

TALLINNA TEHNIKAÜLIKOOL  
Infotehnoloogia teaduskond

Mirjam Pajumägi 192170IAPM

# **Koostöösobilike ettevõtete kaevandamine**

Magistritöö

Juhendaja: Ahti Lohk  
PhD

Tallinn 2021

## **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: Mirjam Pajumägi

10.05.2021

## Annotatsioon

Käesoleva magistritöö eesmärgiks on luua ettevõttele Concise Systems OÜ lahendus, mis aitaks ettevõttel vähendada inimtöötundide hulka, mis kulub hetkel potentsiaalsete uute koostööpartnerite leidmiseks. Tulem aitaks Concise'l laiendada oma kliendibaasi, pakkudes kvaliteetset arendusteenust klientidele, kelle väärtused ja hoiakud peegeldavad Concise'i omasid.

Antud eesmärgi saavutamiseks on ReactJS raamistiku ja TypeScript programmeerimiskeele baasil loodud Chrome laiendus, mis võimaldab kasutajal luua seadistusi sõltuvalt veebilehekülje struktuurist. Seadistuste abil tuvastatakse veebileheküljel olevaid tekstilisi andmeid. Tuvastatud tekstilõikude peal rakendatakse VADER sentimentaalanalüüsi, misjärel kuvatakse kasutajale tekstilõigu meelsushinnang normaliseeritud kujul. Kasutajal on võimalik leitud tekstilõike hilisemaks inimanalüüsiks salvestada.

Selgus, et loodud lahendus ei lahenda täielikult Concise Systems OÜ probleemi. Soovitud tulemi saavutamiseks oleks tarvis kasutada võimekamat tekstianalüüsi algoritmi, mis lisaks meelsushinnangule tuvastaks tekstides ka sisulist informatsiooni. Käesoleva magistritöö raames jäi eesmärk saavutamata ning probleem lahendamata. Samas on lahendus unikaalne ning sellele lisa-arendusi implementeerides võib sellest tulevikus rohkem kasu olla.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 45 leheküljel, 5 peatükki, 35 joonist, 0 tabelit.

## **Abstract**

### **Mining of Cooperative Companies**

In recent times, more and more companies realize the importance of a long-lasting and fruitful partnerships where both parties are content and happy with said partnership. There have been several studies for maintaining relationships after the partnership contract has been signed. However, a successful partnership begins with one party doing extensive research of potential partner *before* joining hands.

This is what Concise Systems OÜ is also doing. The team that is responsible for finding new partners spends a vast amount of time and other resources in order to land a partnership that is enjoyable for both sides. Therefore, a question whether this kind of research could be automated has risen.

The aim of this thesis is to find an answer to that question and create a custom-made application to lessen the manhours Concise uses for research today.

To achieve this, a ReactJS and TypeScript based Chrome Extension was created. The implemented solution allows user to create configurations for webpages that are frequently used for research purposes. Those configurations are based on webpages' internal structure and class names used in HTML elements. Constructed configurations can later be used to find elements from webpage that have matching class names with the ones saved in configuration. Texts contained in found elements are assessed sentiment-wise using VADER sentiment analysis, giving the user an overview of the sentiments connected to possible partner candidate. User can also save more thought-provoking texts for later usage.

It turned out that created solution does not fully solve Concise Systems OÜ's problem as more sophisticated text-analysis algorithm needs to be used. The created solution gives a good overview of the image that a potential partner has left but does not provide a solution that would help to collect factual data about the potential candidate. Therefore, the aim of this thesis was not achieved. However, the approach used for creating the solution was

said to be unique and if some extra development is done in the future, the tool created could help Concise Systems OÜ to solve their problem.

The thesis is in Estonian and contains 45 pages of text, 5 chapters, 35 figures, 0 tables.

