

OSTRAVSKÁ UNIVERZITA  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA INFORMATIKY A POČÍTAČŮ

Adaptivní systém pro automatizovaný  
návrh doporučených produktů v  
internetovém obchodě

DIPLOMOVÁ PRÁCE

Autor práce: 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: Petr Fajmon  
Supervisor: RNDr. Bogdan Walek, Ph.D.

2020

(Zadání vysokoškolské kvalifikační práce)

## **ABSTRAKT**

Český text abstraktu

*Klíčová slova:*

*(klíčová slova vypsaná na řádku, oddělená od sebe čárkami)*

## **ABSTRACT**

The text of the abstract.

*Keywords:*

## 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/ky

## PODĚKOVÁNÍ

Rád bych poděkoval panu RNDr. Bogdanu Walkovi, Ph.D. za pomoc při vytváření této diplomové práce, za odbornou pomoc a konzultace.

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

<b>ÚVOD.....</b>	<b>10</b>
<b>CÍL PRÁCE.....</b>	<b>11</b>
<b>STRUKTURA PRÁCE.....</b>	<b>11</b>
<b>1 DOPORUČOVACÍ SYSTÉM .....</b>	<b>12</b>
<b>2 MOŽNÉ PŘÍSTUPY .....</b>	<b>15</b>
2.1 Kolaborativní filtrování.....	15
2.1.1 Problémy kolaborativního filtrování.....	16
2.1.2 Typy kolaborativního filtrování.....	17
2.2 Filtrování podle obsahu.....	18
2.3 Hybridní přístup .....	19
2.4 Závěr přístupů .....	21
<b>3 NÁVRH.....</b>	<b>23</b>
3.1 Architektura.....	23
3.2 Registrace .....	23
3.3 Historie prohlížení.....	24
3.4 Historie nákupů .....	24
3.5 Přístup pro personalizaci .....	25
<b>4 IMPLEMENTACE.....</b>	<b>28</b>
4.1 Použité technologie .....	28
4.1.1 PHP .....	28
4.2 Databáze .....	28
4.3 Řešení problému.....	29
<b>5 TESTOVÁNÍ.....</b>	<b>42</b>
5.1 Popis testování .....	42
5.2 Průběh .....	42
5.3 Výsledky .....	42
5.4 Závěr .....	42
<b>ZÁVĚR .....</b>	<b>43</b>
<b>RESUMÉ .....</b>	<b>44</b>
<b>SUMMARY .....</b>	<b>45</b>
<b>SEZNAM POUŽITÉ LITERATURY.....</b>	<b>46</b>
<b>SEZNAM POUŽITÝCH SYMBOLŮ.....</b>	<b>47</b>



<b>SEZNAM OBRÁZKŮ .....</b>	<b>48</b>
<b>SEZNAM TABULEK.....</b>	<b>49</b>
<b>SEZNAM PŘÍLOH.....</b>	<b>50</b>

## 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. Každý kdo si v internetovém obchodě něco vybírá, by chtěl, aby se mu ukázaly nejvhodnější produkty.

Tato práce se zabývá návrhem a implementací adaptivního algoritmu, který podle uživatelských dřívějších aktivit na webu bude vybírat nejvhodnější produkty.

## **CÍL PRÁCE**

Cílem práce je návrh a implementace adaptivního systému pro automatizovaný návrh doporučených produktů v internetovém obchodě. Součástí systému bude návrh doporučených produktů pro uživatele na základě informací získaných při registraci, předchozích realizovaných nákupů, preferencí uživatele získaných z předchozích nákupů a prohlížených produktů. Součástí systému bude systém pro podporu rozhodování z výše zmíněných vstupních dat, který nabídne uživateli doporučené produkty.

## **STRUKTURA PRÁCE**

První část práce obsahuje analýza současného stavu, návrh systému. V druhé části se práce věnuje implementaci systému a následnému testování.

# 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.“

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 popsany proces doporučení na streamovacím webu Netflix.

(obrázek)

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 systém doporučovat hudební skladby, není velký problém špatně doručená skladba. Zatímco velkou složitost bude mít systém, který rozhoduje v celoživotních situacích jako například doporučení pěstounů.

## **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 vlastníkovu zisk. Například z prémiových účtů s měsíčním paušálem pro neomezené možnosti na stránkách.

## **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.

## **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).

## **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.

## **Soukromí a důvěryhodnost**

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

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 zákaznickovy doporučovány tak, aby měl pocit, že není manipulován.

## **Rozhraní**

Jaké budou vstupy a výstupy doporučovacího systému? Vstupem může být manuální vkládání dat o uživateli (explicitní) nebo automatický, 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).

## **Algoritmy**

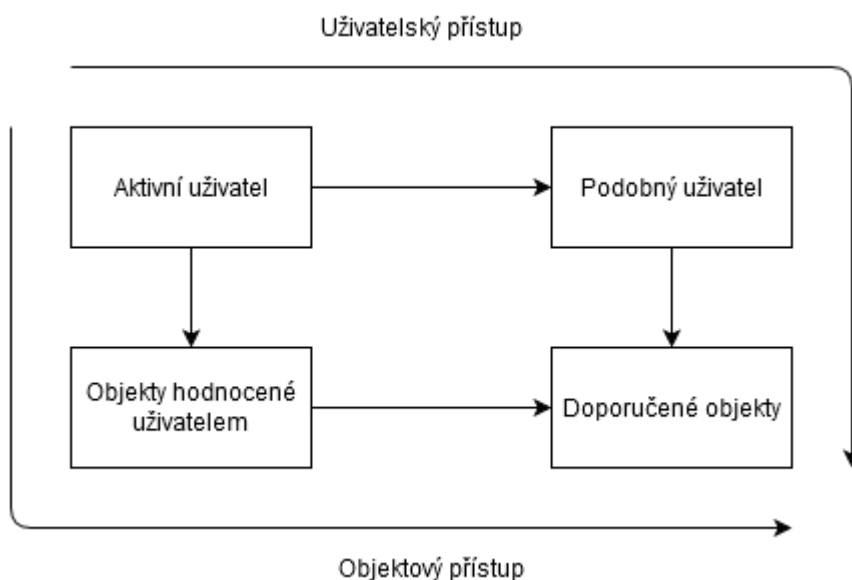
Jaký algoritmus bude systém používat? Algoritmy využívající se pro doporučování se dělí na kolaborativní filtrování, filtrování podle obsahu a hybridní systémy, které kombinují již zmíněné skupiny. Tyto přístupy jsou popsány níže.

## 2 MOŽNÉ PŘÍSTUPY

Text

### 2.1 Kolaborativní filtrování

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



Kolaborativní filtrování založené na podobnosti uživatelů využívá dvě možnosti doručení objektu. První možnost využívá nalezení podobného 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. [Practical Recommended - book]

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. [Practical Recommended - book]

### **2.1.1 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 studenému startu je problém šedé ovce (Gray sheep), kdy uživatel má specifické preference natolik, že je těžké vytvořit analýzu na základě podobných uživatelů.

#### **Studený produkt (Cold product)**

Nová položka, přidána 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.

