Rīgas 64. Vidusskola

**Google Chrome Paplašinājums: Sentimenta Nolasīšana No YouTube Video Komentāru Sadaļām**

Programmatūras prasību specifikācija

**Darba autori:**

Kristers Laganovskis, Kristaps Bukšs

**Darba vadītājs:**

Edvards Bukovskis

Rīga 2023

Saturs

[Definīcijas, akronīmi un saīsinājumi 3](#_Toc158324738)

[1. Ievads 4](#_Toc158324739)

[1.1 Nolūks 4](#_Toc158324740)

[1.2. Darbības sfēra 4](#_Toc158324741)

[1.3. Saistība ar citiem dokumentiem. 4](#_Toc158324742)

[1.4. Pārskats 4](#_Toc158324743)

[2. Problēmas izpēte un analīze 5](#_Toc158324744)

[2.1. Aktualitāte 5](#_Toc158324745)

[2.2. Pētāmā problēma un mērķauditorija 5](#_Toc158324746)

[2.3. Produkta funkcijas: 6](#_Toc158324747)

[2.6. Front-end paplašinājuma skice: 6](#_Toc158324748)

[3. Programmatūras prasību specifikācijas 7](#_Toc158324749)

[3.1. Funkcionālās prasības: 7](#_Toc158324750)

[3.2. Nefunkcionālās prasības: 8](#_Toc158324751)

[3.3. Datu plūsmas shēma: 9](#_Toc158324752)

# Definīcijas, akronīmi un saīsinājumi

|  |  |
| --- | --- |
| **Saīsinājums, termins** | **Skaidrojums** |
| AL | Augstākā līmeņa |
| lietotājs | Persona, kas izmanto programmatūru |
| m.g. | Mācību gads |
| saskarne | Funkciju kopa, kas nodrošina datu apmaiņu starp lietotāju un programmatūru. |
| SNYVKS | Sentimenta Nolasīšana No YouTube Video Komentāru Sadaļām |
| back-end | Programmu kods, kas nav pieejams un ko nevar redzēt lietotāji, tas nodrošina datu apstrādi un uzglabāšanu. |
| VS code | Visual Studio Code - integrētā izstrādes vide |
| IDE | Integrētā izstrādes vide (Integrated Development Environment) |
| front-end | Programmu sastāvdaļa, kas tiek tieši attēlota un ar kuru tieši mijiedarbojas lietotāji. |
| RoBERTa | ”Robustly Optimized BERT Pretraining Approach” sentimenta nolasīšanas modelis. |
| VADER | ”Valence Aware Dictionary for Sentiment Reasoning” sentimenta nolasīšanas modelis. |
| VPS | Virtuālais privātais serveris |
| Ubuntu | operētājsistēma, kas balstās uz Linux un ir plaši izmantota serveru vidē. |
| web | globālais tīmeklis |
| API | Lietotāja programmas saskarne, kas ļauj citām programmām vai sistēmām komunicēt vai integrēties ar programmu. |

# Ievads

## 1.1 Nolūks

Šodienas informācijas plūsma rada izaicinājumus gan saturu patērētājiem, gan veidotājiem, jo cilvēki bieži uzticas subjektīviem satura veidotājiem, kas var manipulēt ar informāciju. Īpaši "YouTube" platformā veidotie video bieži satur subjektīvus viedokļus, kas var tikt uzskatīti par faktiem. Lai risinātu šo problēmu, izstrādāts Google Chrome paplašinājums, kas veic automātisku sentimenta analīzi desmitiem, simtiem vai pat tūkstošiem orģinālu komentāru konkrētā "YouTube" video sadaļā. Paplašinājums sniedz lietotājiem iespēju objektīvāk novērtēt informāciju un novērst viltus informācijas izplatīšanu, uzlabojot patērētāju drošību interneta vidē.

## 1.2. Darbības sfēra

Chrome paplašinājums ir izstrādāts kā publiski pieejama aplikācija, kuru var lietot ikviens, kam ir interese vai nepieciešamība izmantot funkcijas, ko šī programmatūra piedāvā. Paplašinājuma darbības sfēra ir padarīt viegli pārskatāmu lielu informācijas daudzumu. Šis paplašinājums ne tikai novērš potenciālu viltu ziņu izplatīšanu, bet arī nodrošina iespēju viegli pārskatīt kāda konkrēta YouTube video komentāru sadaļu. Komentārus paplašinājums iedala pēc emocijām – pozitīvi, neitrāli un negatīvi –, tādējādi veicinot skaidrību un pārredzamību analīzē

## 1.3. Saistība ar citiem dokumentiem.

1. Dokumenta PPS noformēšanā ievērotas standarta LVS 68:1996 prasības.

2. Dokumenta PPS izveidē ievēroti 2023./2024. m.g. centralizētā eksāmena programmas kritēriji augstākajam mācību satura apguves līmenim programmēšanā.