## Lühendite ja mõistete sõnastik

|       |   |
|-------|---|
| API   | <i>Application Programming Interface</i> , rakendusliides ehk programmiliides ehk rakendustarkvara liides     |
| CRM   | <i>Customer Relationship Management</i> , kliendisuhete haldamine   |
| CSV   | <i>Comma-Separated Values</i> , komaga eraldatud väärtused, failiformaat tabeliandmete hoiustamiseks          |
| DACH  | <i>Germany (D) – Austria (A) – Switzerland (CH)</i> , ühisnimetus Saksamaa-Austria-Šveits ühiskonnaruumile    |
| IDE   | <i>Integrated Development Environment</i> , integreeritud arenduskeskkond                                     |
| MVP   | <i>Minimum Viable Product</i> , vähim elujõuline toode, prototüüp   |
| NER   | <i>Named Entity Recognition</i> , nimeolemite tuvastamine   |
| SVG   | <i>Scalable Vector Graphics</i> , skaleeritav vektorgraafika  |
| URL   | <i>Uniform Resource Locator</i> , internetiaadress, üldine infoallika asukohamääraja                          |
| UUID  | <i>Universally unique identifier</i> , universaalne unikaalne identifikaator                                  |
| VADER | <i>Valence Aware Dictionary for sEntiment Reasoning</i> , valentsiteadlik sõnaraamat meelsuste põhjendamiseks |

## Sisukord

|  |    |
|--|----|
| 1 Sissejuhatus .....   | 10 |
| 1.1 Probleem.....  | 10 |
| 1.2 Töö eesmärk .....  | 11 |
| 2 Metoodika.....   | 13 |
| 2.1 Loodava lahenduse kirjeldus .....  | 13 |
| 2.2 Lahenduse loomiseks kasutatavad raamistikud ja tööriistad.....                                     | 15 |
| 2.2.1 Tekstitöötlus .....  | 15 |
| 2.2.2 Chrome laiendus.....   | 18 |
| 2.3 Protsessi kirjeldus.....   | 20 |
| 2.3.1 Ettevõtte kohta informatsiooni analüüsimine.....   | 20 |
| 3 Lahenduse tehniline dokumentatsioon .....  | 21 |
| 3.1 Eeltöö.....  | 21 |
| 3.2 Chrome laiendus .....  | 22 |
| 3.2.1 Seadistuse loomine .....   | 26 |
| 3.2.2 Meelsusanalüüsi teostamine .....   | 32 |
| 3.2.3 Informatsiooni salvestamine ja kuvamine .....  | 35 |
| 4 Analüüs ja järeldused.....   | 41 |
| 4.1 Autori üldised tähelepanekud .....   | 41 |
| 4.2 Autori tähelepanekud valminud lahenduse funktsionaalsuse ja käitumise osas... 43                   |    |
| 4.3 Concise Systems OÜ tähelepanekud .....   | 50 |
| 4.4 Järeldused tähelepanekutest.....   | 54 |
| 5 Kokkuvõte .....  | 55 |
| Kasutatud kirjandus .....  | 56 |
| Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks ..... | 59 |

