

TALLINNA TEHNIKAÜLIKOOL  
Infotehnoloogia teaduskond

Regina Helena Lõpp-Elmeste 154877IAPB

**PIMEDATE ÖÖDE FILMIFESTIVALI  
VEEBILEHELE FILMISOOVITUSTE  
MOOTORI ARENDAMINE**

Bakalaureusetöö

Juhendaja: Ago Luberg

Tallinn 2019

## **Autorideklaratsioon**

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Regina Helena Lõpp-Elmeste

09.01.2019

## **Annotatsioon**

Paljud ettevõtted seisavad silmitsi probleemiga, kuidas võimalikult palju tooteid võimalikult paljudele klientidele maha müüa. Kuna kliendid on väga palju mõjutatud sellest, mida neile pakutakse, siis toodete soovitude mootorid on ideaalne lahendus mainitud probleemile.

Töö eesmärgiks on arendada Pimedate Ööde filmifestivali kodulehele filmisoovituste mootor, mis tooks festivalile rohkem filmihuvilisi ning aitaks läbi soovitude veebilehe kasutajatel lihtsama vaevaga enda jaoks sobivaid filme valida.

Töö tulemusena loodi skript, mis võtab andmebaasist filmide märksõnad, žanrid ja programmid ning nende põhjal genereerib filmiga sarnaste filmide nimekirja ja lisab selle andmebaasis vastavasse kohta. Samuti valmis filmisoovituste mootor, mis vastab ettevõtte vajadustele ja nõuetele. Kasutajad saavad kodulehe kaudu vaadatud filmidele hindteid anda, filme lisada OMA KAVAsse ning osta pileteid, mille põhjal süsteem genereerib neile filmid, mis võiksid kliendile huvi pakkuda ja meeldida.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 27 leheküljel, 7 peatükki, 7 joonist, 4 tabelit.

## **Abstract**

### **Film recommendations engine development for the Black Nights Film Festival webpage**

A lot of companies often wonder how they can sell a lot of their products to a lot of customers. Since clients are very affected by what they are offered, then product recommendation engines are the ideal solution for this problem.

The aim of this work is to develop a film recommendation engine for the website of NGO Black Night Film Festival, which will bring more customers to the company and help the users better select films which they like.

The result of this thesis was a script that takes keywords or tags, genres and programs from the films database and generates a movies list that are similar to the films and inserts the list to the corresponding field in the database. The result also included a film recommendations engine, which meets the requirements and needs of the company. Users can rate films which they have watched, add films to their personal schedule and buy tickets and based on this information the system generates films that are suitable for the client.

The thesis is in Estonian and contains 27 pages of text, 7 chapters, 7 figures, 4 tables.

## Lühendite ja mõistete sõnastik

PÖFF	Pimedate Ööde Filmifestival
FIAPF	Rahvusvaheline Filmiprodutsentide Liitude Föderatsioon (prantsuse keeles <i>Fédération Internationale des Associations de Producteurs de Films</i> ) [1]
SMS	Lühisõnum
int	Number
string	Tekst, sümbolite jada
varchar	Andmebaasi andmete salvestamise tüüp, mis sisaldab tähti ja numbreid
XML	Laiendatav märgistuskeel (inglise keeles <i>Extensible Markup Language</i> ) [2]
PHP	Skriptimiskeel (inglise keeles <i>Hypertext Preprocessor</i> ) [3]
HTML	Hüperteksti märgistuskeel (inglise keeles <i>HyperText Markup Language</i> ) [4]
DOM	Dokumendi objektimudel (inglise keeles <i>Document Object Model</i> ) [5]
CSS	Küljenduskeel (inglise keeles <i>Cascading Style Sheets</i> ehk astmelised stiililehed) [6]
SASS	Stiililehe keel (inglise keeles <i>Syntactically Awesome Style Sheets</i> ) [7]
JS	JavaScript – objektorienteeritud skriptimiskeel [8]
jQuery	JavaScripti teek [9]
SQL	Struktuurpääringukeel (inglise keeles <i>Structured Query Language</i> ) [10]
IE	Internet Explorer, veebibrauser

## Sisukord

1 Sissejuhatus .....	10
1.1 Taust ja probleem .....	10
1.2 Eesmärgid .....	11
1.3 Metoodika.....	11
1.4 Ülevaade tööst .....	11
2 Filmisoovituste mootori spetsifikatsioon .....	12
2.1 Varasemad lahendused .....	12
2.2 Nõuded ja ettevõtte vajadused .....	12
2.2.1 Funktsionaalsed nõuded .....	13
2.2.2 Mittefunktsionaalsed nõuded.....	14
2.3 Andmebaasi tabelite kirjeldus .....	15
3 Filmisoovituste mootori kirjeldus.....	18
3.1 Kasutatud tehnoloogiad .....	18
3.1.1 PHP.....	18
3.1.2 HTML.....	18
3.1.3 CSS ja SASS.....	18
3.1.4 JavaScript ja jQuery .....	19
3.1.5 SQL.....	19
3.2 Kasutajaliides.....	19
3.2.1 Profiilivaade .....	20
3.2.2 Filmi vaade .....	21
3.2.3 Personaalsete filmisoovituste lehe vaade .....	23
4 Soovitusmootori realiseerimine .....	24
4.1 Andmete kogumine.....	24
4.2 Andmete talletamine.....	24
4.3 Andmete analüüsimine .....	24
4.4 Andmete filtreerimine.....	25
4.4.1 Lihtne filtreerimine.....	25
4.4.2 Sisupõhine filtreerimine .....	26

4.4.3 Koostööpõhine filtreerimine.....	26
4.5 Üksteisega sarnased filmid .....	28
4.6 Filmele antud hinded .....	29
4.7 OMA KAVA .....	29
4.8 Ostetud piletid.....	30
5 Valminud soovitusmootori analüüs .....	31
5.1 Testimine .....	31
5.2 Soovitusmootori kasutamise statistika.....	32
6 Võimalikud edasiarendused.....	33
7 Kokkuvõte .....	34
Kasutatud kirjandus .....	35

## Jooniste loetelu

Joonis 1. Navigatsioonidiagramm. ....	20
Joonis 2. Sisselogitud kasutaja profiili vaade, kust saab personaalsete filmisoovituste vaatesse. ....	21
Joonis 3. Valitud filmi terviklik vaade. ....	21
Joonis 4. Sisselogimata kasutajale valitud filmiga sarnaste filmide kuvamine. ....	22
Joonis 5. Sisse logitud kasutajale personaalsete filmisoovituste ja valitud filmiga sarnaste filmide kuvamine. ....	22
Joonis 6. Filmide hindamise vaade enne sisselogimist, enne hindamist ja pärast hindamist. ....	22
Joonis 7. Personaalsete filmisoovituste vaade. ....	23



## Tabelite loetelu

Tabel 1. Andmebaasi tabeli „ <i>users</i> “ kirjeldus.....	16
Tabel 2. Andmebaasi tabeli „ <i>films</i> “ kirjeldus. ....	16
Tabel 3. Andmebaasi tabeli „ <i>films_screenings</i> “ kirjeldus. ....	17
Tabel 4. Andmebaasi tabeli „ <i>films_tags</i> “ kirjeldus .....	17

# 1 Sissejuhatus

Pimedate Ööde filmifestival ehk PÖFF on 1997. aastal asutatud Põhja-Euroopa üks suurimaid filmifestivale ning kuulub maailma viieteistkümne A-kategooria võistlusprogrammiga festivali hulka, mis on tunnustatud Rahvusvahelise Filmiprodutsentide Liitude Föderatsiooni ehk FIAPF'i poolt. Festivali kavas on enam kui 250 täispikka mängufilmi rohkem kui 70st eri riigist ja umbes 350 lühi- ja animafilmi. Viimastel aastatel on festivali külastajaid olnud rohkem kui 80 000 inimest ja festival on Eestisse toonud üle 300 välismaise filmindustegelase. [11]

Kuna filme on nii palju, siis on festivali külastajatel väga keeruline enda jaoks sobivaid filme välja valida, mis neile meeldida võiks. Selle tõttu tekkiski MTÜ Pimedate Ööde Filmifestivalil mõte, et võiks teha filmisoovituste mootori. Käesoleva bakalaureusetöö eesmärgiks ongi teha PÖFF'i kodulehele filmisoovituste mootor, mis on ette nähtud laiemal projektil osana.