3. Dokumenta PPS kalpo kā atsauce turpmākiem sistēmas uzlabojumiem

## 1.4. Pārskats

1. nodaļa: Šajā nodaļā tiek piedāvāta sākotnējā informācija un vispārējs apraksts, ietverot dokumenta nolūku, definīciju skaidrojumu, savstarpējo saikni ar citiem dokumentiem un sistēmas darbības jomas aprakstu.

2. nodaļa: Otrajā nodaļā notiek problēmas izpēte un analīze.

3. nodaļa: Trešajā nodaļā tiek detalizēti aprakstītas programmatūras funkcionālās un nefunkcionālās prasības.

4. nodaļa: Ceturtā nodaļa piedāvā programmatūras izveides plānu.

5. nodaļa: Piektajā nodaļā notiek programmatūras atkļūdošana un testēšana.

6. nodaļa: Sestajā nodaļā tiek izstrādāts lietotāja ceļvedis.

7. nodaļa: Septītajā nodaļā tiek sniegts apraksts par piemērojamās licences pamatojumu.

# 2. Problēmas izpēte un analīze

## 2.1. Aktualitāte

Šodienas informācijas plūsma strauji pieaug, radot izaicinājumus gan saturu patērētājiem, gan veidotājiem. Cilvēki bieži uzticas satura veidotājiem, kuri ne vienmēr ir objektīvi. Tas rada risku, ka veidotājs var manipulēt ar informāciju, ietekmējot patērētājus. Tāpēc ir svarīgi apsvērt arī citu cilvēku viedokļus par konkrēto saturu kādā konkrētā video, it īpaši platformā "YouTube", kur satura veidotāji bieži pauž subjektīvus viedokļus, ko patērētāji var uzskatīt par faktiem.

## 2.2. Pētāmā problēma un mērķauditorija

Pētāmā problēma ir saistīta ar subjektīvu informācijas nodošanu un tās uztveres risku, it īpaši "YouTube" platformā, kur satura veidotāji bieži pauž savus viedokļus, kas var ietekmēt patērētājus kā faktus. Pētījuma mērķis ir izveidot risinājumu šai problēmai, piedāvājot Google Chrome paplašinājumu, kas veic automātisku sentimenta analīzi orģinālu komentāru sadaļā konkrētā "YouTube" video.

Paplašinājuma mērķis ir atvieglot lietotājiem ne tikai pirkumus internetā, bet arī sniegt iespēju objektīvāk izvērtēt informāciju, ko piedāvā video veidotāji. Tas palīdz novērst viltus informācijas izplatīšanu un uzlabot patērētāju drošību interneta vidē, piedāvājot objektīvākus viedokļus un pamatojumu lietotājiem, kuri grib veikt informētus secinājumus par saturu.

Mērķauditorija veidotajai programmatūrai ir:

* Skolēni un studenti, jo šī vecuma grupas cilvēki ir pakļauti vislielākajam riskam tikt manipulētiem ar nepatiesu informāciju. Berni bieži uztver pat nepatiesu informāciju par faktiem.
* Šī auditorija ir būtiska, jo tie, kuri regulāri veic pirkumus tiešsaistē, ir pakļauti dažādām reklāmām un ietekmējošiem faktoriem, tostarp "YouTube" video saturam. Paplašinājums sniedz iespēju šiem lietotājiem padziļināti novērtēt ne tikai produktus, bet arī saistītos komentārus un viedokļus, palīdzot tiem pieņemt informētākus lēmumus par iepirkumiem.
* Ikviens interneta lietotājs: Jo īpaši tie, kuri bieži lieto "YouTube" platformu. Paplašinājums piedāvā visiem iespēju veikt objektīvākas secinājumus par saturu, un samazināt viltus informācijas ietekmi.

Lai izveidotu produktu, kas pēc iespējas vairāk līdzinātos minētajās mērķauditorijas redzējumam un vajadzībām, nepieciešam aptauja vai datu analīze, lai noskaidrotu tipiskākās problēmas vai neērtības šajā jomā, kas programmatūrai ir jāpanāk, lai atrisinātu šo problēmu vai apmierinātu lietotājus, piemēram, aptaujā tiktu iekļauti šādi jautājumi:

* “Cik bieži jūs lasāt, vai pievēršat uzmanību komentāriem "YouTube" video sadaļās, it īpaši, ja tiek izteikti subjektīvi viedokļi?”
* “Vai esat iegādājies produktu vai pakalpojumu, pamatojoties uz informāciju, kas atrasta "YouTube" video komentāros?”
* “Vai jums ir bijusi pieredze ar viltus informāciju vai maldinošu saturu kādā "YouTube" video?”