## Jooniste loetelu

|   |    |
|---|----|
| Joonis 1. Parempoolse tööriistariba kuvatõmmis. ....  | 23 |
| Joonis 2. Vasakpoolse tööriistariba kuvatõmmis. ....  | 24 |
| Joonis 3. Seadistuste JSON kuju programmikoodi näide. ....  | 25 |
| Joonis 4. Seadistuste TypeScripti tüübi programmikoodi näide. ....                                  | 25 |
| Joonis 5. Sõnumi <i>content</i> 'ilt <i>background</i> 'ile saatmise programmikoodi näide. ....     | 26 |
| Joonis 6. Sõnumi <i>background</i> 'ilt <i>content</i> 'ile saatmise programmikoodi näide. ....     | 26 |
| Joonis 7. Sõnumi <i>content</i> 'is ja <i>background</i> 'is vastuvõtmise programmikoodi näide. ... | 26 |
| Joonis 8. Laienduse sisselülitamisel kuvatavate tööriistaribade kuvatõmmis. ....                    | 29 |
| Joonis 9. Teksti selekteerimise kuvatõmmis. ....  | 29 |
| Joonis 10. Selekteeritud teksti märgendamise ja HTML klassistruktuuri kuvamise<br>kuvatõmmis. ....  | 29 |
| Joonis 11. Sobiva seadistuse märgendamise kuvatõmmis. ....  | 30 |
| Joonis 12. Valiidse seadistuse kuvatõmmis. ....   | 31 |
| Joonis 13. Loodud seadistuse parempoolsel tööriistaribal kuvamise kuvatõmmis. ....                  | 32 |
| Joonis 14. Laienduse komponentide kuvamise ja peitmise <i>expand/hide</i> nupu<br>kuvatõmmis. ....  | 32 |
| Joonis 15. Meelsusanalüüsi üldhinnangu tekstilõikudel kuvamise kuvatõmmis. ....                     | 33 |
| Joonis 16. Lehekülje keskmise meelsusskoori kuvamise kuvatõmmis. ....                               | 34 |
| Joonis 17. Meelsusanalüüsi koondtulemused hajuvusdiagrammil kuvatõmmis. ....                        | 34 |
| Joonis 18. Teksti hilisemaks töötlemiseks salvestamise nupu kuvatõmmis. ....                        | 35 |
| Joonis 19. Teksti salvestamise hüpinkakna kuvatõmmis. ....  | 35 |
| Joonis 20. Huvipakkuva teksti TypeScripti tüübi programmikoodi näide. ....                          | 36 |
| Joonis 21. Salvestatud tekstilõikude üldvaate lehekülje kuvatõmmis. ....                            | 36 |
| Joonis 22. Chrome laienduse eraldi leheküljele pääsemise kuvatõmmis. ....                           | 37 |
| Joonis 23. Chrome laienduse hüpinkmenüü kuvatõmmis. ....  | 37 |
| Joonis 24. Ettevõtte kohta käivate salvestatud tekstilõikude kuvamise kuvatõmmis. ....              | 38 |
| Joonis 25. Tekstilõikude võtmesõna alusel filtreerimise kuvatõmmis. ....                            | 39 |
| Joonis 26. Salvestatud tekstilõigu informatsiooni muutmise kuvatõmmis. ....                         | 39 |
| Joonis 27. Salvestatud tekstilõigu kustutamise kinnitamise kuvatõmmis. ....                         | 40 |



|  |    |
|--|----|
| Joonis 28. Veebilehel facebook.com samade klassinimedega elementide hulga näite<br>kuvatõmmis. ....                              | 44 |
| Joonis 29. Veebilehel bbc.com ebapiisava täpsusega määratletud uudisartikli sisu<br>kuvatõmmis. ....                             | 45 |
| Joonis 30. Veebilehel news.crunchbase.com ebapiisava täpsusega määratletud<br>uudisartikli sisu kuvatõmmis. ....                 | 46 |
| Joonis 31. Veebilehel gizmodo.com heledal taustal tekstide meelsusanalüüsi<br>kuvatõmmis. ....                                   | 47 |
| Joonis 32. Veebilehel linkedin.com vasakpoolse tööriistariba kuvatõmmis. ....  | 48 |
| Joonis 33. VADER'i neutraalse hinnangu kuvamine inimese jaoks negatiivsele lausele<br>kuvatõmmis. ....                           | 49 |
| Joonis 34. Veebilehel play.google.com kasutajate hinnangu ning VADER<br>sentimentaalhinnangu vahelise ebakõla kuvatõmmis. ....   | 52 |
| Joonis 35. Veebilehel google.com Concise Systems OÜ poolt otsitava võimaliku<br>koostööpartneri meelsusanalüüsi kuvatõmmis. .... | 53 |

