina B | 20 | 38 | 98% |
| Skupina C | 20 | 36 | 97% |
| Skupina D | 20 | 36 | 98% |
| **Přístup 4** | | | |
| Skupina A | 20 | 82 | 96% |
| Skupina B | 20 | 61 | 97% |

Respondenti po testování odpověděli na několik otázek, týkajících se druhého testování. Jak je zmíněno výše, dotazník obsahoval 7 otázek.

Obrázek 39: Dotazník - Druhé testování, první otázka

První otázka byla stejná jako v předchozím dotazníku a opět všichni testeři byli s navrženými produkty spokojeni.

Následující dvě otázky se týkaly jednotlivých skupin přístupů 1 a 2. Jelikož se seznam doporučených produktů liší v několika produktech, je odpověď na tyto otázky individuální. Někomu více vyhovovaly produkty podobné k prohlíženým a jiným podobné produkty k hodnoceným.

Obrázek 40: Dotazník - Druhé testování, druhá otázka

Obrázek 41: Dotazník - Druhé testování, třetí otázka

Obrázek 42: Dotazník - Druhé testování, čtvrtá otázka

Čtvrtá otázka se týkala doporučení produktů podle historie nákupů. Většina respondentů zvolila skupinu B. Ovlivňuje to především menší počet doporučených produktů s větším procentem relevantnosti.

Obrázek 43: Dotazník - Druhé testování, pátá otázka

Otázka byla zaměřená na seřazení produktů. Většina produktů do seznamu patřila, některé produkty však měly vyšší podobnost, i když byly podobné produktům zakoupeným dříve. To způsobilo změnu v seřazení doporučených produktů a častější odpověď „Spíše ano“.

Obrázek 44: Dotazník - Druhé testování, šestá otázka

Jak bylo zaznamenáno v tabulkách č. 19 až č. 27, výsledky testování dopadly velmi dobře, tudíž většina respondentů odpověděla kladně.

**„Systém u navržených produktů v Přístup 1, 2 a 4 vysvětloval, proč daný produkt navrhl (např. V mých top produktech, Vysoká podobnost, Stejná kategorie k nakoupeným). Byly pro Vás tyto informace přínosné pro rozhodnutí, zda je produkt relevantní/nerelevantní?“**

Obrázek 45: Dotazník - Druhé testování, sedmá otázka

Tato otázka měla zjistit, zda popisky u produktů proč se produkt doporučuje, mají vliv na rozhodnutí uživatele. Zda je produkt relevantní nebo nerelevantní. Několik uživatelů odpovědělo „Ne“ a „Spíše ne“. Bylo zjištěno, že tyto popisky někteří uživatelé vůbec nečetli a rozhodovali se na základě toho, zda je produkt z oblíbených kategorií, nebo podobný k prohlíženým.

* + 1. Expertní systém

Testování expertního systému bylo provedeno v jedné fázi. Výstup expertního systému byl přístup 1 skupina B s přístupem 4. Tyto skupiny byly vybrány z důvodu nejlepších výsledků. Testování se účastnilo 20 respondentů z řad studentů a kamarádů, kteří se účastnili fáze 2. V následujících tabulkách jsou výsledky vybraných uživatelů.

Tabulka 25: Třetí testování - testování ES

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 80 | 0 | 100% |
| User 2 | 81 | 0 | 100% |
| User 3 | 78 | 0 | 100% |
| User 4 | 81 | 2 | 98% |
| User 5 | 79 | 2 | 97% |
| User 6 | 74 | 3 | 96% |

Z tabulky výše můžeme vidět, že počet doporučených produktů se lehce liší. Důvodem je zvolený počet nerelevantních produktů ve skupině B přístupu 1 a skupině B přístupu 4. Počet nerelevantních produktů je velice nízký na počet doporučených produktů, jelikož produkty, které uživatel označil ve vybraných přístupech za nerelevantní, nejsou doručeny pomocí ES.

V následující tabulce lze vidět souhrn všech testujících třetího testování.

|  |  |  |  |
| --- | --- | --- | --- |
| **Skupina přístupu** | **Počet uživatelů** | **Průměrný počet doporučených produktů** | **Průměrný procentuální podíl** |
| Skupina A | 16 | 78 | 99% |

Dotazník týkající se tohoto testování obsahoval dvě otázky. První otázka se ptala uživatelů, zda byly doporučené produkty relevantní.