#### **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.

#### **Šedá ovce (Gray sheep)**

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

#### **Řídkost (Sparsity problem)**

Uživatelé a produkty jsou reprezentovány user-item maticí. Řádky matice jsou uživatelé a sloupce produkty, buňky matice znamenají jednotlivé hodnocení uživatele na daný produkt. Většina obchodů však 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.

		Produkty									
Uživatelé	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										
	11										
	12										

### 2.1.2 Typy kolaborativního filtrování

#### Memory-based

Tento přístup se rozděluje na dvě hlavní sekce, tím je user-item filtrování a item-item filtrování. User-item filtrování bere aktivního uživatele, najde podobné uživatele a na základě podobnosti hodnocení doručí uživateli produkty, které podobní uživatelé hodnotili pozitivně. Na druhou stranu item-item filtrování bere produkt, najde uživatele, kteří tento produkt hodnotili pozitivně, a najde podobné produkty, které uživatel nebo podobní uživatelé hodnotili pozitivně. Pro výpočet takové podobnosti se využívá kosinová podobnost (cosine similarity) nebo Pearsonův korelační koeficient (Pearson correlation coefficient).

#### Model-based

Tento přístup využívá pro vývoj modelů machine learning algoritmy k získání hodnot neohodnocených produktů uživatele. Snaží se zaplnit matici chybějícími daty. Takové modely jsou například Bayesianovy sítě, Singular Value Decomposition (SVD) a pravděpodobnostní latentní sémantické modely.

## 2.2 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. [Practical Recommended - book]

Filtrování podle obsahu vybírá objekty podle podobnosti obsahu objektu a preferencích uživatele, na rozdíl od kolaborativního filtrování, které vybírá objekty podle podobnosti uživatelů. [Robin van Meteren<sup>1</sup> and Maarten van Someren]

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

Pro vytvoření fungujícího algoritmu je potřeba vymezit části algoritmu.

- **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.

[Practical Recommended - book]

### 2.2.1 Typy filtrování podle obsahu

#### **Content only**

Tento přístup doporučí uživateli podobné produkty již dříve hodnoceným, nebo zakoupeným. 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.

#### **User profile**

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

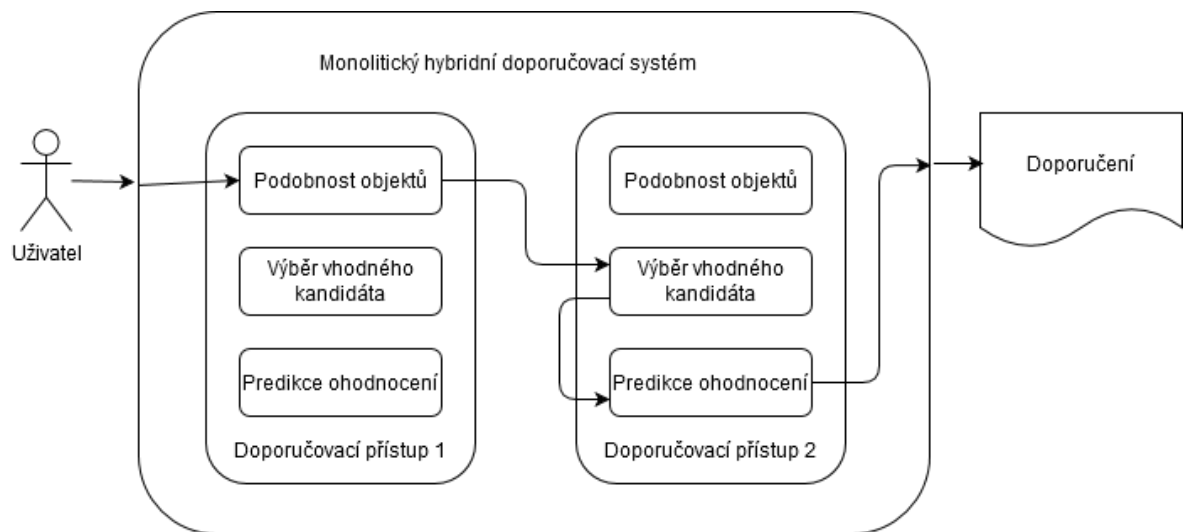
## 2.3 Hybridní přístup

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

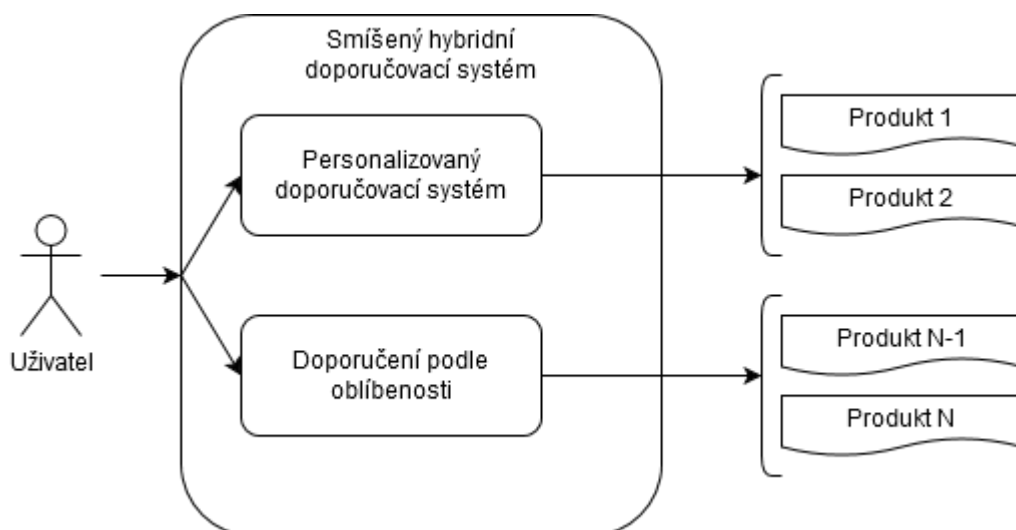
### Monolitické

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. [Practical Recommended - book]



### Smíšené

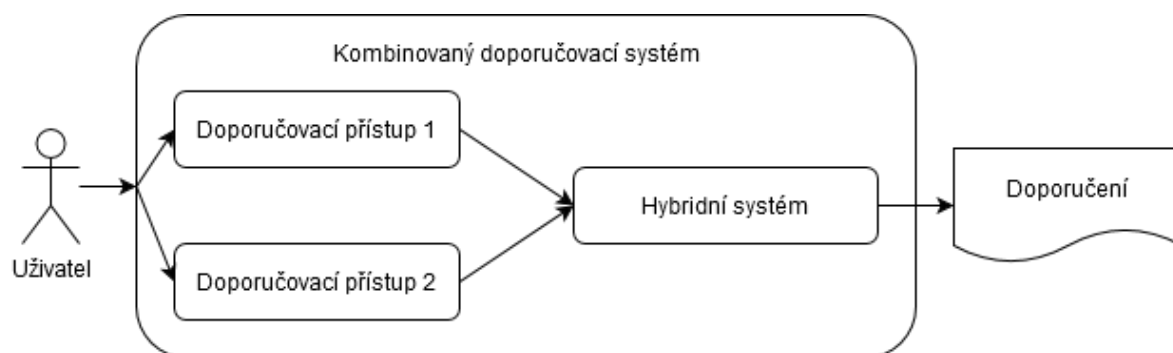
Smíšené doporučovací systémy doplňují výsledky jednoho vybraného algoritmu dalšími výsledky z jiného vybraného algoritmu. Například pokud algoritmus doporučující personalizované produkty uživateli doručí N výsledků, tak je doplněn do M požadovaných výsledků jiným algoritmem doporučujícím například oblíbené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