## 1.1 Taust ja probleem

Paljud ettevõtted seisavad silmitsi probleemiga, kuidas võimalikult palju tooteid võimalikult paljudele klientidele maha müüa. Kuna PÖFF-il on erinevaid filme väga palju (kokku üle 500), siis on filmihuvilistel väga keeruline omale sellest suurest hulgast seda õiget filmi välja valida ja mõned inimesed võib-olla jätavad üldse filmi vaatamata, sest ei tea, kas see võiks neile meeldida või mitte.

Kliendid on väga palju mõjutatud sellest, mida neile pakutakse. 35 protsenti Amazoni ostudest ja 75 protsenti Netflix'i filmide vaatamistest on tehtud just toodete soovitusel algoritmi abiga [12]. Sellest võib järeldada, et toodete soovitusel moodustatud lahendus on ideaalne lahendus mainitud probleemile. Teiseks probleemiks on PÖFF-i jaoks asjaolu, et publiku lemmikfilmi hääletamine on siiani ainult paberil (või SMSi teel) toimunud ja pärast hääletuse lõppu peavad meeskonna liikmed kõik hääled käsitsi ükshaaval kokku lugema. See lahendus on mõlema osapoole jaoks üsna tülikas ja aeganõudev.

## **1.2 Eesmärgid**

Käesoleva bakalaureusetöö eesmärgiks on arendada filmisoovituste mootor, mis aitab inimestel teatud kriteeriumite järgi leida just neile sobivad filmid. Soovitatavad filmid peaks võimalikult kiiresti üles leidma ja peavad üksteisega mingite ühiste parameetrite põhjal sarnanema.

Veebilehel peaks olema ka võimalik filmidele hindeid anda (näiteks 5-palli süsteemis). Seda peaks iga inimene saama teha iga filmi jaoks ainult ühe korra ja ainult juhul, kui ta on seda konkreetset filmi käinud kinos vaatamas. Filmidele hinnete andmine peaks olema võimalikult lihtne ja mugav kasutada, et ka vanemad inimesed saaksid sellega probleemivabalt hakkama.

## **1.3 Metoodika**

Filmisoovituste mootori arendamisel kasutab töö autor PÖFF-i andmebaasis filmide juures olevaid märksõnu, mis on eelnevalt imporditud Eventivalist (filmifestivalide haldussüsteemist). Tulemuse sobivuse kontrolliks tuleb vaadata, kas teatud filmi lehte vaadates soovitatakse piisavalt sarnaseid filme.

Kui kasutaja on lehele sisse loginud, siis tuleks vaadatud filmide juurde lisada hinnangu andmise võimalus. Tulemuste sobivuse kontrolliks tuleb teha näidistabel, kus on testandmed selle kohta, mis filme keegi vaatamas on käinud. Seejärel tuleks logida testkontodega PÖFFi veebilehele sisse ja üritada filmidele hinnanguid anda. Lõpuks tuleks vaadata, kas hinne salvestus andmebaasi ja kontrollida, kas kliendil välja- ja uuesti sisse logides on antud hinnangud veel nähtavad.

## **1.4 Ülevaade tööst**

Töö sisu koosneb viiest osast. Esmalt annab autor ülevaate filmi soovitude varasematest lahendustest, süsteemi nõuetest ja ettevõtte vajadustest ning kirjeldab süsteemi arhitektuuri ja andmebaasi tabeleid. Järgmisena kirjeldatakse filmisoovituste mootori loomise protsessi ja realiseerimist, sellele järgnev peatükk räägib valminud süsteemi analüüsist ja testimisest. Viimane peatükk keskendub võimalikele edasiarendustele.

## **2 Filmisoovituste mootori spetsifikatsioon**

Käesolev peatükk toob välja filmisoovituste mootori varasemad lahendused, ettevõtte vajadused, funktsionaalsed ja mittefunktsionaalsed nõuded ning kirjeldab süsteemi arhitektuuri ja annab ülevaate andmebaasi tabelitest.

### **2.1 Varasemad lahendused**

Kui inimesed varasematel aastatel ei osanud endale festivali suure filmivaliku hulgast sobilikke filme valida, siis käisid nad PÖFF-i infopunktides infotöötajalt või piletimüüjalt uurimas, milliseid filme nemad soovitada oskavad. Infotöötajad ei ole võimelised kogu mahukat programmi (ligi 500) filmi meelde jätma, seega on oluline leida filmisoovituste jaoks innovaatilisemad lahendused. Automaatne süsteem, mis lähtuks iga inimese personaalsest maitsest, oleks aga tunduvalt efektiivsem ja täpsem.

### **2.2 Nõuded ja ettevõtte vajadused**

MTÜ Pimedate Ööde Filmifestivali eesmärk on, et tema kliendid leiaksid omale kiiresti ja mugavalt sobilikud filmid ning seega käiksid rohkem kinodes filme vaatamas. Ettevõtte soovib ka seda, et kliendid saaksid läbi kodulehe filmidele 5-punkti süsteemis võimalikult mugavalt hindeid anda. See vähendaks meeskonna tööd, sest seni toimus kogu hääletus paberil. Samuti on oluline tagasiside programmi meeskonnal, millised filmid küllastajatele meeldivad ja millised mitte. Järgmise aasta programmi koostamiseks on see info väga vajalik.

Süsteemi nõuded on jaotatud kaheks: funktsionaalsed ja mittefunktsionaalsed nõuded. Funktsionaalseid nõudeid tuleb järgida filmisoovituste mootori arenduse käigus, sest need näitavad mida süsteem peaks teatud olukorras tegema. Mittefunktsionaalsed nõuded aga kirjeldavad detailselt kuidas süsteem neid tegema peaks ning sisaldavad kasutatavust, käideldavust, jõudlust ja toetatavust. [13]

### 2.2.1 Funktsionaalsed nõuded

Filmisoovituste mootor peab soovitama süsteemi kasutajale filme kahel eri viisil:

- Kui avada ühe filmi vaade, siis kõige all peaks olema nähtavad valitud filmiga sarnased filmid.
  - Kui klient on sisse loginud, siis kuvatakse ühe filmi vaate all ka kliendi personaalsed filmisoovitused.
  - Kui klient ei ole sisse loginud, siis kuvatakse ühe filmi vaate all ainult sarnaseid filme.
- Klient peab nägema ka kõiki oma personaalseid filmisoovitusi.
  - Personaalsed filmisoovitused peavad olema kliendi jaoks lihtsalt üles leitavad (näiteks profiilivaatest või menüüribalt).

Personaalsed soovitused koostatakse kliendile kahte erinevat aspekti arvestades:

- Filmidele antud hinnete põhjal:
  - Klient peab saama filme hinnata 5 palli süsteemis.
  - Ühele filmile peab saama iga klient anda hinnet ainult ühe korra.
  - Klient võiks saada oma antud hinnet ka hiljem muuta. Sellisel juhul läheb arvesse viimane hinne.
  - Klient võiks saada ka oma hinnet kustutada (näiteks, kui kogemata vajutas mõne hinde peale).
  - Filmidele antud hinded salvestatakse andmebaasi listina, mis on muudetud komadega eristatud stringiks ning võtme ja väärtuse vahel on koolon.
  - Kui klient on mõnda filmi hinnanud, siis personaalsete soovituste all seda filmi ei näidata, sest klient on arvatavasti seda filmi juba näinud.
  - Filmidele antud hinded peaksid olema kõige tähtsamal kohal personaalsete soovituste tegemisel.

- OMA KAVAsse lisatud filmid:
  - Klient peab saama filme lisada OMA KAVAsse ning nende põhjal tehakse talle ka soovitusi.
  - OMA KAVAsse lisatud filme ei tohiks personaalsete soovitude nimekirjas enam kuvada.
- Filmid, millele on ostetud pilet:
  - Klient peab saama vaadata, millistele filmidele tal on pilet ostetud (näiteks profiilivaatest).
  - Nendele filmidele, kuhu on klient ostnud pileti arvestatakse personaalsete soovitude tegemisel.
  - Filme, mille pilet on kliendil juba olemas, ei tohiks personaalsete soovitude nimekirjas enam kuvada.
- Kõikide personaalsete soovitude vaates peab klient vahet tegema filmidel, millel on mitu talle sobilikku märksõna ning filmidel, millel on ainult üks sobilik märksõna.
- Kõikide personaalsete soovitude vaates peavad olema nii hinnete põhjal kui ka märksõnade põhjal soovitatavad filmid.
- Kõikides filmisoovitude kuvamise kohtades peaksid olema soovitatavad filmid kindlas järjekorras (tähtsamad eespool ja vähemtähtsad tagapool).
- Soovitatavale filmile peale vajutades peab olema kliendil võimalus selle filmi kohta rohkem infot näha.