Šie jautājumi varētu palīdzēt iegūt sīkāku izpratni par mērķauditorijas uzskatiem un vajadzībām, lai veidotu paplašinājumu, kas efektīvi risina problēmu un sniedz vērtību lietotājiem. Kā arī varētu, veikt datu analīzi saistībā, kas izpēta, cik uzticīgi ir YouTube satura veidotāji.

## 2.3. Produkta funkcijas:

Šī paplašinājuma galvenā funkcija ir rast iespēju lietotājam efektīvi filtrēt video komentāru sadaļas YouTube platformā. Šis paplašinājums sadala komentāru sadaļu trīs klasēs – negatīvi, neitrāli un pozitīvi. Kur pēc tam lietotājs var iet cauri un attiecīgi novērtēt, kādus viedokļus komentētāji izsaka. Kā arī paplašinājums parādīs procentuāli, cik pozitīvi, neitrāli un negatīvi kopumā ir konkrētā video komentāru sadaļa.

2.4. Vispārējie ierobežojumi:

Šī produkta ierobežojumi ir dažādi, taču tos iespējams, ka varētu novērst ar laiku. Viens no ierobežojumiem ir sarkasms. Ar šo problēmu saskaras ikviens sentimenta nolasīšanas modelis, jo tie nespēj uztvert sarkasmu, tādējādi rezultātā tas atgriež viltus pareizas atbildes. Otrais ierobežojums ir “emoji” lietojums komentāru sadaļās, jo šo arī nespēj noteikt sentimenta modelis. Taču šos ierobežojumus ir iespējams novērst vismaz daļēji ieviešos sarkasma nolasītāju un “emoji” emociju nolasītāju. Trešais ierobežojums ir pats sentimenta modelis, lai kādu sentimenta modeli izmantos, tā precizitāte nebūs 100%, tāpēc var paredzēt, ka būs pāris nepareizi nolasīti noskaņojumi. Visbeidzot, apstrādāšanas ātrums – tas noteikti ir viens no ierobežojumiem, jo ne visi sentimentu modeļi spēj nolasīt sentimentu ātri. Tā nav liela problēma, ja tie ir vairāk nekā 200 komentāru, bet, ja tie ir mazāk par 50 komentāriem, tad izdevīgāk ir pašam iziet cauri komentāru sadaļai. Šo problēmu varētu novērst ar jaudīgu datoru un labu internetu, kas paātrinātu datu sagatavošanas un nolasīšanas soļus.

## 2.6. Front-end paplašinājuma skice:

A diagram of a computer

Description automatically generated with medium confidence

1. Attēls: Front-end paplašinājuma skice, kurā tiek parādīts kā paplašinājums varētu vizuāli izskatīties.)

# Programmatūras prasību specifikācijas

Programmatūras produkta mērķauditorijai ir nepieciešams rīks, kas spēj analizēt sentimentu lietotāja izvēlētam YouTube video komentāru sadaļai. Šim paplašinājumam ir jābūt iespēja attēlot, cik procentuāli ir katra kategorija komentāru sadaļa, piemēram – “Negatīvi: 8%, Neitrāli: 21% un Pozitīvi: 71%”. Paplašinājumam ir jānodrošina interaktīva vide, lietotājam ir jābūt iespējai nolasīt sentimentu, lietotājam ir jābūt iespējai izvēlēties kādu no trīs kategorijām - negatīvi, neitrāli, pozitīvi. Kur lietotājam ir iespēja interaktīvi “skrullēt” cauri atlasītajiem komentāriem, kā arī pēc tam atgriezties atpakaļ galvenajā izvelmē.

SNYVKS ir Chrome paplašinājuma aplikācija, kas rada iespēju novērst viltu ziņu izplatīšanu, paātrina informācijas uztveres ātrumu, sadala ērti un uzskatāmi komentāru sadaļu.

SNYVKS tiks izstrādāts uz VS code IDE. Sentimenta noteikšanā tiks izmantoti divi modeļi VADER un RoBERTa, kura tiks salikta kopā un veidos hibrīda modeli. VADER ir sentimenta nolasīšanas modelis, kas darbojas uz leksikonu bāzes, kamēr RoBERTa ir balstīta uz mašīnmācīšanās algoritmiem. Šie abi modeļi tiks izmantoti, lai iegūtu pēc iespējas precīzākus rezultātus. Back-end kods tiks rakstīts Python 3.8 programmēšanas valodā, jo tā ir salīdzinoši ātra valoda, kā arī ērti un plaši pielietojama. Kamēr paplašinājuma izveidei (front-end) tiks izmantotas valodas kā HTML 5, CSS un JS (paplašinājuma back-end daļai), jo šīs valodas tiek izmantotas, lai veidotu web platformu interfeisu.