## Kombinované

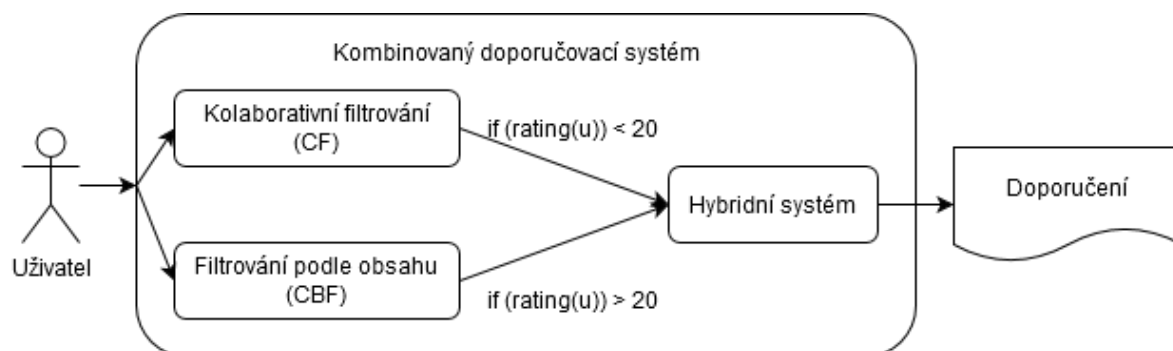
Rozdíl mezi kombinovaným přístupem a smíšeným hybridním přístupem je, že smíšený hybridní přístup vždy ukazuje vše z jednoho doporučovacího algoritmu doplněného jiným doporučovacím algoritmem. [Practical Recommended - book]

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. [Practical Recommended - book]

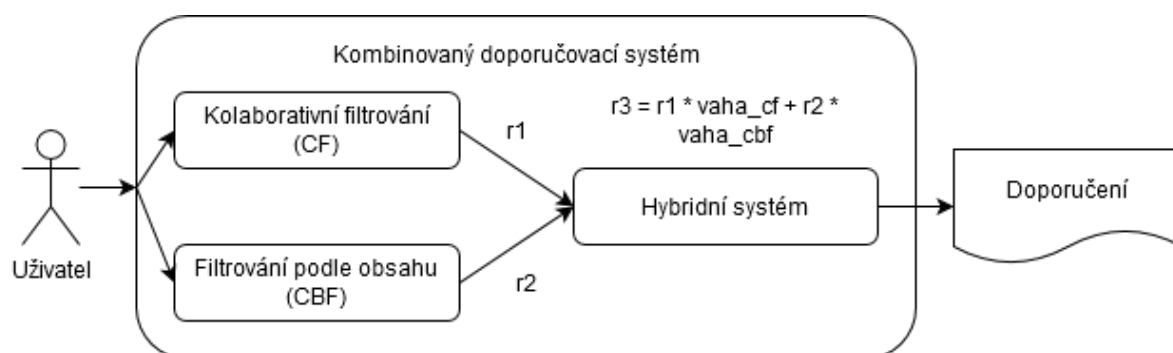


**Přepínací kombinovaný systém** přepíná mezi doporučovacími systémy na základě kontextu a dodává všechny výsledky jednoho doporučovacího systému. Například podle

země původu, ze které uživatel pochází, nebo podle denní doby. [Practical Recommended - book]



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



## 2.4 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 uživateli 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 obchodu.

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 s těmito problémy 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 obchodě, který má minimum dat a málo uživatelů.

### **Filtrování podle obsahu**

Váhodou 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.

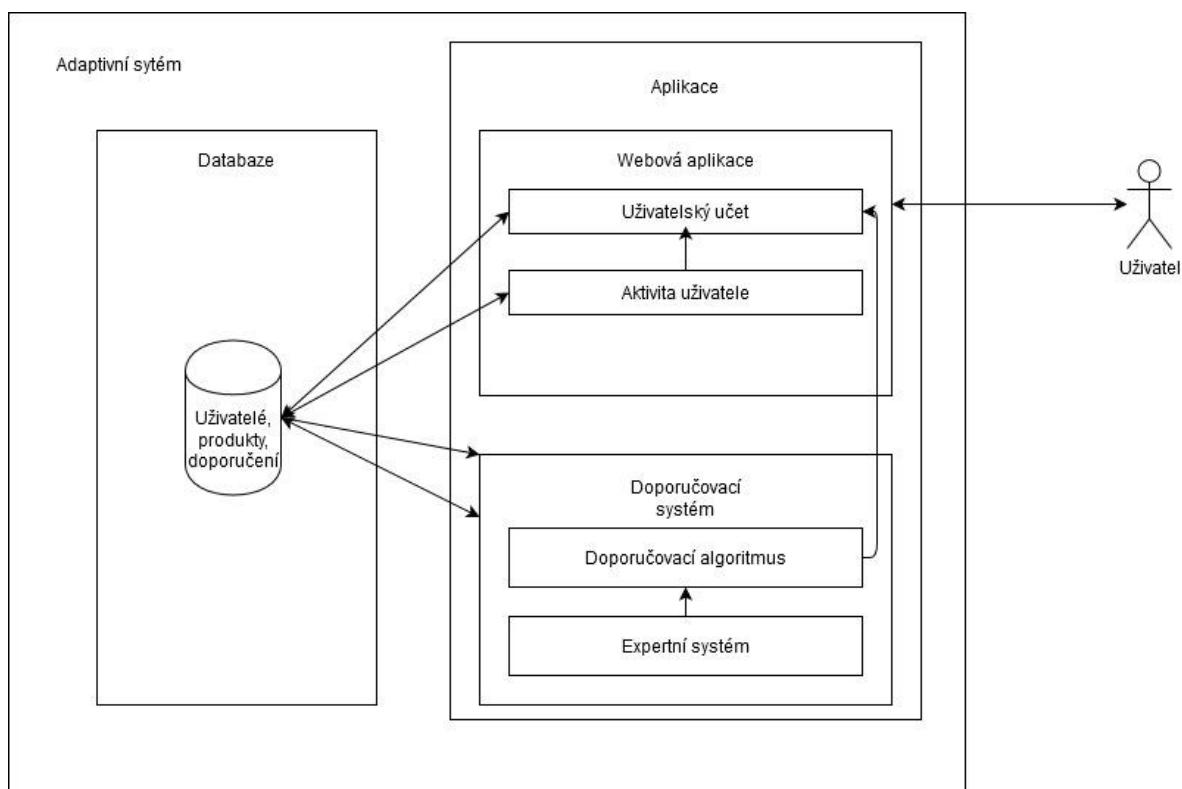
Filtrování podle obsahu se lépe využije na obchodech s menším množstvím dat a uživatelů, nebo na začínajících obchodech, kdy obchod má pouze několik produktů a minimum uživatelů. Dobré využití je také na doménách, které nepoužívají hodnocení, například informační portály nebo zpravodajství ([www.novinky.cz](http://www.novinky.cz),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

### **Hybridní systémy**

### 3 NÁVRH

Navržený adaptivní systém, bude zaměřen na doručení obsahu uživateli pomocí dat získaných z 3 vstupů. Systém bude využívat data získána od uživatele při vstupním nastavení při registraci. Nejprve bude pracovat s daty získaných od uživatelů s podobnými vstupními daty. Následně pomocí prohlížení jednotlivých produktů a historie nákupu bude uživateli doporučován personalizovaný obsah.