### **2.2.2 Mittefunktsionaalsed nõuded**

Kasutatavus:

- Filmisoovitude mootorit peavad saama kasutada kõik PÖFF-i kliendid, kes tahavad enda jaoks sobivaid filme leida.

- Filmisoovituste mootor peab olema üheselt mõistetav ja kasutajale võimalikult mugav ja lihtne.
- Filmisoovituste mootor peab olema kasutatav kõikides enamkasutatavates brauserites ja kõikide operatsioonisüsteemidega.
- Filmisoovitusi peab olema võimalik vaadata nii arvutiga kui ka mobiiliga.
- Filmisoovitused peavad olema loetavad nii eesti kui ka inglise keeles.

Käideldavus:

- Kui filmisoovituste mootor mingil põhjusel ei peaks teatud hetkel korralikult töötama, siis kõik muu lehel olev selle tõttu ei kannata ja töötab korralikult edasi filmisoovituste mootorist sõltumata.

Jõudlus:

- Filmisoovituste mootori algoritmide arvutamine ei tohi võtta aega kauem kui 1 sekund.

Toetatus:

- Filmisoovituste mootorile peab olema hiljem võimalik täiendavat funktsionaalsust juurde lisada.

### **2.3 Andmebaasi tabelite kirjeldus**

Filmisoovituste mootori andmebaasi jaoks on vajalikud 3 tabelit. Järgnevalt tuuakse välja nende tabelite oluliste veergude nimed, tüübid, pikkused, kohustuslikkuse ja näidisväärtused.

Tabel 1. Andmebaasi tabeli „users“ kirjeldus.

Nimi	Tüüp	Pikkus	Kohustuslik	Näidisväärtus
id PK	int	11	jah	123
omakava	text		ei	90701,10801,11202,21102
bought_films	text		ei	070101,151001,170509
ratings	text		ei	416972:2,416632:5
personal_films	text		ei	{"14":[418245,416766,417083], "13":[416527,416072,420230, 422083], ...}
personal_films_timestamp	timestamp		ei	2018-12-03 18:55:57

Tabel 2. Andmebaasi tabeli „films“ kirjeldus.

Nimi	Tüüp	Pikkus	Kohustuslik	Näidisväärtus
id PK	varchar	6	jah	302516
sub_films	text			422208, 422210, 422211, 422213, 422214
sub_film	int	1	jah	1
title_original	text		jah	Brama
title_eng	text		jah	The Gateway
title_est	text		jah	Värav
completion_year	year		jah	2017
countries	text		ei	UA,US
directors	text		jah	Adrian Panek
categories	text		jah	Just Film
programs	text		ei	Põhivõistlusprogramm
tags_est	text		ei	humor, põnevus, armastus
genres	text		ei	draama
screenings	text		ei	121502, 140503, 151902
silimar_films	text		ei	310632, 320583, 321359, 322462



Tabel 3. Andmebaasi tabeli „films\_screenings“ kirjeldus.

Nimi	Tüüp	Pikkus	Kohustuslik	Näidisväärtus
id PK	varchar	6	Jah	010206
film_id FK	varchar	6	jah	414955
date	date		jah	2018-11-24
time	time		jah	18:30:00
normal_seats_available	int	1	jah	1
double_seats_available	int	1	jah	0

Tabel 4. Andmebaasi tabeli „films\_tags“ kirjeldus

Nimi	Tüüp	Pikkus	Kohustuslik	Näidisväärtus
id	int	11	Jah	12
tag_name	varchar	100	Jah	Draama
film_ids	text		Jah	309091, 313837, 321002, 322691, 345678

## 3 Filmisoovituste mootori kirjeldus

Töö autor arendas lõputöö käigus MTÜ Pimedate Ööde Filmifestivali kodulehele filmisoovituste mootori, et klientidel oleks lihtsam enda jaoks sobivaid filme valida ja ettevõtte saaks omale rohkem kliente, kes käivad filme vaatamas. Käesolevas peatükis kirjeldatakse selle mootori loomise protsessi.

### 3.1 Kasutatud tehnoloogiad

Filmisoovituste mootor on üks veebilehe osa, mis on kirjutatud kasutades PHP, HTML, CSS ja JavaScripti programmeerimiskeelt. Lisaks on kasutatud ka SASS-i ja jQuery-t. Andmebaasiga suhtluseks on kasutatud MySQL-i.

#### 3.1.1 PHP

PHP (inglise keeles *Hypertext Preprocessor*) on skriptimiskeel, mida saab väga lihtsalt siduda HTML-iga. PHP-d kasutatakse põhiliselt serveripoolsetes lahendustes dünaamiliste veebilehtede loomisel ning see toetab objektorienteeritud programmeerimist ning andmebaasiga suhtlemist. [3]

#### 3.1.2 HTML

HTML (inglise keeles *HyperText Markup Language* ehk „hüperteksti märgistuskeel“) on programmeerimiskeel, mida kasutatakse veebilehtede valmistamiseks. HTML paneb paika veebilehe struktuuri ning dokumentide vaatamiseks kasutatakse veebibrauserit. HTML failid koosnevad elementidest, mis on ümbritsetud noolsulgudega („<“ ja „>“) ja annavad brauserile teada, mida ja kuidas selle sisuga teha. [4]

#### 3.1.3 CSS ja SASS

CSS (inglise keeles *Cascading Style Sheets* ehk astmelised stiililehed) on küljenduskeel, mida kasutatakse veebilehe kujundamiseks. CSS-iga saab muuta HTML elementide paigutust, suurust, värvi ja palju muid omadusi. [6]

SASS (inglise keeles *Syntactically Awesome Style Sheets*) on stiililehe keel, mida kasutatakse samuti veebilehtede kujundamiseks ning see on CSS-i laiendus. SASSis

kirjutatud kood teisendatakse edasi CSS-i. SASSi kasutatakse, sest sellel on CSS-koodi ees päris palju eeliseid: muutujate defineerimine (ing. k. *variables* - toetatud andmetüübid on number, sõne, värv, tõeväärtus), pesastamine (ing. k. *nesting* - võimalus kirjutada CSS-i reeglid üksteise sisse), mikserid (ing. k. *mixins*) ja selektorite pärilus (ing. k. *inheritance*). Antud töös kasutatakse SASSi uuemat süntaksit SCSS. [7]

### 3.1.4 JavaScript ja jQuery

JavaScript (lühend JS) on objektorienteeritud skriptimiskeel, mille abil saab veebilehele efekte lisada. HTML, CSS ja JS kolmekesi on veebilehtede põhitehnoloogiad. JavaScripti peetakse maailma populaarseimaks programmeerimiskeeleks. [8]