Na otázku odpověděli všichni testující uživatelé kladně. Většina z nich zvolila možnost Ano a pouze 4 zvolili možnost spíše ano.

Druhá otázka k testování ES se zabývala, zda produkty doporučené pomocí ES odpovídají jejich aktivitě na e-shopu. Jedná se o oblíbené kategorie, prohlížení, hodnocení a nákupy.

Odpovědi na druhou otázku jsou opět všechny kladné. Většina testujících zvolila možnost Ano, tedy produkty doporučené pomocí ES jsou relevantní k jejich aktivitě na e-shopu.

* + 1. Košík.cz

Poslední fáze testování měla porovnat doporučení produktů na e-shopu košík.cz a testovacím e-shopem. Uživatelé porovnávali doporučování produktů na základě zakoupeného zboží, v internetovém obchodě košík.cz vloženého zboží do košíku. Výsledky testování jsou znázorněny v následujících tabulkách.

Tabulka 26: Čtvrté testování - nákup v e-shopu www.kosik.cz

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 36 | 14 | 61% |
| User 2 | 36 | 16 | 56% |
| User 3 | 36 | 14 | 61% |
| User 4 | 36 | 13 | 64% |
| User 5 | 36 | 15 | 58% |
| User 6 | 36 | 15 | 58% |

Z tabulky výše můžeme vidět, že e-shop www.kosik.cz podobné produkty doporučuje. Nicméně doporučuje také jiné produkty z jiných kategorií.

Tabulka 27: Čtvrté testování - nákup v testovacím e-shopu

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Počet doručených produktů** | **Nerelevantní produkty** | **Procentuální podíl** |
| User 1 | 16 | 0 | 100% |
| User 2 | 16 | 0 | 100% |
| User 3 | 16 | 0 | 100% |
| User 4 | 16 | 0 | 100% |
| User 5 | 16 | 0 | 100% |
| User 6 | 16 | 0 | 100% |

V tabulce č. 26 jsou výsledky z testovacího e-shopu, kdy byl uživatelům vygenerován stejný nákup, který provedli v e-shopu www.kosik.cz. Navržený systém doporučil méně produktů, avšak všechny produkty spadaly do stejných kategorií jako zakoupené produkty.

|  |  |  |  |
| --- | --- | --- | --- |
| **Skupina přístupu** | **Počet uživatelů** | **Průměrný počet doporučených produktů** | **Průměrný procentuální podíl** |
| **www.kosik.cz** | | | |
| Skupina A | 16 | 36 | 59% |
| **Testující e-shop** | | | |
| Skupina A | 16 | 0 | 100% |
| Skupina B | 10 | 0 | 100% |

Testující odpovídali v dotazníku na dvě otázky týkající se tohoto testování. První otázka se ptala uživatelů, zda byly doporučené produkty na e-shopu www.kosik.cz relevantní k vybranému nákupu.

Obrázek 46: Dotazník - čtvrté testování, první otázka

Z obrázku č. 46 lze vidět, že polovina uživatelů označila doporučené produkty v internetovém obchodě Košík za spíše relevantní. To odpovídá také výsledkům, kdy většině uživatelů bylo doručeno přes polovinu relevantních produktů. 5 uživatelů označila doporučené produkty za spíše nerelevantní.

Druhá otázka se týkala porovnání doporučení produktů v testovacím e-shopu a internetovým obchodem www.kosik.cz.

Obrázek 47: Dotazník - čtvrté testování, druhá otázka

* 1. Závěr testování

Z výše provedeného testování na reálných uživatelích můžeme říct, že navržený a implementovaný systém doporučování produktů je hodnocen ve většině případů kladně. Respondenti hodnotí systém jako intuitivní v zobrazování relevantních položek a doporučené produkty odpovídají jejich preferencím.

První fáze testování se zúčastnilo 30 respondentů. Z této fáze byli odebráni ti, kterých výsledné hodnoty byly moc nízké, nebo moc vysoké. Jedná se o eliminaci uživatelů, kteří by byli se systémem vždy spokojeni, nebo nespokojeni. Druhá fáze doporučovala všem respondentům stejné produkty. Za cíl bylo zjistit rozdíly v relevantnosti u jednotlivých uživatelů. Z provedeného testování bylo zjištěno, že se výsledky lišily pouze v několika položkách (u přístupu 1 a 2, byl rozdíl maximálně 3 položky).