#### 3.1 Architektura



#### 3.2 Registrace

Uživatel při registraci zvolí oblíbené kategorie, podle kterých systém vyhodnotí prvotní doručení obsah na základě těchto kategorií. Na základě prohlížení jiných uživatelů s podobnou kategorií bude vybrán doručení obsah uživateli.

### 3.3 Historie prohlížení

Uživateli bude doručován obsah také na historii prohlížení produktů. Je potřeba si uvědomit jaké data budou 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.

*Délka návštěvy produktu* může napovědět, jestli má uživatel o daný produkt zájem. Musíme brát v potaz fakt, že uživatel by při prohlížení produktů neměl být vyrušován vnějšími faktory a ukončit délku prohlížení kliknutím na další odkaz. V následující tabulce je znázorněna délka návštěvy produktu s jeho významem.

Délka návštěvy produktu	Význam
Méně než 5 sekund	Uživatel se nezajímá o produkt
Více než 5 sekund	Uživatel se zajímá o produkt
Více než minuta	Uživatel se zajímá o produkt velmi
Více než 5 minut	Pravděpodobně dělá něco jiného
Více než 10 minut	Opustil produkt bez překliknutí na jiný

*Rozšířené kliky* jsou další možnosti 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 se uživateli mohou doručovat produkty. Aby si uživatel vytvořil takový seznam, může být nazván „seznam přání“, nebo „seznam oblíbených produktů“.

### 3.4 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. Uživateli bude doručen podobný produkt na základě content-based algoritmu. Podle výpočtu podobnosti produktů pomocí TF-IDF vektoru.

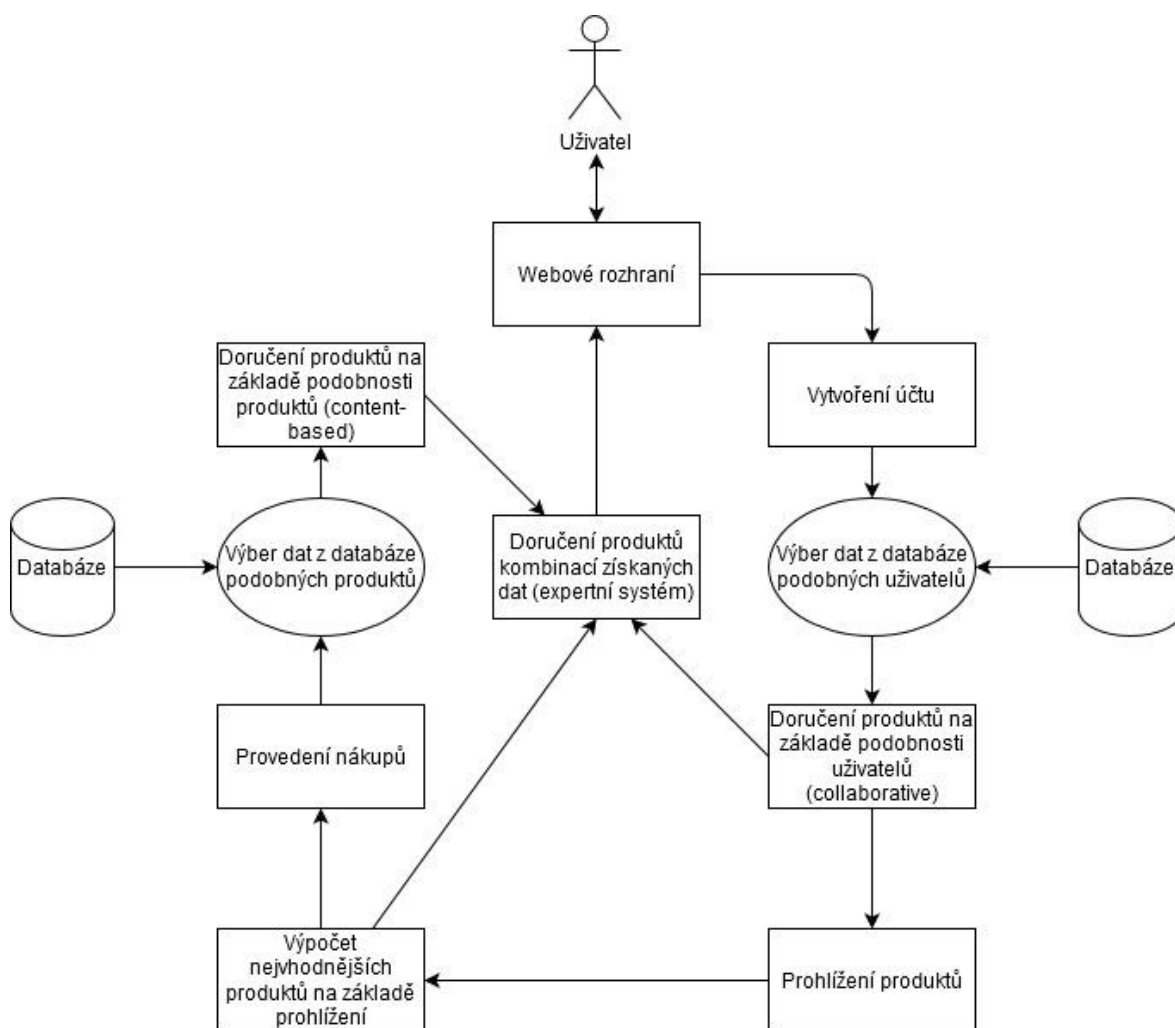


*Produkty* budou nejdůležitější složkou pro výpočet doručení podobných produktů. Bude důležité, jaké produkty si uživatel zakoupil a také jak často, nebo v jakém množství. Pokud například uživatel nakupuje mléčné výrobky především jogurty, budou mu doručeny podobné produkty s těmito parametry.