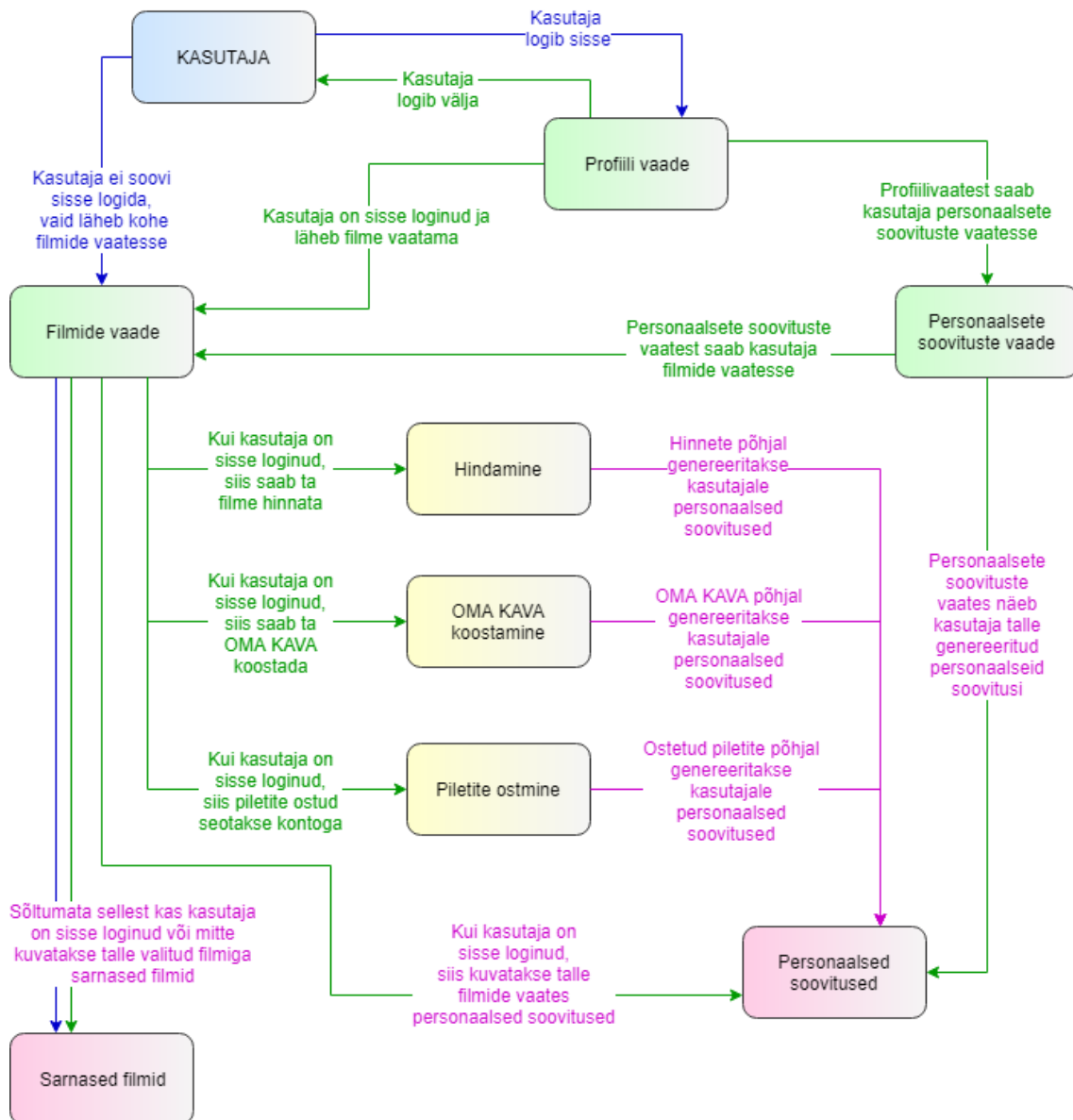
jQuery on JavaScripti teek, mis on loodud selleks, et kasutajal oleks mugavam dokumendis navigeerida ning võimaldada kiiret juurdepääsu kõikidele DOM-elementidele, nende sisule ja atribuutidele ning neid muuta. [9]

### 3.1.5 SQL

SQL (inglise keeles *Structured Query Language* ehk struktuurpäringukeel) on andmebaasi päringukeel, mida saab koos PHP-ga kasutada. SQL-iga saab teha andmete päringuid, uuendamisi, struktuuri loomist ja muutmist ning andmetele juurdepääsu kontrolli. [10]

## 3.2 Kasutajaliides

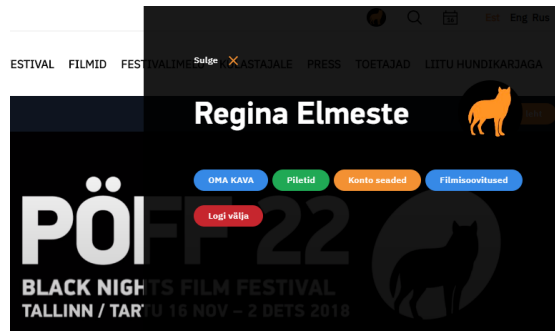
Kasutaja navigeerimisvõimalusi iseloomustava navigatsioonidiagramm on välja toodud joonisel 1. Sinine kast märgib kasutaja objekti, rohelised kastid on erinevad vaated, kollased kastid kasutaja tegevused, roosad kastid filmi objektid. Sinised jooned kirjeldavad sisse logimata kasutaja tegevusi, rohelised jooned sisse logitud kasutaja tegevusi ja roosad süsteemi tegevusi.



Joonis 1. Navigatsioonidiagramm.

### 3.2.1 Profiilivaade

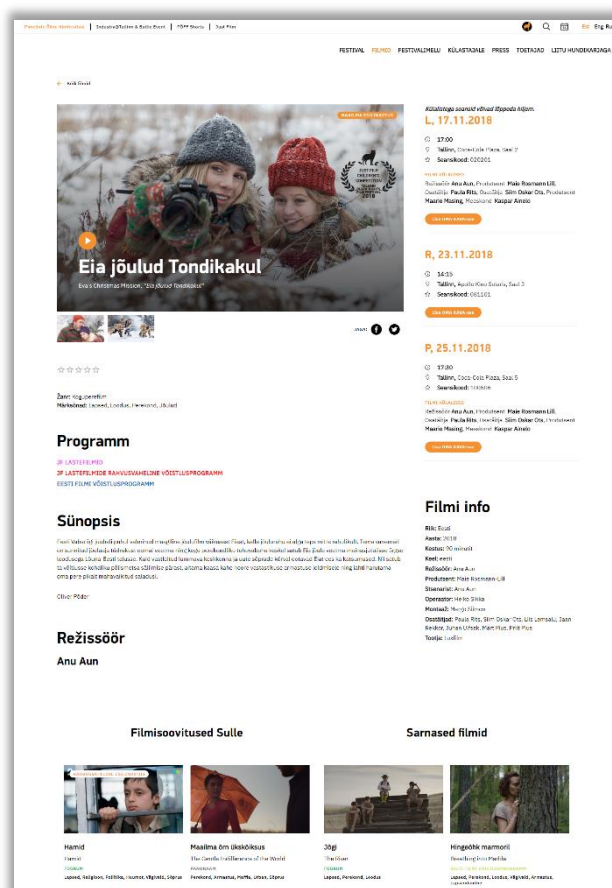
Kui kasutaja avab veebilehe, siis ekraani ülemises paremas nurgas on profiiliikoon, millele vajutades on kasutajal võimalik veebilehele sisse logida. Sisselogimine on võimalik e-maili, facebook'i ja Google-i kaudu. Kui kasutaja on sisse loginud, siis on tal sealt samast profiilivaatest (Joonis 2) leida link, millele vajutades näeb ta kõiki oma personaalseid filmisoovitusi.



Joonis 2. Sisselogitud kasutaja profiili vaade, kust saab personaalsete filmisoovituste vaatesse.

### 3.2.2 Filmi vaade

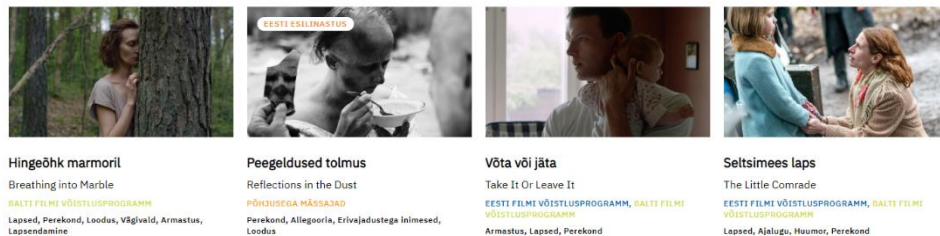
Antud vaade kuvab kasutajale kogu valitud filmi info (Joonis 3). Filmi piltide all on näha hindamise võimalust ning selle all on märgitud ära ka filmi žanr, märksõnad ning programmid, mille põhjal kasutajale filme soovitatakse.



Joonis 3. Valitud filmi terviklik vaade.

Lehe kõige alumises osas on võimalik näha ka valitud filmiga teisi sarnaseid filme. Kui kasutaja on sisse loginud, siis talle kuvatakse samas vaates ka personaalsed filmisoovitused (Joonis 5). Kui kasutaja ei ole sisse loginud, siis ta personaalseid soovitusi ei näe (Joonis 4).