# 1 Sissejuhatus

Koostöö on mõiste, mis tänapäeva ühiskonnas omab üha suuremat mõju. See on oluline ka ettevõtete tasandil, kuna heast koostööst sünnivad lahendused, mis lihtsustavad paljude inimeste elu. Üha enam *startup*'ilikumaks muutuv maailmas on oluline leida enda ettevõtte kõrvale ettevõtteid, kellega saab teha kvaliteetset koostööd ning kelle väärtushinnangud on kooskõlas partnereid otsiva ettevõtte omadega. Oma koostööpartneriteks oodatakse pakkujaid, kes suudavad mõista, mida ettevõtte vajab ning mis on tema ootused ja soovid. Samas on väga olulisel kohal ka koostööst saadava tulemi kvaliteet.

Ettevõtetes on populaarne mõiste CRM (*Customer Relationship Management*), mis käsitleb juba olemasolevate partneritega suhtlust pikaajase koostöö eesmärgil. See on väga oluline ka IT-ettevõtetes. Ettevõtte juhtidele suunatud CRM'i teostamise juhendeid leiab juba 2002. aastast [1]. Tänapäeval on väga tavaline, et informatsiooni saamiseks kasutatakse interneti ja andmebaase ning jõudsalt turundust. Aastal 2004 oli see alles tõusev trend, mille edukus oli juba siis tajutav [2].

Tihti peale algab hea partnerlus siiski aga koostööeelsest tegevusest. Seetõttu tuleb ettevõttel, kes soovib enda toodet või teenust väljapoole pakkuda, enne selle tegemist teha suur hulk eeltööd, et leida partnerid, kellele pakkumise tegemine võib tuua kõige potentsiaalsemat kasu mõlemale osapoolele. See eeltöö sisaldab endas paljude andmete kogumist, potentsiaalselt väga mitmetest allikatest, ning kogutud andmete põhjal teostatud analüüsi. Analüüsi tulemused viivad otsusteni.

Ajastul, kus andmete hulgid internetis ulatuvad miljonite terabaititeni [3], muutub väga oluliseks vajaliku ja relevantse informatsiooni leidmine usaldusväärsetest allikatest. "Vajalik" on tihti peale pandud paika aga otsija subjektiivse hinnangu alusel.

## 1.1 Probleem

Infotehnoloogiliste valdkondade arengust lähtuvalt on ka Concise Systems OÜ-l tekkinud huvi võimalike lahenduste osas oma järgmisele probleemile. Nimelt kulutab ettevõtte

inimtööjõudu ja aega, et leida uusi võimalikke koostööpartnereid ning hinnata nende sobivust. Concise's on loodud n-ö *Growth Team* (ettevõtte kasvule orienteeritud meeskond), mis koosneb inimestest, kelle ülesandeks on leida uusi põnevaid ettevõtteid ja inimesi, kellele saaks Concise'i arendusteenuseid pakkuda. Selleks kulutatakse igapäevaselt palju aega, sest enne pakkumise tegemist tuleb potentsiaalsed ettevõtted kaardistada, uurida välja nende olukord ja murekohad ning selgitada välja, mida need ettevõtted arendusvaldkonnas vajavad ja millises mahus. Kaardistamisel võetakse tihti aluseks tehnoloogia-üritustel (nt WebSummit [4]) osalenute nimekiri, kuna sealsed osalevad ettevõtted otsivad tõenäolisemalt endale partnereid. Otsitava informatsiooni hulka kuuluvad näiteks ettevõtte poolt üles pandud töökuulutuste arv ja seal otsitavad positsioonid; finantsseis – kas on saadud investoritelt ressursse, millises mahus; kuidas on kajastus uudisvoogudel; mis regioonis ettevõtte tegutseb jpm. Otsus, kas teha pakkumine või mitte, tuleb paljude allikate uurimisest, leitud info analüüsimisest ning lõpuks pingerea koostamisest. Mida kõrgemal on ettevõtte pingereas, seda potentsiaalsemalt on see sobiv Concise'i poolt pakutavatele teenustele ning seda võimalikum on koostöö ja selle püsivus.