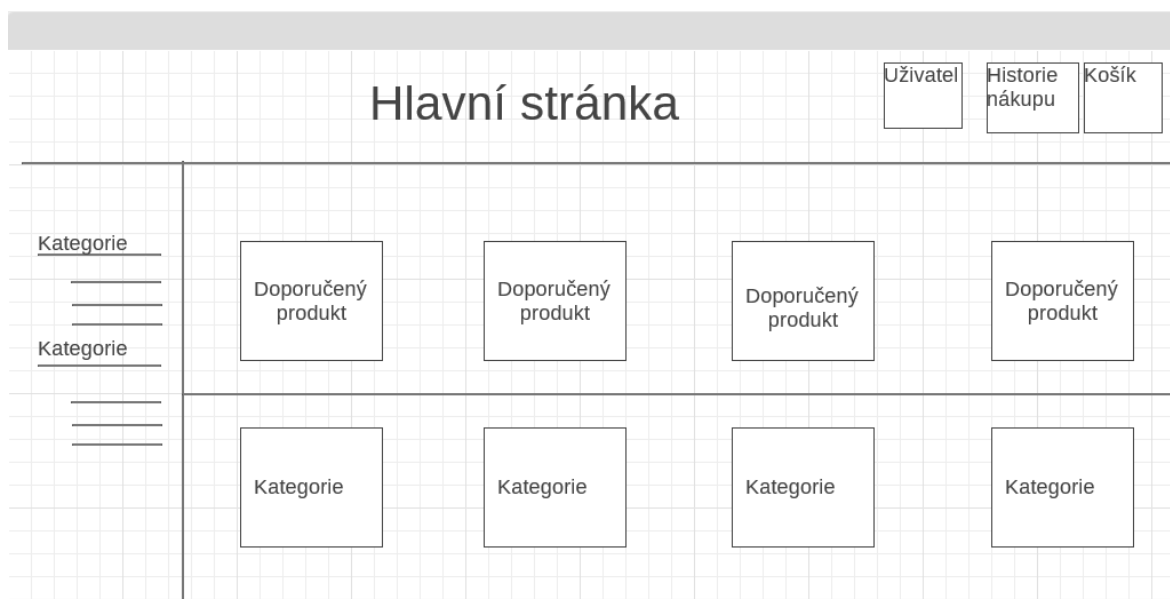
*Uskutečnění nákupu* bude určovat váhu zakoupených produktů pro výpočet. Pokud si uživatel daný produkt nezakoupil za posledních pár nákupů, může to znamenat, že o daný produkt nemá zájem. Takový produkt s danými parametry tedy nebude brán do hodnocení pro podobnost produktů.

### 3.5 Přístup pro personalizaci

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



Navržený systém je automatizovaný, kdy jsou uživatelům doručovány produkty na základě jeho oblíbených kategorií, chování na webu a podobnosti s jinými uživateli. Uživatel při se prvním kontaktu se systémem zaregistruje a zvolí oblíbené kategorie. Systém vybere jiné uživatele podobné aktuálnímu uživateli podle kategorií. Uživatelům jsou při první návštěvě hlavní stránky automatizovaně doručeny první produkty, které odpovídají produktům zvolených podobnými uživateli.



Dalším aspektem automatizovaného doručování je chování uživatele na webu. Jak je uvedeno výše, uživatel se může produkty pouze “proklikávat”, proto je důležité stanovit intervaly, podle kterých bude systém zaznamenávat chování uživatele na webu a doručovat mu materiál, který uživatele zajímá. Pokud systém zaznamená uživatelský zájem o nějaký produkt, bude mu ho následně doručovat spolu s produkty oblíbených kategorií. Produkty z oblíbených kategorií budou mít větší váhu při doručení, než produkty, které do oblíbených kategorií nespádají.

Uživatel

Historie nákupu

Košík

Detail produktu

Kategorie

Kategorie

Produkt

Název produktu

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis bibendum, lectus ut viverra rhoncus, dolor nunc faucibus libero, eget facilisis enim ipsum id lacus. Praesent id justo in neque elementum ultrices. Etiam dui sem, fermentum vitae, sagittis id, malesuada in, quam. Pellentesque pretium lectus id turpis. Nullam rhoncus aliquam metus. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Přidat do košíku

Doporučený produkt

Doporučený produkt

Doporučený produkt

Doporučený produkt

Posledním vstupním parametrem bude historie nákupu uživatele. Pokud si uživatel daný produkt zakoupil, mohl by mít zájem i o produkty podobného typu. Systém tedy doručí takové produkty, které mají podobné parametry jako zakoupený produkt. Produkty zakoupené uživatelem nejpozději, budou mít největší prioritu, pokud se tedy nějaký produkt v historii nákupů nebude dlouho objevovat, uživatel o něho mohl mít přestat mít zájem a systém tedy takový produkt vyřadí.

Výsledným řešením bude kombinovaný přístup k doručování pomocí expertního systému. Data získána z výše popsanych akcí budou vloženy do expertního systému a z nich vybrány nejvhodnější produkty. Vstupní proměnné budou hodnota získána z výsledků podobnosti uživatelů, hodnota získána z výpočtu chování uživatele na webu a hodnota získána z podobnosti produktů. Pomocí IF-THEN pravidel bude vypočítána výstupní proměnná. Výstupní proměnná bude nejvhodnější doručení uživateli.

## **4 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. V poslední části této kapitoly je popsáno řešení navrženého systému.

### **4.1 PHP**

PHP (zkratka pro PHP Hypertext Preprocessor) je skriptovací programovací jazyk zabudovaný na straně serveru, všechny úkony jsou prováděny na hostingu, tudíž není zatíženo uživatelské 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, PostgreSQL, Oracle). Díky své jednoduchosti a bohaté zásobě funkcí se stal jazyk oblíbeným. V PHP jsou napsány například Wikipedie nebo Facebook.

### **4.2 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.

### **4.3 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 rychlosti díky omezením, které ostatní databázové systémy nemají. Zkratka znamená „My Structured Query Language“ v překladu

system pro řízení databází. Do databáze lze ukládat různá data (text, obrázky, hodnoty atd.) s nimiž se dá jednoduše pracovat (třídít, řadit, filtrovat atd.). System se nejčastěji používá v kombinaci s programovacím jazykem PHP, které 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.

MyISAM formát je využívám především pro velké množství čtením 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 jednoduchá, jelikož se záznamy pouze zařadí na konec datového souboru. Operace mazání a aktualizace jsou však problematičtější, odstranění musí zanechat prázdné místo, jinak by se změnila posuny řádků. Hlavní nevýhodou MyISAM je absence transakcí a nejsou podporovány cizí klíče. MyISAM podporuje FULLTEXT indexování.

## **4.4 Řešení problému**

Aby mohl algoritmus doporučit produkty personalizované přímo pro aktivního uživatele, je nutné zvolit vhodný způsob. Při prvních návštěvách e-shopu o aktivním uživateli nevíme zhora nic. Nevíme, jaké jsou oblíbené kategorie aktivního uživatele, ani o jaké produkty má zájem. Tento problém řeší první algoritmus.

### **4.4.1 Algoritmus 1**

Uživatelův první kontakt je pro systém náročný, jelikož systém nemá o uživateli žádné informace a nemůže mu tedy doporučit žádný personalizovaný produkt. První algoritmus řeší tento problém vyžádáním uživatele o zvolení oblíbených kategorií. Podle zvolených kategorií je uživatel zařazen do skupiny uživatelů s podobnými kategoriemi. Na základě této skupiny jsou uživateli doručovány první personalizované produkty.