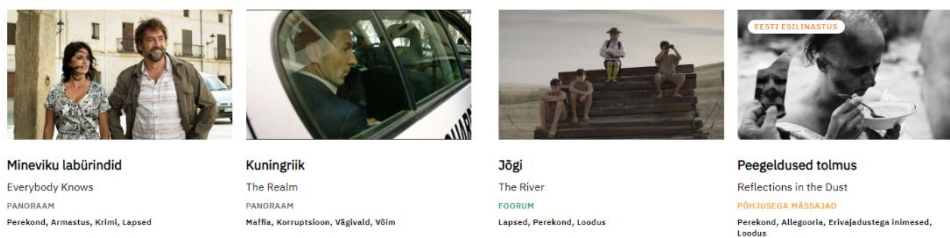
#### Sarnased filmid



Joonis 4. Sisselogimata kasutajale valitud filmiga sarnaste filmide kuvamine.

#### Filmisoovitused Sulle

#### Sarnased filmid



Joonis 5. Sisse logitud kasutajale personaalsete filmisoovituste ja valitud filmiga sarnaste filmide kuvamine.

Filmi vaates on kasutajal võimalik ka filmidele hindeid anda (Joonis 3). Kasutajale on näha 5 tähte, kui neile hiirega peale liikuda, siis muudavad need tähed vastavalt värvi. Valitud hinne jääb kasutajale näha. Kui kasutaja ei ole sisse loginud, siis veebileht suunab kasutajat sisse logima (Joonis 6).

#### Filmi hindamine

Filmi hindamiseks logi sisse!

#### Filmi hindamine



#### Filmi hindamine



Joonis 6. Filmide hindamise vaade enne sisselogimist, enne hindamist ja pärast hindamist.

### 3.2.3 Personaalsete filmisoovituste lehe vaade

Personaalsete filmisoovituste lehe vaates näeb sisseloginud kasutaja filmisoovitusi, mis on koostatud tema filmidele antud hinnete, OMA KAVAs olevate filmide ja ostetud piletite põhjal. (Joonis 7). Pildil olevad numbrid näitavad sobivuse protsenti ehk mida suurem on number, seda suurem on tõenäosus, et need filmid võiksid kasutajale meeldida. Kasutajale kuvatakse ainult kaheksa blokki neid filme, millel on kõige suurem sobivusprotsent.

Pimedate Ööde Filmifestival | Industry@Tallinn & Baltic Event | PÖFF Shorts | Just Film

FESTIVAL FILMID FESTIVALIMELU KÜLASTAJALE PRESS TOETAJAD LIITU HUNDIKARJAGA

## Filmisoovitused

Filmisoovitused on koostatud OMA KAVAs olevate filmide, ostetud piletite ja filmide hinnangute põhjal.

Allolevad numbrid näitavad Sulle sobivuse protsenti.  
Mida suurem on number, seda suurem on tõenäosus, et need filmid Sulle meeldida võiksid.

**50.72 %**

**Kõik on hästi**  
All is Good  
PANORAAM, GÖTTE-INSTITUUT ESITLES: IUS  
SILKJA STUM  
Armastus, Seks, Eneseotsing, Vägivald

**Armastus viimasest pilgust**  
An Unexpected Love  
PANORAAM  
Armastus, Perekond, Inimsuhted, Meeste keskekrilis

**Tavaline naine**  
An Ordinary Woman  
TV BEATS  
Huumor, Põnevus, Armastus

**Pärijannad**  
The Heiresses  
PANORAAM  
LGBT, Armastus, Huumor

**49.28 %**

**Rinnahoidja**  
The Bra  
FORUM  
Armastus, Huumor, Teekond

**Mind ei huvita, kui me ka läheme ajalukku barbaritena**  
I Do Not Care If We Go Down in History as Barbarians  
PANORAAM  
Satir, Poliitika, Ajalugu, Huumor

**Kärva maha, paps!**  
Why Don't You Just Die!  
DEBÜÜTFILMIDE VÕESTLUSPROGRAMM  
Vägivald, Satir, Põnevus, Huumor

Joonis 7. Personaalsete filmisoovituste vaade.

## **4 Soovitusmootori realiseerimine**

Soovitusmootorid koosnevad tüüpiliselt neljast eri faasist: andmete kogumine, talletamine, analüüsimine ja filtreerimine. [14]

### **4.1 Andmete kogumine**

Andmeid saab koguda kahel viisil: vastavalt kasutaja käitumisele või kasutaja sisestatud info põhjal. Esimesel viisil andmete kogumine on tunduvalt lihtsam, sest see ei nõua kasutajalt mingeid lisa tegevusi. Süsteem jälgib ainult kasutaja tegevust ja vastavalt sellele salvestab andmed logisse. Sellisel viisil kogutud andmete analüüsimine on aga väga keeruline, sest süsteem peab vahet tegema asjadel, mis kasutajale meeldisid ja mis mitte. [14] Sel põhjusel on antud töö raames kasutatud teist andmete kogumise viisi, mille puhul kasutaja peab ise info süsteemi sisestama. Sellise variandi puhul on küll andmete kogumise protsess kasutaja jaoks keerulisem, kuid lõpptulemus on tunduvalt täpsem ja seda lisa liigutust väärt.

### **4.2 Andmete talletamine**

Mida rohkem andmeid on kasutaja kohta salvestatud, seda paremaid soovitusi saab talle nende andmete põhjal teha. Arvestades seda, milliseid andmeid süsteem kogub, siis sõltuvalt sellest tuleks ka kohe paika panna, et mis kujul need andmed salvestatakse. [14] Antud töös salvestatakse kasutajate tabelisse (Tabel 1) kasutaja poolt filmidele antud hinded, OMA KAVAsse lisatud filmid ning filmid, millele kasutaja on piletid ostnud. Kõik kolm on talletatud andmebaasis listina, mis on teisendatud komadega eraldatud stringiks. Hinnete puhul on filmikood ja antud hinnang eraldatud kooloniga.

### **4.3 Andmete analüüsimine**

Salvestatud andmete analüüsi faas nõuab arusaamist süsteemi nõuetest. Kui kiiresti peab süsteemi suutma kasutajale soovitusi teha sõltub sellest, kas kasutaja peab neid soovitusi nägema kohe või alles hiljem, kui talle soovitused näiteks e-mailile saadetakse. Tähtis on ka see, kas filtreerimine põhineb ainult kasutaja maitsetel või on oluline arvestada ka teiste kasutajate maitseid või arvestatakse ainult neid tooteid, mis loogiliselt võttes üksteisega kokku peaksid sobima. [14]



On olemas kolme sorti analüüse. **Reaalajas** (inglise keeles *real-time*) süsteemid peavad suutma anda kasutajale soovitusi momentaalselt ja seega peavad olema suutelised saadud andmeid väga kiirelt analüüsima. **Peaaegu reaalajas** (*near-real-time*) süsteemid analüüsivad andmeid mõne sekundi või minuti tagant ning on sobilikud andma soovitusi, mis on mõeldud sama brauseri sessiooni jaoks. **Partii või portsjoni analüüs** (*batch analysis*) toimub perioodiliselt ehk süsteem ootab nii kaua kuni on kogutud piisav kogus andmeid ning alles siis hakkab neid analüüsima. See sobib väga hästi soovitude jaoks, mida on vaja saata kasutajatele e-maili peale. [14]

Käesolev töö kasutab filmide soovitamiseks peaaegu reaalajas analüüsi. Kui kasutaja hindab filme, siis saadud andmed salvestatakse kohe andmebaasi ja selle põhjal genereeritakse kasutajale koheselt uued filmisoovitused. Sama kehtib ka filmide OMA KAVAsse lisamise puhul ning piletiostude puhul.

## 4.4 Andmete filtreerimine

Andmete filtreerimine on soovitusmootori juures kõige olulisem. Kõige levinumad filtreerimisviisid on sisupõhised, rühmapõhised ja koostööpõhised.

Töö käigus loodud filmisoovituste mootor kasutab sisupõhist filtreerimist. Kasutajad, kes ei ole sisse loginud ja kelle kohta ei ole süsteemil veel piisavat infot näevad ainult mõne filmiga teisi sarnaseid filme. Sisse loginud kasutajatele soovitude tegemiseks arvestab süsteem kasutaja hinnatud filme, OMA KAVAs olevaid filme ning ostetud pileteid.