Back-end koda daļa tiks uzstādīta un uz ORACLE piedāvātā VPS servisa, nodrošinot stabilu un ilgstošu darbību. Šis VPS serviss tiks izvēlēts, lai nodrošinātu augstu pieejamību un elastīgumu, kas nepieciešams, lai apkalpotu lielu apmeklētāju skaitu un nodrošinātu nepārtrauktu pakalpojumu sniegšanu. Šajā VPS tiks izmantota Linux Ubuntu 20 operētājsistēma, kas ir plaši izplatīta un piedāvā stabilu un drošu darbību. Tas nodrošinās optimālu vidi back-end koda darbībai un vispārējam projekta veiksmīgam darbībai.

## 3.1. Funkcionālās prasības:

SNYVKS galvenā pamatfunkcija ir nodrošināt lietotājam vieglu un ērti lietojamu paplašinājumu, kas spēj nolasīt sentimentu no lietotāja izvēlēta YouTube video komentāru sadaļas. Paplašinājumam ir piecas pamata funkcionālās prasības interfeiss (front-end) daļai:

1. Spēja iegūt datus no izvēlētā video. Paplašinājumam jānodrošina, kā tas nolasa datus ar API no konkrēti izvēlētā YouTube video.
2. Spēja attēlot procentuālo sadalījumu, nodrošinot skaidru un saprotamu informāciju lietotājam.
3. Spēja lietotājam izvēlēties vēlamo kategoriju – negatīvi, neitrāli vai pozitīvi.
4. Lietotājam jābūt iespējai ērti un vienkārši pārlūkot komentārus, izmantojot “skrullēšanas” funkciju vai citas navigācijas iespējas.
5. Lietotājam jābūt ērtai un saprotamai iespējai pārveidoties starp dažādām paplašinājuma funkcijām un darbības režīmiem.

Kā arī ir savas funkcionālās prasības paplašinājuma back-end daļai:

1. Datu saņemšana un nolasīšana no API.
2. Katram komentāram tiek noteikta valoda.
3. Ja komentārs nav angļu valodā, tam ir jābūt pārtulkotam uz angļu valodu.
4. Komentāri tiek apstrādāti jeb sagatavoti sentimenta nolasīšanas modelim.
5. Tiek nolasīts sentiments no katra komentāra.
6. Tiek iedalīti komentāri attiecīgajās kategorijās pēc tā sentimenta.
7. Tiek izrēķināti, cik procentuāli ir katra kategorija komentāru sadaļā.
8. Tiek padoti atpakaļ iegūtie dati uz front-end paplašinājumu caur API.

Back-end daļas shēma:

A diagram of a software company

Description automatically generated with medium confidence

(2. Attēls: Back-end daļas shēma, kurā tiek attēlota datu plūsma caur back-end daļu.)

## 3.2. Nefunkcionālās prasības:

1. **Efektivitāte un veiktspēja**: Paplašinājumam jāspēj efektīvi apstrādāt lielu datu apjomu, lai nodrošinātu neierobežotu komentāru skaitu.
2. **Drošība**: Paplašinājumam jānodrošina droša datu pārraide starp klienta un servera pusi, lai novērstu jebkādas datu noplūdes vai manipulācijas riskus. Jānodrošina, ka izmantotās sentimenta nolasīšanas metodes un algoritmi ir uzticami un neatklāj lietotāju datus vai citu konfidenciālu informāciju.
3. **Lietojamība un interfeiss**: Interfeisam jābūt viegli uztveramam, lai lietotājiem nebūtu grūtību izmantot paplašinājumu. Jānodrošina, ka interfeiss ir atbilstoši pielāgots dažādām ekrāna izmēru un izšķirtspēju ierīcēm, lai nodrošinātu labu lietojamību visās situācijās.
4. **Uzticamība un atbalsts**: Paplašinājumam jābūt stabiliem un uzticamiem, lai lietotāji varētu paļauties uz to nepārtrauktu darbību. Jānodrošina nepārtraukts atbalsts un regulāras atjaunināšanas, lai novērstu jebkādas iespējamās kļūdas vai drošības trūkumus.
5. **Pārnesamība**: Paplašinājumam jābūt pārnēsājamam un atbalstītam dažādās Google Chrome versijās, kā arī dažādās operētājsistēmās, kurās darbojas Google Chrome.

Šīs nefunkcionālās prasības nodrošina, ka paplašinājums darbojas efektīvi un droši, lietotājam nodrošinot labu lietojamību un uzticamību.

## 3.3. Datu plūsmas shēma:

A diagram of a square with arrows

Description automatically generated

(3. Attēls: Datu plūsma starp paplašinājumu, serveri un back-end kodu.)