Skupiny:

**U** - seznam uživatelů podobných aktivnímu uživateli.

**P** - seznam top 100 produktů ze skupiny **U**.

**NP** - naposledy prohlížené produkty.

1. část algoritmu:

- 1) Uživatel se zaregistruje a zvolí si oblíbené kategorie.
- 2) Algoritmus aktivního uživatele zařadí do skupiny **U** podobných uživatelů na základě podobných kategorií.
  - a) Vytvoří se pole kategorií aktivního uživatele, pole kategorií ostatních uživatelů a je vypočtena podobnost, na základě počtu podobných prvků v poli.
  - b) 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, hodnocení produktu a vybere top 100 produktů.
- 4) Jestliže existuje skupina **P**, algoritmus vyfiltruje ze skupiny pouze produkty odpovídající uživatelským kategoriím. V opačném případě jsou aktivnímu uživateli doručeny všechny produkty z jeho kategorií.

V první části je vyřešeno zařazení uživatele do skupiny podobných uživatelů a doporučení produktů z této skupiny. Tyto produkty jsou však top produkty uživatelů, kteří jsou podobní aktivnímu uživateli. Produkty jsou zřídka personalizované aktivnímu uživateli. Aktivní uživatel si na e-shopu prohlíží produkty, které ho zajímají. Algoritmus zaznamenává, které produkty aktivní uživatel navštívil a ukládá data o prohlížení. Následně pomocí algoritmu TF-IDF jsou aktivnímu uživateli doporučeny podobné produkty k jeho prohlíženým, tudíž více personalizované.

Algoritmus TF-IDF je metodika hodnocení relevance při vyhledávání v textu. Každý produkt je reprezentován jak TF-IDF vektor. Název je spojením zkratk 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.

$$TF(\text{slovo}) = \frac{\text{Výskyt slova v aktuálním dokumentu}}{\text{Celkový počet slov v aktuálním dokumentu}}$$

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.

$$IDF(slovo) = \log_{10} * \left( \frac{\text{Celkový počet dokumentů}}{\text{Počet dokumentů obsahující dané slovo}} \right)$$

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.

$$TFIDF = TF * IDF$$

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, které se v popise vyskytují, ale nepřidávají mu žádný význam. Takovými slovy mohou být například předložky, spojky a další vybraná slova (např.: aby, který, jsem, zda).

Tento problém s využitím algoritmu TF-IDF řeší druhá část algoritmu.

## 2. část algoritmu

1) Uživatel navštíví produkt, tento produkt se uloží do skupiny **NP**.

- a) 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“.
- b) 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 seřazena podle několika kritérií. Nejdůležitější kritérium značí počet návštěv produktu, následně poslední návštěva a nakonec jsou produkty seřazeny podle hodnocení produktu uživatelem.

3) Pomocí algoritmu **TF-IDF** jsou ke skupině **NP** vybrány podobné produkty seřazené podle podobnosti.

a) Pro každý produkt je získán index [index], jeho popis a počet slov v popisu.

i) Pro každé slovo v popisu produktu je vypočten koeficient TF-IDF.

$$TFIDF[slovo] = TF * \log\left(\frac{docCount}{DF}\right)$$

ii) Pro popis je získána hodnota TF-IDF[index] jako součet TF-IDF vektoru všech slov v popise.

iii) Následně je vypočtena podobnost pro každý produkt ze skupiny **NP**.

$$Podobnost[index] = \frac{TFIDF[index]}{\text{počet slov v popisu}[index]}$$

iv) Celková podobnost je vypočtena jako součet podobností pro daný produkt.

$$Celková\ podobnost = celková\ podobnost + podobnost[index]$$

V druhé části jsou aktivnímu uživateli doručeny produkty, podobné prohlíženým produktům. Je využit algoritmus TF-IDF, jehož hlavním úkolem je hodnocení relevance při vyhledávání v textu. Z výše uvedených částí byly získány 2 skupiny produktů. První skupina obsahuje top produkty všech uživatelů, kteří patří do skupiny aktivního uživatele, vyfiltrované podle kategorií aktivního uživatele. Druhá skupina obsahuje top produkty podobné prohlíženým aktivním uživatelem.

### Příklad k Algoritmu 1

#### 1. Uživatel se zaregistruje a vybere si oblíbené kategorie

<input checked="" type="checkbox"/> Alkoholické nápoje	<input type="checkbox"/> Dětská výživa	<input type="checkbox"/> Houby
<input checked="" type="checkbox"/> Koření	<input type="checkbox"/> Maso	<input type="checkbox"/> Mléčné výrobky
<input checked="" type="checkbox"/> Nápoje	<input type="checkbox"/> Obilniny	<input type="checkbox"/> Ořechy
<input checked="" type="checkbox"/> Ovoce	<input type="checkbox"/> Pečivo	<input type="checkbox"/> Pivo
<input checked="" type="checkbox"/> Ryby	<input type="checkbox"/> Trvanlivé potraviny	<input type="checkbox"/> Zelenina
<input type="button" value="Upravit"/>		

Moje kategorie
1. Ryby
2. Ovoce
3. Nápoje
4. Koření
5. Alkoholické nápoje



Zde má aktivní uživatel navolené kategorie Alkoholické nápoje, Koření, Nápoje, Ovoce a Ryby. Podle uživatelových zvolených kategorií je zařazen do skupiny.

2. Algoritmus aktivního uživatele zařadí do skupiny podobných uživatelů.

Uživatel	Kategorie
1. Teodor	Ryby, Ořechy, Mléčné výrobky, Alkoholické nápoje, Nápoje, Ovoce,
2. Alfréd	Trvanlivé potraviny, Ořechy, Ryby, Alkoholické nápoje,

Ve skupině aktivního uživatele je Teodor a Alfréd. Teodor má společně 4 kategorie ze svých 6, jeho podobnost s aktivním uživatelem je tedy 66,6%. Alfréd má společně 2 kategorie ze svých 4 a jeho podobnost s aktivním uživatelem je tedy hraničních 50%.

3. Podle skupiny, do které je aktivní uživatel zařazen, je vytvořen seznam top 100 produktů této skupiny.

Název	Počet hodnocení	Hodnocení	Typ
1. Vlašské ořechy	25	4.9	Ořechy
2. Kokosové chipsy	25	4.8	Ořechy
3. Směs ořechů	25	4.8	Ořechy
4. Arašídý	25	4.2	Ořechy
5. Kešu	25	4.2	Ořechy
6. Pekanové ořechy	25	4.1	Ořechy
7. Para ořechy	25	4.0	Ořechy
8. Mandle	25	3.9	Ořechy
9. Mandle loupané	25	3.7	Ořechy
10. Lískové ořechy	25	3.5	Ořechy
11. Švestky	24	4.6	Ovoce
12. Jagermeister	22	4.8	Alkoholické nápoje
13. Pavolín Pálava	22	4.6	Alkoholické nápoje
14. Vaječný likér	22	4.6	Alkoholické nápoje
15. Becherovka	22	4.5	Alkoholické nápoje
16. Fernet	22	4.4	Alkoholické nápoje
17. Jack Daniels Honey	22	4.3	Alkoholické nápoje
18. Malibu	22	4.1	Alkoholické nápoje
19. Beefeater Gin Pink	22	3.8	Alkoholické nápoje
20. Tullamore Dew	22	3.8	Alkoholické nápoje

V obrázku výše, můžeme vidět vybraných top 20 produktů ze skupiny uživatelů aktivního uživatele. Produkty jsou seřazeny podle počtu hodnocení a následného hodnocení daného produktu. Můžeme vidět, že v seznamu jsou produkty, které nespádají do kategorií uživatele. K tomu slouží další krok, který vyfiltruje pouze produkty z kategorií uživatele.