### 4.4.1 Lihtne filtreerimine

Lihtsa filtreerimise korral saab teha soovitusi kõikidele kasutajatele, ükskõik kas nad on sisse loginud või ole ning kas süsteemil on nende kohta andmeid või ei ole. Neid soovitusi tehakse kas populaarsuse ja/või mingi kindla tootomaduse põhjal. Selle meetodi puhul lähtutakse asjaolust, et populaarsel tootel on suurem tõenäosus meeldida keskmisele kasutajale. Sellist süsteemi kasutab näiteks IMDB Top 250. [15]

Antud töös ei ole kasutatud lihtsat filtreerimist, sest selle puhul on tegemist populaarsete filmide soovimisega. Filmide populaarsust ei ole just selle pärast arvestatud, sest sellisel juhul jääksid vähempopulaarsed filmid tähelepanuta ja keegi ei ostaks sinna pileteid.

#### **4.4.2 Sisupõhine filtreerimine**

Sisupõhise (*content-based*) filtreerimise puhul on tähtis teada iga pakutava toote omadusi. Selle meetodi põhiideeks on see, et kui sulle meeldib mingi toode, siis sulle ilmselt meeldib ka selle sarnane toode. Populaarne ja soovitatud toode omab sarnaseid tunnuseid nagu need tooted, mida kasutaja vaatab või mis talle meeldivad. [14]

Sisupõhine filtreerimine töötab kasutaja käest saadud andmete põhjal, näiteks kui kasutaja hindab mõnda toodet. Nende andmete põhjal saab teha kasutajale personaalseid soovitusi ja pakkuda meeldinud toodetega teisi sarnaseid tooteid. Mida rohkem kasutaja enda eelistuste kohta teada annab, seda targemaks süsteem saab ning seda paremaks ja täpsemaks ka soovitused muutuvad. [16]

Sellisel filtreerimismeetodil on palju eeliseid võrreldes teistega. See on väga efektiivne, kui on palju uusi kasutajaid või uusi tooteid ja seega saab ka kohe neid uusi ja mitte nii populaarseid tooteid soovitada. Unikaalse maitsega kasutajatele saab väga hästi temale sobivaid tooteid soovitada. Kohati võib aga sarnaste toodete leidmine olla väga keeruline, kui ei ole piisavalt omadusi üles märgitud ning kasutajal võib seetõttu jääda mõni tema jaoks vägagi sobilik toode märkamata. [16]

Antud töös kasutatakse sisupõhist filtreerimist. Kui kasutaja tunneb huvi mõne filmi kohta, siis näidatakse talle selle filmiga teisi sarnaseid filme, mis on seotud märksõnade, žanrite ja programmidega. Kui kasutaja on sisse loginud, hinnanud mõnda filmi, lisanud filmi OMA KAVAsse või ostnud pileteid, siis talle soovitatakse nende filmide põhjal teisi filme, mis talle meeldida võiksid.

#### **4.4.3 Koostööpõhine filtreerimine**

Koostööpõhine (*collaborative*) filtreerimine põhineb täielikult minevikus tehtud valikutele ja mitte toodete eripäradele. [16] See on täielikult seotud kahe kasutaja sarnasusega. Soovitatud toode on meeldinud teistele kasutajatele, kellele on meeldinud need samad tooted, mida antud kasutaja vaatab või mis talle on meeldinud. [14] Näiteks kui kasutajale X meeldivad tooted A, B ja C ning kasutajale Y meeldivad tooted A, B ja D, siis eeldatakse, et kasutajale X võiks meeldida toode D ning kasutajale Y võiks meeldida omakorda toode C. See on üks kõige populaarsemaid filtreerimismeetodeid, mis ei põhine sisul. [16]

Koostööpõhised filtreerimissüsteemid jagunevad kaheks:

- **Mälul põhinev** filtreerimine arvutab sarnasusi erinevate toodete ja kasutajate vahel. [14] See jaguneb veel omakorda kaheks:
  - **Kasutajate vahelise** filtreerimise puhul leitakse sarnasusi kahe kasutaja vahel ja soovitatakse tooteid, mis kellelegi teisele on meeldinud. See algoritm on väga efektiivne, aga väga kulukas, sest nõuab üsna palju aega ja ressursse, sest tuleb arvutada kõikide kasutajapaaride kohta eraldi tulemused. [16]
  - **Toodete vahelise** filtreerimise puhul leitakse sarnasusi kahe toote vahel ja kui kasutaja on mõnda toodet hinnanud, siis soovitatakse kasutajale selle tootega sarnaseid tooteid. See algoritm võtab palju vähem aega, sest tihti on toodete arv piiratud ning pika peale jääb see sarnasuste-maatriks muutumatuks. Uute kasutajate jaoks on see tunduvalt efektiivsem, sest ei pea hakkama arvutama iga kasutajapaari jaoks eraldi tulemusi. [16]
- **Mudelil põhinev** filtreerimine üritab õppida mustrit, mis näitab kuidas kasutajad hindavad või toodetesse suhtuvad. Näiteks üritab tuvastada kasutajate antud hindeid nende vanuse, soo või asukoha järgi. [14]

Koostöö põhine filtreerimine annab küll väga täpseid tulemusi, kuid teatud juhtudel on see väga ebaefektiivne. Kui on palju uusi kasutajaid või palju uusi tooteid, siis tekib külmakäivitamise probleem (ehk ei ole piisavalt tingimusi täidetud, et süsteem saaks tööle hakata) [17]. Suurte andmemahtude puhul võib olla keeruline leida kasutajaid, kes on hinnanud samu tooteid. Tihti soovitatakse ainult kõige populaarsemaid tooteid ja mõningad tooted jäävad seetõttu tähelepanuta. [16]

Külmakäivituse probleemi tõttu ei ole antud töös koostöö põhised filtreerimised kasutatud. Kuna enne töö tegemist ei olnud võimalik kasutajatel filme hinnata, siis ei olnud ka teada mitte ühegi kasutaja eelistusi ehk siis süsteemi mõistes olid kõik kasutajad uued. Teiseks põhjuseks on asjaolu, et iga aasta on festivalil uued filmid ja seega on väga keeruline soovitada filme, kui on pidevalt ainult uued tooted.

Kuna paljud külastajad on soovinud jääda anonüümseteks ja ostavad pileteid ilma sisse logimata, siis PÖFFil puudub selle kohta täpne statistika, kuid arvatakse, et keskmine

festivalikülastaja vaatab 17 päeva (~2,5 nädala) jooksul 3-5 filmi. Festivalipassi omanikud ja tõelised kinofännid jõuavad selle ajaga vaadata ära umbes 30 filmi. Kuna filme on palju – peaaegu 600 filmi ja üks kasutaja vaatab ära vaid mõned filmid sellest suurest hulgast, siis on üsna keeruline teiste kasutajate arvamuse põhjal soovitusi teha. Lisaks tuleb arvestada veel asjaolu, et väga väike protsent kasutajatest hindab filme ning mõni hindab ainult mõnda üksikut filmi, mitte kõiki, mida ta vaadanud on.

#### 4.5 Üksteisega sarnased filmid

Kui tavaliselt kasutatakse üksteisega sarnaste filmide leidmisel ka tähtsamate näitlejate, režissööri ja võib-olla ka produtsendi, stsenaaristi või mõne muu tähtsa filmitegija nime, [15] siis antud töös need nimed ei ole tähtsad. Kuna PÖFFi filmid on väga erinevatest riikidest üle maailma ja mõnel režissööril on alles esimene film, siis ei ole ühtegi filmi, millel need nimed kattuksid. Kui isegi leidub mõni üksik film, millel on sama režissöör, siis on neid väga vähe ja seega ei ole need PÖFFi filmisoovituste mootori jaoks olulised.

Antud töös on kasutatud filmide sarnasuste leidmiseks kolme omadust:

- **märksõnad:** lapsed, jõulud, tants, perekond, loodus
- **žanrid:** õudusfilm, komöödia, draama
- **programmid:** Põhivõistlusprogramm, Debüütfilmide programm, Eesti filmide võistlusprogramm, Lastefilmid

Kõik PÖFFi filmide andmed on imporditud Eventivalist XML-i kujul ja salvestatud PÖFFi andmebaasi (Tabel 2). Et nende andmete põhjal teada saada, mis filmid on omavahel seotud, koostas töö autor skripti, mis loob andmebaasi tabelisse „*films\_tags*“ (Tabel 4) märksõnade, žanrite, programmide ja vastavate filmide vahel seosed.