Další fáze testování se zúčastnilo 17 respondentů. Všichni se museli zúčastnit předešlých fází testování, aby měli vytvořený profil s aktivitou (prohlížení, nákup, hodnocení). Systém doporučil finální seznam produktů pomocí ES ze vstupních parametrů získaných z aktivit uživatele (vstupní proměnné - hodnocení produktu, koeficient historie nákupu, podobnost k již hodnoceným produktům). Výsledky testování dopadly velmi dobře. Systém doporučil velké množství produktů s malým nebo žádným množstvím nerelevantních produktů.

Poslední testování se týkalo porovnání testovacího e-shopu s internetovým obchodem www.kosik.cz. Výsledky ukázaly, že www.kosik.cz doporučuje podobné produkty k nakoupeným, ale také jiné produkty. Internetový obchod tak nabídne větší množství produktů ze svého sortimentu, ale méně personalizovaného obsahu. Na druhou stranu navržený systém doporučuje menší množství produktů s větším procentuálním podílem relevantnosti.

ZÁVĚR

Cílem této diplomové práce byl návrh a implementace vlastního doporučovacího systému v internetovém obchodě zaměřeného na potraviny. Systém je navržen tak, aby uživateli doporučoval produkty postupně, podle toho jak dlouho a jakým způsobem se systémem interaguje.

Nejprve jsou uživateli doručeny produkty na základě zvolených oblíbených kategorií. Následně systém zaznamenává aktivitu uživatele (prohlížení a hodnocení produktů) a pomocí algoritmu TF-IDF doporučí nejrelevantnější produkty k prohlíženým. Dále pak jsou uživateli doporučené podobné produkty na základě již hodnocených produktů pomocí algoritmu SVD. Tyto dvě skupiny doporučených produktů jsou zkombinovány do přístupů, ze kterých je nejlepší přístup vybrán jako vstup pro konečný expertní systém. Další aktivita zaznamenávaná systémem je historie nákupů uživatele. Podle kterých jsou uživateli doručeny podobné produkty pomocí upraveného algoritmu TF-IDF. Finálním řešením doporučení je modifikace pomocí expertního systému využívající získaná data z nejlepšího přístupu, hodnocení produktů a koeficientu podobnosti k historii nákupů. Produkty získávají hodnotu doporučení, podle které jsou seřazeny a doporučeny uživateli.

Doporučovací systém byl postupně testován uživateli. Uživatelé hodnotili systém jako takový, doručení relevantních položek a porovnání s internetovým obchodem kosik.cz.

Výsledky jednotlivých testování dopadly ve většině případů pozitivně. Uživatelé hodnotili navržený systém jako intuitivní doručující relevantní výsledky. Horšími výsledky disponoval pouze přístup 3, který byl modifikován a upraven do přístupu 4, jehož výsledky dopadly dobře.

**Možnosti rozšíření práce**

Možné rozšíření práce je možné hned při získávání informací o uživateli. Například pomocí aktivity uživatele na e-shopu a jeho hodnocení jednotlivých produktů vytvořit profil uživatele s oblíbenými a neoblíbenými kategoriemi. Systém by tyto typy kategorií rozpoznal sám a nemusel by uživatele žádat o zvolení oblíbených kategorií.

Další možností je modifikace systému pro využití kombinace přístupu 1 (pevný procentuální podíl kombinace algoritmů TF-IDF a SVD) a přístupu 2 (procentuální podíl kombinace algoritmů TF-IDF a SVD na základě hodnocených produktů). Zvolení přepínacího hybridního systému se zjištěním v jaké situaci je vhodné použít výstup přístupu 1 a v jaké situaci je vhodné použití přístupu 2.

Další možností je také upravení webového rozhraní. Zvolení vhodných spotů na jednotlivých stránkách (hlavní stránka, stránka s produkty, detail produktu) pro výsledné seznamy doporučených produktů.

RESUMÉ

V první kapitole této diplomové práce je analýza současného stavu. Popisují se zde základní pojmy doporučovacích systému a možné přístupy k doporučování. Algoritmy využívané v doporučovacích systémech a jejich výhody či nevýhody. V závěru této kapitoly je zhodnocení možných přístupů k doporučování.

Druhá kapitola se věnuje návrhu doporučovacího systému. Nejprve definuje vstupní parametry do expertního systému. Popisuje jednotlivé části systému, kterými uživatel prochází. Dále popisuje použité algoritmy a jejich modifikace využité v této práci. V další části této kapitoly jsou popsány přístupy k doporučování produktů. Závěr této kapitoly se věnuje samotnému expertnímu systému, který využívá vstupní parametry pro doporučení produktů.