Päevakorda on kerkinud küsimus, kas antud protsessi oleks võimalik mingis mahus või täielikult automatiseerida ning kui see on võimalik, siis kuidas seda teha.

## **1.2 Töö eesmärk**

Antud magistritöö eesmärk on uurida välja, kas potentsiaalseid ettevõtteid, kellega koostööd teha, oleks võimalik kaardistada mingis mahus või täielikult automatiseeritult ning saadud informatsiooni põhjal luua erilahendusega programm või rakendus, mis aitaks Concise'l efektiivsemalt välja valida ettevõtteid, kellega kontakteeruda võimaliku koostöö tarvis. Tulem aitaks Concise'l laiendada oma kliendibaasi, pakkudes kvaliteetset arendusteenust klientidele, kelle väärtused ja hoiakud peegeldavad Concise'i omasid.

Peamiseks probleemiks on otsuse tegemiseks vajalike andmete kättesaamine ja kogumine, kuna informatsioon pole struktureeritud ühes allikas, vaid tuleb kokku koguda mitmete allikate kaudu. Allikaid on erinevaid ja kõikides allikates ei pruugi leiduda informatsiooni kõikide uuritavate ettevõtete kohta. Samuti on vajalik tekstidest olulise väljafiltreerimine ning selle hindamine.

Antud töö esialgne eesmärk on luua lahendus või prototüüp, mis vähendaks informatsiooni otsimisele kuluvat aega. Seetõttu toimub tulemi valideerimine Concise'i meeskonna eksperthinnangu kaasabil ning varasemate koostöö- ja muude kogemuste põhjal. Võimalik on võrrelda manuaalselt otsitava informatsiooni ning antud magistritöös valmiva süsteemi poolt informatsiooni leidmiseks ja analüüsimiseks kulunud aega.

## 2 Metoodika

Efektiivsemaks informatsiooni kogumiseks potentsiaalsete koostööpartnerite kohta, eelkõige otsingule kuluva aja vähendamise vaatepunktist, on mitmeid võimalusi. Otsuste tegemiseks vajalike andmete kogumiseks on võimalik kasutada erinevate veebis leitavate infoallikate poolt pakutavaid API (*Application Programming Interface*) põhiseid lahendusi, mis allikas sisalduva informatsiooni kontsentreeritult kasutajale annavad. Tihtipeale selgub aga tõsiasi, et kõik veebiallikad ei võimalda sedalaadi informatsiooni kogumist.

### 2.1 Loodava lahenduse kirjeldus

Töö käigus loodav lahendus aitab kasutajal hinnata erinevatel veebilehtedel sisalduvat informatsiooni meelsuse seisukohast.

Lahendus jaguneb kasutajapoolsest protsessivaatest lihtsustatult kaheks töövooks:

- Huvipakkuva veebilehekülje struktuurielementide salvestamine hilisemaks mugavamaks kasutamiseks ja kiiremaks sama tüüpi informatsiooni leidmiseks.
- Huvipakkuvate tekstilõikude hindamine ning hinnangute kuvamine kasutajale.

Esimene töövoog, struktuurielementide salvestamine võimaldab järgnevat:

- Navigeerides huvipakkuvale veebileheküljele, on kasutajal võimalik selekteerida tekstilõik, milles sisalduva informatsiooni „meelsus“ kasutajat võiks huvitada.
- Selekteeritud tekstilõigu kaudu on kasutajal võimalik valida atribuudid, mis teda antud lehel kõige enam huvitavad. Leheküljel leitakse üles kõik samaväärse identiteediga elemendid ning kuvatakse neid kasutajale. Kasutajal on võimalus salvestada sobiv seadistus hilisema kasutuse tarvis. Kasutaja saab salvestatud konfiguratsioonile anda nime ning veebilehekülje aadressi, kus antud konfiguratsioon töötama peab.