Kui filmide andmed on andmebaasi imporditud, siis on need algselt kujul filmikood => (märksõnade, žanrite, programmide list). Et need saada vastupidisele kujule (märksõnade, žanrite, programmide list => filmikoodid), siis selleks on esialgu vaja saada „*films*“ tabelist kätte filmikood ning sellele koodile vastavad märksõnad, žanrid ja programmid. Seejärel tuleb *for*-tsükliga ükshaaval läbi käia kõik filmid ja selle sees omakorda eraldi kõik märksõnad, žanrid ja programmid, teha need 3 ühiseks listiks ning lisada kõik

andmed uude listi kujul (märksõnade, žanrite, programmide list) => list filmikoodidest. Seejärel saab alles saadud andmed andmebaasi lisada.

Kui märksõnade tabelisse (Tabel 4) on lisatud märksõnad, žanrid ja programmid ja neile vastavate filmide koodid, siis saab lisada ka filmide tabelisse iga filmiga sarnaste filmide nimekirja. Selleks on töö autos loonud skripti, mis tuvastab ühe korraga filmide vahelised sarnasused ära. Kui nüüd tahta mõne filmiga sarnaseid filme saada, siis on vaja ainult andmebaasist küsida ja on kohe nimekiri olemas.

#### **4.6 Filmidele antud hinded**

Kui kasutaja on mõnda filmi näinud, siis saab ta filmidele hindeid anda 5-palli süsteemis, kus 5 näitab, et film väga meeldis ning 1 näitab, et film kasutajale ei meeldinud. Kui kasutaja on mõne filmi ära hinnanud, siis see hinne salvestatakse koheselt andmebaasi ja selle põhjal genereeritakse kasutajale uued personaalsed filmisoovitused.

Kui kasutaja ei ole ühtegi filmi hinnanud, ei ole lisanud ühtegi filmi OMA KAVAsse ning ei ole ka ühtegi piletit veel ostnud, siis talle personaalseid soovitusi ei näidata ja kasutaja suunatakse ühte neist tegema. Kui kasutaja on filme lisanud OMA KAVAsse, aga ei ole veel ühtegi piletit ostnud, ega filme hinnanud, siis tehakse soovitusi ainult OMA KAVAs olevate filmide põhjal (ja nii ka teiste tingimuste põhjal).

Kui kasutaja on kõik 3 tingimust täitnud, siis tehakse soovitusi kõikide põhjal, kuid hinnatud filmid on kõige suurema tähtsusega, sest selle põhjal saab täpselt teada, mis kasutajale meeldis ning mis mitte. OMA KAVAs olevatele filmidele ning filmidele, mille on ostetud piletid arvestatakse hindeks kõikidele filmidele antud hinnete keskmine.

#### **4.7 OMA KAVA**

OMA KAVA on kasutaja personaalne festivali kava, mille ta on ise koostanud endale välja valitud filmide seanssidest. Kui filmid veel kinodes ei linastu, siis kasutajad ei saa ka filme hinnata ja seega hinnangute põhjal soovitusi teha ei saa. Kui on aga mõned filmid OMA KAVAsse lisatud, siis selle põhjal saab küll juba soovitusi teha.

OMA KAVA on neist kolmest kõige populaarsem vahend, mida inimesed kasutavad. Seda koostatakse juba enne festivali algust. Kuid selle põhjal ei saa kindel olla, kas OMA KAVAs olev film kasutajale meeldib või ei meeldi enne kui ta ei ole filmi hinnanud.

#### **4.8 Ostetud piletid**

Profiilivaatest näeb kasutaja enda poolt välja valitud filmidele ostetud pileteid. Piletite ostmine toimub Piletilevi kaudu ning kui kasutaja läheb PÖFFi kodulehel oma pileteid vaatama, siis küsitakse see info Piletilevi serveritest ja saadakse JSONi kujul andmed, mida kuvatakse kasutajale arusaadaval kujul. PÖFFi andmebaasi salvestatakse seejärel ka ostetud piletite andmed. Kui kasutaja ei ole ise läinud PÖFFi kodulehel oma pileteid vaatama, siis andmebaasis neid ka ei ole. Seega koostööpõhisel meetodil soovitude tegemine oleks ostetud piletite põhjal väga keeruline.

## 5 Valminud soovitusmootori analüüs

Lõputöö tulemuse ja filmisoovituste mootori toimimise hindamiseks pidas töö autor vajalikuks süsteemi testimist erinevate brauserite, erinevate kasutajate, erinevate filmide ja erinevate hinnangutega. Samuti oli vajalik teha statistika analüüs, mis näitaks kui palju reaalselt kasutajad soovitusmootorit kasutasid ja kas sellest oli üldse kasu.

### 5.1 Testimine

Enne kui soovitusmootori sai PÖFFi veebilehel avalikustada, pidas töö autor vajalikuks süsteemi testimist erinevate brauserite, erinevate kasutajate, erinevate filmide ja erinevate hinnangutega.

Tänapäeval on kõige populaarsemateks brauseriteks arvutis Google Chrome, Safari, Internet Explorer & Edge, Firefox ja Opera [18]. Kuna Microsoft lõpetas 2016 aasta jaanuaris IE vanemate versioonide toetamise [19] ja paljud CSS-i kujundamise atribuudid ei ole IE-s toetatud, siis käesoleva töö autor ei pidanud oluliseks Exploreris oma loodud süsteemi testida. Filmisoovituse mootori arenduse käigus on jooksvalt testitud toimimist erinevates brauserites: Google Chrome'is, Operas, Firefoxis, Edge'is ja Safaris. Samuti on testitud kasutamist mobiilis ning kõikides testitud seadmetes ja brauserites toimib süsteem tõrgeteta.

Töö autor testis ka süsteemi erinevate kasutajatega valides erinevaid filme ja hinnates filme erinevalt ning vastavalt valikutele tulid ka erinevad filmisoovitused. Kui erinevate testkasutajatega valida samad filmid OMA KAVAsse ja hinnata samasid filme täpselt samade hinnetega, siis tulevad ka täpselt samad soovitused, mida võiski oodata. Näiteks kui hinnati filmi hindegaga 5, mille märksõnadeks olid „põnevus“, „õudus“, „fantaasia“ ja „loodus“ ning teist filmi hindegaga 2, mille märksõnadeks olid „elulood“, „perefond“ ja „loodus“, siis kasutajale soovitati nende põhjal filme, mille märksõnadeks olid „põnevus“, „õudus“ ja / või „fantaasia“. Testimise käigus selgus, et filmisoovituste mootor töötab korrektselt ja töös kasutatud vahendid on selle jaoks sobilikud.

## 5.2 Soovitusmootori kasutamise statistika

2018. aastal toimus PÖFF 16. novembrist kuni 2. detsembrini ehk 17 päeva (~2,5 nädalat). Statistika on kogutud kasutades Google Analytics'it ning siin on välja toodud mõningad arvud ajavahemikust 1. september kuni 31. detsember:

- Lehe külastajate/isikute arv – 116 680
- Lehe külastamiste arv – 2 035 959
- Külastajate riike - 157
- Kontosid loodud - 3827
- OMA KAVA koostajaid - 1977
- Filme hinnanud - 381
- Sarnaste filmisoovituste peale klikatud – 10 399 korda
- Personaalsete filmisoovituste peale klikatud - 705 korda

Nagu statistika andmetest näha, siis iga filmi all olevatele sarnastele filmidele on vajutatud peaaegu 10.5 tuhat korda, mis on päris suur arv. Personaalsete soovitude peale on vajutatud aga tunduvalt vähem – ainult 700. Sarnaste filmide kuvamine oli nähtaval juba üsna festivali alguses, kuid personaalsed soovitud sai lisatud alles festivali keskel. Sarnaste filmide nägemiseks ei pea olema sisse logitud ja ei pea hindama filme ega lisama filme OMA KAVAsse ega ostma pileteid. Seetõttu on ka personaalsete soovitude peale vajutatud tunduvalt vähem.

Statistikast võib veel näha, et kontosid on tehtud peaaegu 4000, kuid OMA KAVA on neist koostanud ainult pooled. Filme on hinnanud ainult ligi 400 kasutajat ehk 10% kõikide kasutajate hulgast ning 20% OMA KAVA koostajatest. Sellest võib ka järeldada, et algselt oletatud asjaolud peavad paika, et OMA KAVA ei koosta kõik kasutajad ning filme hindab veel väiksem hulk inimesi.