Ve třetí kapitole jsou popsány využité technologie využité v této práci. Kapitola také popisuje webové rozhraní aplikace - uživatelskou část a administrační část.

Poslední kapitola se zabývá samotnému testování na reálných uživatelích a analýze získaných dat. Také se věnuje otázkám z dotazníků týkajících se testování. Závěrem této kapitoly je zhodnocení testování.

SUMMARY

The first chapter of this thesis contains an analysis of the current state. There are described basic concepts of recommendation systems and possible approaches to recommendations. Algorithms used in recommendation systems and their advantages or disadvantages. In final part of this chapter is an evaluation of possible approaches to recommendations.

The second chapter focuses on design of a recommendation system. At first, it defines the input parameters to the expert system. Describes the various parts of the system that the user goes through. It also describes the used algorithms and their modifications used in this thesis. The next section of this chapter describes product recommendation approaches. The conclusion of this chapter deals with the expert system itself, which uses input parameters for product recommendations.

The third chapter describes the technologies used in this thesis and also the web interface of the application - the user part and the administration part.

The last chapter consists the testing on real users and analysis of the obtained data. It also focuses on questions from questionnaires used in testing. The conclusion of this chapter is the evaluation of testing.

SEZNAM POUŽITÉ LITERATURY

1. FALK, Kim. *Practical Recommender Systems*. First edition. Shelter Island: Manning, 2019. ISBN 1-61729-270-2.

2. Robin Van Meteren , Maarten Van Someren, Using Content-Based Filtering for Recommendation <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.25.5743&rep=rep1&type=pdf>

3. Matthew J. Lavin, "Analyzing Documents with TF-IDF," The Programming Historian 8 (2019), zdroj: <https://doi.org/10.46430/phen0082>.

4. Kevin Luk, Introduction to TWO approaches of Content-based Recommendation System , publikováno 2019, [cit. 9. 9. 2020], zdroj: <https://towardsdatascience.com/introduction-to-two-approaches-of-content-based-recommendation-system-fc797460c18c>

5. DVOŘÁK, Jiří, *Expertní systémy*. Brno: Vysoké učení technické v Brně, 2004

6. CASTAGNETTO, Jesus. PHP: programujeme profesionálně. Vyd. 1. Přeložil Ludvík Roubíček. Praha: Computer Press, 2001, xxiv,656 s. ISBN 80-7226-310-2

7. DUBOIS, Paul. *MySQL*. Fourth edition. Developer's Library, 208. ISBN 9780132704649

8. KOFLER, Michael a Bernd ÖGGL. PHP 5 a MySQL 5: průvodce webového programátora. Vyd. 1. Brno: Computer Press, 2007, 607 s. ISBN 978-80-251-1813-9

9. Proč používat Nette ?. Nette.org [online]. 2008-2021 [cit. 12. 1. 2021]. Dostupné z: <https://doc.nette.org/cs/3.1/why-use-nette>

10. Various Implementations of Collaborative Filtering. Towardsdatascience.com [online]. 28. 12. 2017 [cit. 3. 2. 2021]. Dostupné z:

<https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0>

11. KUMAR, Vaibhav. Singular Value Decomposition and Its Application In Recommender Systém. Analyticsindiamag.com [online]. 25. 3. 2020 [cit. 5. 2. 2021] <https://analyticsindiamag.com/singular-value-decomposition-svd-application-recommender-system/>

12. YANAI, Haruo, TAKEUCHI, Kei, TAKANE Yoshino. *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. Statistics for Social and Behavioral Sciences. Springer, New York, NY, 2011. ISBN: 978-1-4419-9887-3.

13. WALEK, Bogdan. FOJTÍK, Vladimír. *A hybrid recommender system for recommending relevant movies using and expert system.* Expert Systems With Appliacitons. Katedra informatiky a počítačů, Ostravská Univerzita, Ostrava, Česká Republika.