4. Ze seznamu top produkty naší skupiny jsou vyfiltrovány pouze produkty, které odpovídají kategoriím uživatele.

Název	Počet hodnocení	Hodnocení	Typ
1. Švestky	24	4.6	Ovoce
2. Jagermeister	22	4.8	Alkoholické nápoje
3. Vaječný likér	22	4.6	Alkoholické nápoje
4. Pavolín Pálava	22	4.6	Alkoholické nápoje
5. Becherovka	22	4.5	Alkoholické nápoje
6. Fernet	22	4.4	Alkoholické nápoje
7. Jack Daniels Honey	22	4.3	Alkoholické nápoje
8. Malibu	22	4.1	Alkoholické nápoje
9. Tullamore Dew	22	3.8	Alkoholické nápoje
10. Beefeater Gin Pink	22	3.8	Alkoholické nápoje
11. Božkov Republica	22	3.7	Alkoholické nápoje
12. Slámové víno	22	3.7	Alkoholické nápoje
13. Absinth	22	3.6	Alkoholické nápoje
14. Whiskey	22	3.6	Alkoholické nápoje
15. Moët & Chandon Rosé	22	3.6	Alkoholické nápoje
16. Caleo Primitivo Salento	22	3.5	Alkoholické nápoje
17. Moët & Chandon Impérial	22	3.4	Alkoholické nápoje
18. Chardonnay Pays	22	3.1	Alkoholické nápoje
19. Rum	22	3.1	Alkoholické nápoje
20. Pellegrino Marsala	22	3.1	Alkoholické nápoje

Na obrázku lze nyní vidět, že jsou vybrány pouze produkty spadající do kategorií, které si uživatel vybral.

5. Uživatel se pohybuje na e-shopu a prohlíží si produkty.

Aby se využil potenciál prvního přístupu, je nutné, aby v seznamu TF-IDF bylo alespoň 30 produktů.

	Název	Navštíveno	Naposledy navštíveno	Moje hodnocení	Typ
1.	Kofola	3x	17.12.2020 08:31	4	Nápoje
2.	Kokosový ořech	2x	17.12.2020 08:30	nehodnoceno	Ovoce
3.	Losos s kůží	2x	17.12.2020 08:29	4	Ryby
4.	Pomeranč	2x	17.12.2020 08:27	5	Ovoce
5.	Kapr porce	1x	17.12.2020 08:30	nehodnoceno	Ryby
6.	Švestky	1x	17.12.2020 08:28	nehodnoceno	Ovoce
7.	Ananas	1x	17.12.2020 08:27	nehodnoceno	Ovoce
8.	Ledový čaj broskvev	1x	17.12.2020 08:27	nehodnoceno	Nápoje

V tabulce výše můžeme vidět, jaké produkty si uživatel prohlížel. Produkty jsou seřazeny podle počtu návštěv. Dále podle poslední návštěvy a nakonec zda uživatel produkt již hodnotil, anebo nehodnotil. Parametr hodnocení má vliv na následný výpočet podobných produktů, podle algoritmu TF-IDF, pokud uživatel daný produkt nehodnotil, tento produkt má poloviční vliv na výpočet doporučených produktů.

Prohlížených produktů je dostatek, aby bylo pomocí TF-IDF nalezeno alespoň 30 produktů.

6. Pomocí algoritmu TF-IDF jsou k prohlíženým produktům nalezeny podobné produkty.

















Název	Podobnost	Hodnocení	Typ
1. Filet s kůží	4.424882710254	4.6	Ryby
2. Ledový čaj citron	3.3775900661759	3.3	Nápoje
3. Kofola plech	1.9790065002738	3.9	Nápoje
4. Becherovka	1.7263394812486	4.5	Alkoholické nápoje
5. Grep	1.7239122722763	3.5	Ovoce
6. Kiví	1.6830546216917	4.2	Ovoce
7. Hruška	1.055048139204	3.6	Ovoce
8. Citron	1.0458072769412	4.9	Ovoce
9. Bluma	0.96938273441051	3.7	Ovoce
10. Halibut filet	0.96288750607591	4.3	Ryby
11. Treska filet	0.90357012153681	4.7	Ryby
12. Rohlík	0.86327371390909	3.9	Pečivo
13. Energy drink Semtex	0.82816919186921	3.7	Nápoje
14. Fazole	0.71077241818196	5	Trvanlivé potraviny
15. Mango	0.69224972202449	4.1	Ovoce
16. Staropramen nealko	0.67491000771599	4.1	Pivo
17. Coca-cola	0.66685958396135	4.8	Nápoje
18. Vepřové mleté	0.66477686804481	3.8	Maso
19. Červené fazole	0.64422127620619	4.2	Trvanlivé potraviny
20. Tilápie filet	0.62348075870128	4.5	Ryby
21. Fanta	0.61671668373926	4.8	Nápoje
22. Coca-cola plech	0.60338488962321	4.8	Nápoje
23. Žampiony Portobello	0.59969945454557	4.8	Houby
24. Hrozno	0.59304745995364	3.6	Ovoce
25. Pletenec	0.5895933262333	5	Pečivo
26. Energy drink Tiger	0.57160104449747	4.2	Nápoje
27. Březňák	0.57117255873352	3.8	Pivo
28. Vepřová bez kosti	0.55398072337067	3.4	Maso
29. 7 Up	0.52477397895703	3.8	Nápoje

V tabulce výše můžeme vidět všechny doporučené produkty pomocí algoritmu TF-IDF k prohlíženým produktům aktivního uživatele. Nejdůležitější je podobnost, která udává jak moc je daný produkt podobný k prohlíženým produktům. Dále je zde vidět celkové

hodnocení produktu a typ. V seznamu 29 produktů jsou pouze 3 produkty, které nepatří do kategorií aktivního uživatele. Můžeme zde vidět několik produktů, které nepatří do kategorií aktivního uživatele. Z 29 produktů je 6 produktů, které nevyhovují, nicméně většina je zařazena až v druhé části seznamu.

















Skupiny obou částí se prolínají a doporučují uživateli top produkty. Je vytvořeno několik přístupů, které kombinují tyto dvě skupiny. První přístup využívá procentuální poměr skupin, ze kterého se následně vytvoří seznam top produktů. Tento přístup je rozdělen do tří skupin, kde každá skupina znázorňuje procentuální poměr skupin. Skupina A má rozdělení 30% produktů z první části algoritmu (top produkty mé skupiny) a 70% z druhé části algoritmu (doporučené produkty podle TF-IDF). Skupina B má rozdělení 50% z první části algoritmu a 50% z druhé části algoritmu. Skupina C je opak skupiny A, tedy 70% z první části algoritmu a 30% z druhé části algoritmu.