## 6 Võimalikud edasiarendused

Kuna soovitusmootor on üsna mahukas projekt, siis on autoril mitmeid ideid, kuidas seda edasi arendada.

Esiteks, võiks juurde arendada koostööpõhise mudelil põhineva andmete filtreerimise. See tähendab, et kasutajatele soovitatakse filme, mis teistele kasutajatele on meeldinud. Selle abil saaks süsteem koostada mustri, mille põhjal saab teada, mis filmid millistele klientidele meeldida võiks.

Lisaks võiks süsteem arvesse võtta ka seda, mitu korda mõni klient teatud filmi peale vajutanud on. Kui klient on mitu korda teatud filmi uurinud, siis järelkult pakub see film talle huvi ja need filmid, millele ta on ainult üks või kaks korda peale vajutanud, ei paku huvi. See lähenemisviis ei nõuaks kliendilt filmide hindamist ega märksõnade valikut ning kasutaja ei pea selle jaoks lisaliigutust tegema. Negatiivse poole peale võib see aga ebatäpseid tulemusi anda, sest üldiselt ei saa sellise käitumise korral 100 protsendiliselt kindel olla, kas kasutajale see film meeldib või mitte.

Kuna filmisoovituste mootor on juba ühe festivali ajal kasutusel olnud, siis saab järgmiseks festivaliks juurde lisada funktsionaalsuse, mis soovitab kasutajale filme tema eelmise aasta valikute põhjal. Selle lisamine on mainitud edasiarenduste variantidest kõige lihtsam ning praegu olemasolevale koodile on vaja juurde kirjutada ainult mõned read ja eelmise aasta andmed järgmise aasta andmetest eraldi hoida.

## 7 Kokkuvõte

Käesolevas töö eesmärgiks oli arendada ettevõtte MTÜ Pimedate Ööde Filmifestival kodulehele filmisoovituste mootor, mis tooks ettevõttele rohkem kliente ning aitaks veebilehe kasutajatel paremini enda jaoks sobivaid filme valida.

Eesmärkide saavutamiseks pandi kõigepealt kirja ettevõtte vajadused ja nõuded, et neist lähtuvalt oli võimalik süsteemi looma hakata. Esmalt loodi filmidele hinde andmise võimalus. Seejärel koostati sarnaste filmide kuvamine märksõnade, žanrite ja programmide abil. Järgmiseks mõtles autor välja algoritmi, mille põhjal saaks filmidele antud hinnangutest teha filmisoovitusi. Kui need kõik olid tehtud, siis ühendas töö autor OMA KAVAs olevad filmid, ostetud piletid ja antud hinnangud soovitusmootoriga.

Töö käigus valmis filmisoovituste mootor, mis analüüsib andmeid peaaegu reaalajas ning filtreerib saadud infot sisupõhiselt. Veebilehele sisse logimata kasutajad näevad ainult filmidega sarnaseid filme ning sisse loginud kasutajad saavad filme hinnata ning näevad lisaks sarnastele filmidele ja personaalseid filmisoovitusi, mis on tehtud filmidele antud hinnangute, OMA KAVAs olevate filmide ja ostetud piletite põhjal.

Autor testis loodud süsteemi Google Chrome'is, Firefoxis, Edge'is ja Safari brauseritega ning proovis saada erinevaid filmisoovitusi kasutades selleks erinevaid testkontosid, erinevaid filme OMA KAVAsse lisades ja andes filmidele erinevaid hindeid. Kõik toimis vastavalt ettevõtte nägemusele ning süsteem avalikustati PÖFFi veebilehel. Seega võib öelda, et püstitatud eesmärk sai täidetud.

Isikliku arvamusena oli antud tööd väga huvitav teha, sest sain teada kuidas soovitusmootorid töötavad ja uurimise käigus selgus ka kui palju neist võib ettevõtete ja klientide jaoks kasu olla. Töö autoril on plaanis loodud filmisoovituste mootorit vastavalt käesoleva töö peatükis 6 nimetatud edasiarendusi realiseerida.

## Kasutatud kirjandus

- [1] „FIAPF,“ Wikipedia, 1 2018. [Võrgumaterjal]. Available: <https://en.wikipedia.org/wiki/FIAPF>. [Kasutatud 6 5 2018].
- [2] „XML,“ Wikipedia, 7 2017. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/XML>. [Kasutatud 5 5 2018].
- [3] „PHP,“ Wikipedia, 10 2017. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/PHP>. [Kasutatud 5 5 2018].
- [4] „HTML,“ Wikipedia, 12 2017. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/HTML>. [Kasutatud 5 5 2018].
- [5] „Dokumendi objektimudel,“ Wikipedia, 2 2014. [Võrgumaterjal]. Available: [https://et.wikipedia.org/wiki/Dokumendi\\_objektimudel](https://et.wikipedia.org/wiki/Dokumendi_objektimudel). [Kasutatud 7 5 2018].
- [6] „CSS,“ Wikipedia, 5 2018. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/CSS>. [Kasutatud 5 5 2018].
- [7] „SASS (stiililehe keel),“ Wikipedia, 5 2018. [Võrgumaterjal]. Available: [https://et.wikipedia.org/wiki/Sass\\_\(stiililehe\\_keel\)](https://et.wikipedia.org/wiki/Sass_(stiililehe_keel)). [Kasutatud 5 5 2018].
- [8] „JavaScript,“ Wikipedia, 2 2018. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/JavaScript>. [Kasutatud 5 5 2018].
- [9] „jQuery,“ Wikipedia, 5 2018. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/JQuery>. [Kasutatud 5 5 2018].
- [10] „Struktuurpäringukeel,“ Wikipedia, 10 2016. [Võrgumaterjal]. Available: <https://et.wikipedia.org/wiki/Struktuurpäringukeel>. [Kasutatud 5 5 2018].
- [11] MTÜ Pimedate Ööde Filmifestival, „PÖFF'i koduleht,“ 2018. [Võrgumaterjal]. Available: <https://2018.poff.ee>. [Kasutatud 26 4 2018].
- [12] C. M. S. N. Ian Mackenzie, „How retailers can keep up with consumers,“ McKinsey & Company, 10 2013. [Võrgumaterjal]. Available: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>. [Kasutatud 30 4 2018].
- [13] P. Potter, „Süsteemi nõuete esiletoomine ja analüüs,“ 9 2012. [Võrgumaterjal]. Available: <http://maurus.ttu.ee/sts/wp-content/uploads/2012/09/Süsteemi-nõuete-esiletoomine-ja-analüüs.pdf>. [Kasutatud 1 5 2018].
- [14] „Using Machine Learning on Compute Engine to Make Product Recommendations,“ Google Cloud, 19 3 2018. [Võrgumaterjal]. Available: <https://cloud.google.com/solutions/recommendations-using-machine-learning-on-compute-engine>. [Kasutatud 22 4 2018].
- [15] R. Banik, „Recommender Systems in Python: Beginner Tutorial,“ DataCamp, 16 1 2018. [Võrgumaterjal]. Available: <https://www.datacamp.com/community/tutorials/recommender-systems-python>. [Kasutatud 15 5 2018].

- [16] J. Le, „The 4 Recommendation Engines That Can Predict Your Movie Tastes,“ Medium, 22 4 2018. [Võrgumaterjal]. Available: <https://towardsdatascience.com/the-4-recommendation-engines-that-can-predict-your-movie-tastes-109dc4e10c52>. [Kasutatud 14 5 2018].
- [17] H. Gaspar, „The Cold Start Problem for Recommender Systems,“ Yuspify, 14 7 2015. [Võrgumaterjal]. Available: <https://www.yuspify.com/blog/cold-start-problem-recommender-systems/>. [Kasutatud 17 6 2018].
- [18] „Mobile Browser Market Share Worldwide,“ statcounter, 4 2018. [Võrgumaterjal]. Available: <http://gs.statcounter.com/browser-market-share/mobile/worldwide>. [Kasutatud 15 5 2018].
- [19] Microsoft, „Support for older versions of Internet Explorer ended,“ Microsoft, [Võrgumaterjal]. Available: <https://www.microsoft.com/en-us/windowsforbusiness/end-of-ie-support>. [Kasutatud 20 5 2018].
- [20] „Browser & Platform Market Share,“ W3Counter, 4 2018. [Võrgumaterjal]. Available: <https://www.w3counter.com/globalstats.php>. [Kasutatud 15 5 2018].