SEZNAM POUŽITÝCH SYMBOLŮ

|  |  |  |
| --- | --- | --- |
| MyISAM |  | My Indexed Sequential Access Method |
| ES |  | Expertní systém |
|  |  |  |
|  |  |  |

SEZNAM OBRÁZKŮ

[Obrázek 1: Doporučení produktů na stránce Netflix (zdroj [1]) 13](#_Toc71467598)

[Obrázek 2:Kolaborativní filtrování 15](file:///C:\Users\Fajmy\Desktop\Škola\Mgr\Diplomka\Fajmon_dipl.docx#_Toc71467599)

[Obrázek 3: Problém prázdnoty (zdroj:[1]) 17](file:///C:\Users\Fajmy\Desktop\Škola\Mgr\Diplomka\Fajmon_dipl.docx#_Toc71467600)

[Obrázek 4: Maticový rozklad SVD (zdroj: vlastní) 18](#_Toc71467601)

[Obrázek 5: Monolitický hybridní doporučovací systém (zdroj: [1]) 21](file:///C:\Users\Fajmy\Desktop\Škola\Mgr\Diplomka\Fajmon_dipl.docx#_Toc71467602)

[Obrázek 6: Smíšený hybridní doporučovací systém (zdroj: [1]) 21](file:///C:\Users\Fajmy\Desktop\Škola\Mgr\Diplomka\Fajmon_dipl.docx#_Toc71467603)

[Obrázek 7: Kombinovaný doporučovací systém (zdroj: [1]) 22](#_Toc71467604)

[Obrázek 8: Přepínací kombinovaný doporučovací systém (zdroj: [1]) 22](#_Toc71467605)

[Obrázek 9: Váhový kombinovaný doporučovací systém (zdroj: [1]) 22](#_Toc71467606)

[Obrázek 10: Architektura navrhovaného systému (zdroj: vlastní) 25](#_Toc71467607)

[Obrázek 11: Návrh systému (zdroj: vlastní) 34](#_Toc71467608)

[Obrázek 12: Diagram doporučení podle oblíbených kategorií (zdroj: vlastní) 35](#_Toc71467609)

[Obrázek 13: Nastavení oblíbených kategorií (zdroj: testovací e-shop) 35](#_Toc71467610)

[Obrázek 14: Hlavní stránka s doporučenými produkty (zdroj: testovací e-shop) 36](#_Toc71467611)

[Obrázek 15: Detail produktu (zdroj: testovací e-shop) 37](#_Toc71467612)

[Obrázek 16: Diagram doporučení na základě prohlížení 38](#_Toc71467613)

[Obrázek 17: Příklad prohlížených produktů (zdroj: testovací e-shop) 38](#_Toc71467614)

[Obrázek 18: Výsledky doporučení na základě prohlížení (zdroj: testovací e-shop) 39](#_Toc71467615)

[Obrázek 19: Diagram doporučení podle hodnocení produktů (zdroj: vlastní) 40](#_Toc71467616)

[Obrázek 20: Příklad hodnocených produktů (zdroj: testovací e-shop) 41](#_Toc71467617)

[Obrázek 21: Výsledky doporučení podle hodnocených produktů (zdroj: testovací e-shop) 42](#_Toc71467618)

[Obrázek 22: Diagram doporučení podle historie nákupů (zdroj: vlastní) 43](#_Toc71467619)

[Obrázek 23: Výstupní proměnná expertního systému 52](#_Toc71467620)

[Obrázek 24: Administrativní část - seznam uživatelů (zdroj: testovací e-shop) 59](#_Toc71467621)

[Obrázek 25: Oblíbené kategorie vybraného uživatele 60](#_Toc71467622)

[Obrázek 26: Podobní uživatelé k vybranému uživateli (zdroj: testovací e-shop) 60](#_Toc71467623)

[Obrázek 27: Historie nákupu vybraného uživatele (zdroj: testovací e-shop) 60](#_Toc71467624)

[Obrázek 28: Prohlížené produkty vybraného uživatele (zdroj: testovací e-shop) 61](#_Toc71467625)

[Obrázek 29: Detail výsledků vybraného přístupu (zdroj: testovací e-shop) 61](#_Toc71467626)

[Obrázek 30: Uživatelský panel (zdroj: testovací e-shop) 62](#_Toc71467627)

[Obrázek 31:Doručení produktů u přístupu 1 a přístupu 2 (zdroj: testovací e-shop) 63](#_Toc71467628)

[Obrázek 32: Doručení produktů u přístupu 4 (zdroj: testovací e-shop) 64](#_Toc71467629)

[Obrázek 33: Doporučené produkty pomocí expertního systému (zdroj: testovací e-shop) 65](#_Toc71467630)

[Obrázek 34: Dotazník - První testování, první otázka 73](#_Toc71467631)

[Obrázek 35: Dotazník - První testování, druhá otázka 74](#_Toc71467632)

[Obrázek 36: Dotazník - První testování, třetí otázka 75](#_Toc71467633)