Skupina A má poměr doporučených produktů 30% ze seznamu „TF-IDF“ a 70% ze seznamu „Moje top produkty“. Skupina A pro aktivního uživatele z výše uvedených nastavení a vytvořených seznamů vypadá následovně. Jedná se o náhled prvních 12 produktů ze skupiny A.

Skupina A	Skupina B	Skupina C	
			
Citron	Jagermeister	Treska filet	Filet s kůží
Relevantní	Relevantní	Relevantní	Relevantní
			
Pavolín Pálava	Vaječný likér	Švestky	Tilápie filet
Relevantní	Relevantní	Relevantní	Relevantní
			
Becherovka	Fernet	Kapr porce	Jack Daniels Honey
Relevantní	Relevantní	Relevantní	Relevantní
			
Halibut filet	Losos s kůží	Butterfish	Červené fazole
Relevantní	Relevantní	Relevantní	Nerelevantní

Na obrázku je pouze výběr top 12 produktů z možných 40 ze skupiny A. Můžeme vidět, že všechny produkty, až na jeden produkt, odpovídají uživatelským kategoriím a není třeba produkty označit za nerelevantní. Vzhledem k prohlíženým produktům však tato skupina nabízí alkoholické nápoje, které má aktivní uživatel v kategoriích, ale při své aktivitě na eshopu si žádný takový produkt neprohléhl.

















Skupina B má poměr z uvedených seznamů 50% na 50%. Opět je náhled na prvních 12 produktů skupiny B.

Skupina A	Skupina B	Skupina C	
			
Citron Relevantní	Coca-cola plech Relevantní	Jagermeister Relevantní	Coca-cola Relevantní
			
Treska fileť Relevantní	Pavolín Pálava Relevantní	Vaječný likér Relevantní	Švestky Relevantní
			
Fileť s kůží Relevantní	Tilápie fileť Relevantní	Becherovka Relevantní	Fernet Relevantní
			
Halibut fileť Relevantní	Jack Daniels Honey Relevantní	Červené fazole Nerelevantní	Losos s kůží Relevantní

Ve výběru top 12 produktů z možných 40 můžeme vidět, že se produkty lehce pozměnily. Do výběru se však dostaly i produkty z kategorie nápojů, které uživatel prohlížel a také je má zvolené jako oblíbené.

Skupina C má opačný poměr jako skupina A, tedy 70% ze seznamu „TF-IDF“ a 30% ze seznamu „Moje top produkty“. Tato skupina by měla doporučovat nejvíce personalizovaných produktů.



Skupina A	Skupina B	Skupina C	
			
Pletenec Nerelevantní	Citron Relevantní	Coca-cola plech Relevantní	Jagermeister Relevantní
			
Žampiony Portobello Nerelevantní	Fanta Relevantní	Coca-cola Relevantní	Treska fileť Relevantní
			
Pavolín Pálava Relevantní	Vaječný líkér Relevantní	Švestky Relevantní	Fileť s kůží Relevantní
			
Tilápie fileť Relevantní	Becherovka Relevantní	Fernet Relevantní	Jack Daniels Honey Relevantní

Ve skupině C se vyskytuje na prvním místě produkt z kategorie pečiva. Je to ovlivněno jeho hodnocením a oblíbeností všemi uživateli na eshopu. Tato skupina má obecně nejvíce nerelevantních produktů, to je však způsobeno algoritmem TF-IDF, který hledá podobné produkty na základě podobnosti obsahu.

Další přístup vytváří skupinu top produktů podle počtu hodnocených produktů uživatelem. Systém o uživateli nemá dostatek informací při jeho prvních návštěvách, nejprve systém shromažďuje data o chování uživatele na webu a na základě jeho hodnocení vytváří seznam top produktů v určitém poměru. Pokud si uživatel pouze prohlíží produkty bez jejich ohodnocení, je seznam vytvořen z top produktů skupiny. Následně kdy uživatel začne hodnotit produkty, je výsledný seznam vytvořen procentuální kombinací obou částí algoritmu.



- Alespoň 5 produktů - 90% *Moje top produkty* a 10% *TF-IDF*.
- 11 až 20 produktů - 70% *Moje top produkty* a 30% *TF-IDF*.
- 21 až 30 produktů - 50% *Moje top produkty* a 50% *TF-IDF*.
- 31 a více produktů - 30% *Moje top produkty* a 70% *TF-IDF*.

Třetí přístup je k dispozici pouze v okamžiku, kdy uživatel provedl nějaký nákup. Tento přístup vybere všechny produkty z uživatelské historie nákupů a nalezne k nim podobné produkty podle algoritmu TF-IDF. Každý produkt získá koeficient na základě umístění v historii nákupů. Produkty, které budou zakoupeny naposledy, budou mít větší koeficient a tedy větší vliv na doporučení podobných produktů podle algoritmu TF-IDF.

- Poslední nákup - koeficient 1.
- Nákupy za poslední týden - koeficient 0,7.
- Nákupy za poslední měsíc - koeficient 0,5.
- Nákupy delší jak měsíc - koeficient 0,2.
- Nákupy delší jak 3 měsíce - koeficient 0 (nebrány v potaz do výpočtu).

Čtvrtý přístup pracuje opět s historií uživatele, avšak využívá k tomu rozšířená data o historii nákupů podobných uživatelů. Algoritmus nejprve nalezne podobné uživatele k aktivnímu uživateli, následně získá jejich historii nákupů a vytvoří user-item matici. Kde řádky značí produkty, sloupce znázorňují uživatele a hodnoty jsou vyplněny hodnocením produktu daného uživatele. Pokud uživatel produkt nehodnotil, je hodnota nastavena na 0. Tato matice je důležitá pro aplikování algoritmu SVD (Singular Value Decomposition).

## **5 TESTOVÁNÍ**

### **5.1 Popis testování**

### **5.2 Průběh**

### **5.3 Výsledky**

### **5.4 Závěr**

## **ZÁVĚR**

## **RESUMÉ**

## **SUMMARY**

## SEZNAM POUŽITÉ LITERATURY

1. **Příjmení, Jméno.** *Název knihy.* Město vydání: Vydavatelství, 2003. 123-4-56-789123-4.

2. **Příjmení1, Jméno1 a Příjmení2, Jméno2.** *Název webové stránky. Název webu.*

[Online] Produkční společnost, 23. Zář 2006. [Citace: 19. Zář 2008.]

<http://www.urladresa.cz>. 12-3456-789-12.

1. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.19.777&rep=rep1&type=pdf>

2. <https://1url.cz/DM9HQ>

## SEZNAM POUŽITÝCH SYMBOLŮ

ABC Význam první zkratky.

B Význam druhé zkratky.

C Význam třetí zkratky.

## **SEZNAM OBRÁZKŮ**



## **SEZNAM TABULEK**

## **SEZNAM PŘÍLOH**