[Obrázek 37: Dotazník - První testování, čtvrtá otázka 75](#_Toc71467634)

[Obrázek 38: Dotazník - První testování, pátá otázka 76](#_Toc71467635)

[Obrázek 39: Dotazník - Druhé testování, první otázka 81](#_Toc71467636)

[Obrázek 40: Dotazník - Druhé testování, druhá otázka 82](#_Toc71467637)

[Obrázek 41: Dotazník - Druhé testování, třetí otázka 82](#_Toc71467638)

[Obrázek 42: Dotazník - Druhé testování, čtvrtá otázka 83](#_Toc71467639)

[Obrázek 43: Dotazník - Druhé testování, pátá otázka 83](#_Toc71467640)

[Obrázek 44: Dotazník - Druhé testování, šestá otázka 84](#_Toc71467641)

[Obrázek 45: Dotazník - Druhé testování, sedmá otázka 84](#_Toc71467642)

[Obrázek 46: Dotazník - čtvrté testování, první otázka 88](#_Toc71467643)

[Obrázek 47: Dotazník - čtvrté testování, druhá otázka 89](#_Toc71467644)

SEZNAM TABULEK

[Tabulka 1: Hodnocené produkty aktivního uživatele 30](#_Toc71467645)

[Tabulka 2: Podobné produkty k hodnoceným produktům aktivního uživatele 30](#_Toc71467646)

[Tabulka 3: Příklad výstupu přístupu 4 48](#_Toc71467647)

[Tabulka 4: Vstupní proměnné expertního systému 50](#_Toc71467648)

[Tabulka 5: Rozložení fuzzy množin vstupních proměnných expertního systému 51](#_Toc71467649)

[Tabulka 6: Seznam několika pravidel expertního systému 53](#_Toc71467650)

[Tabulka 7: Testování expertního systému 54](#_Toc71467651)

[Tabulka 8: Výpočet hodnoty doporučení 55](#_Toc71467652)

[Tabulka 9: První testování - Přístup 1, Skupina A 69](#_Toc71467653)

[Tabulka 10: První testování - Přístup 2, Skupina A 70](#_Toc71467654)

[Tabulka 11: První testování - Přístup 3 70](#_Toc71467655)

[Tabulka 12: První testování - Přístup 3, noví uživatelé 71](#_Toc71467656)

[Tabulka 13: První testování - Přístup 4, Skupina A 72](#_Toc71467657)

[Tabulka 14: První testování - Přístup 4, Skupina B 72](#_Toc71467658)

[Tabulka 15: Druhé testování - Přístup 1, Skupina A 77](#_Toc71467659)

[Tabulka 16: Druhé testování - Přístup 1, skupina B 77](#_Toc71467660)

[Tabulka 17: Druhé testování - Přístup 1, Skupina C 78](#_Toc71467661)

[Tabulka 18: Druhé testování - Přístup 2, Skupina A 78](#_Toc71467662)

[Tabulka 19: Druhé testování - Přístup 2, Skupina B 79](#_Toc71467663)

[Tabulka 20: Druhé testování - Přístup 2, Skupina C 79](#_Toc71467664)

[Tabulka 21: Druhé testování - Přístup 2, Skupina D 80](#_Toc71467665)

[Tabulka 22: Druhé testování - Přístup 4, Skupina A 80](#_Toc71467666)

[Tabulka 23: Druhé testování - Přístup 4, Skupina B 81](#_Toc71467667)

[Tabulka 24: Třetí testování - testování ES 85](#_Toc71467668)

[Tabulka 25: Čtvrté testování - nákup v e-shopu www.kosik.cz 87](#_Toc71467669)

[Tabulka 26: Čtvrté testování - nákup v testovacím e-shopu 87](#_Toc71467670)

SEZNAM PŘÍLOH

Výsledky testování všech uživatelů v souboru Microsoft Excel

Seznam uživatelů s přístupovými loginy v souboru Microsoft Excel

Znalostní báze expertního systému ve formátu RB

Přehled všech IF-THEN pravidel v souboru pdf

První testování - instrukce ve formátu PDF

Druhé testování - instrukce ve formátu PDF

Třetí testování (ES) - instrukce ve formátu PDF

Čtvrté testování (www.kosik.cz) - instrukce ve formátu PDF

1. https://github.com/d3veloper/SVD [↑](#footnote-ref-1)
2. https://github.com/sciphp/numphp [↑](#footnote-ref-2)