- Kasutajal on võimalik samale leheküljele luua mitmeid seadistusi ning neid vajadusel muuta või eemaldada.

Selle töövoos eesmärk on mugandada hilisemat samalt veebileheküljelt informatsiooni otsimise protsessi. Eeldades, et veebilehekülg vahepeal oma struktuuri ei muuda, saab sama konfiguratsiooni samade elementide leidmiseks kasutada mitmeid kordi. Antud töövoog on oluline pigem esmakordsel veebilehe külastusel.

Näiteks võib kasutaja otsustada, et LinkedIn'is on ettevõtete kohta huvipakkuvat informatsiooni. Kuna LinkedIn'i lehekülje struktuur on enamjaolt samasugune kõikidel ettevõtetel, on kasutajal võimalik ühe ettevõtte näitel seadistada LinkedIn'i struktuurile vastav konfiguratsioon, mida hiljem teiste ettevõtete puhul samuti kasutada. Seega seadistab kasutaja näiteks konfiguratsiooni, mis leiab ettevõtte uudisvoost kõik seal kuvatavad postitused. Sama konfiguratsiooniga saab kasutaja kiirelt leida kõik postitused ka teistelt LinkedIn'is olevate ettevõtete uudisvoogudelt.

Teine töövoog on seotud selekteeritud või valitud teksti hindamisega ning võimaldab teha järgnevat:

- Navigeerides huvipakkuvale veebileheküljele, on kasutajal võimalik valida eelnevalt loodud seadistus või luua uus seadistus ning seadistuse poolt leitavatel tekstilõikudel teostada meelsusanalüüsi.
- Meelsusanalüüsi tulemusi kuvatakse kasutajale visuaalselt. Lahendus kuvab ka kokkuvõtvat meelsushinnangut kõikide veebileheküljelt leitud elementide kohta. Lisaks võimaldab lahendus salvestada huvipakkuvamaid tekstilõike hilisema inimanalüüsi tarbeks.

Töövoos eesmärk on anda kasutajale tekstilõikude meelsuse kohta koondav ülevaade, millest oleks lihtne ja intuiitiline teha kasutajapoolseid järeldusi. Antud töövoog on oluline erinevate ettevõtete kohta informatsiooni kogumisel juba varem külastatud veebilehtedelt.

Näiteks saab kasutaja LinkedIn'is varasema seadistuse põhjal leida kõik ettevõtte uudisvoos kuvatavad postitused ning automaatselt kuvatakse nendega koos ka postituste sisu meelsusanalüüs. See annab kasutajale kiire ülevaate üldisest postituste meelsusest, mida saab omakorda kasutada hilisemate otsuste vastuvõtmisel.

Kokkuvõtvalt võimaldab antud lahendus kasutajal ise otsida soovitud allikatest informatsiooni, kuid on samal ajal piisavalt automaatne, et kasutaja ei pea end informatsiooniga üle külvama ning saab üleüldise hinnangu ja meelsuse ettevõtte kohta jooksvalt informatsiooni otsimise käigus. Samuti vähendab see kasutajapoolset tööd veebilehtede korduval külastamisel, kuna lehe struktuuri muutumatuse korral on eelmisest külastusest huvipakkuvad elemendid teada ning nende meelsus kuvatakse kasutajale nupuvajutuse järel kohe.

Antud lahendus on kasutajale oluline üldise ettevõtte kohta käiva meelsuse teada saamiseks võimalikult vähese ajakuluga.

## **2.2 Lahenduse loomiseks kasutatavad raamistikud ja tööriistad**

Käesolevas alapeatükis käsitletakse raamistikke ja tööriistu, mida kasutati lahenduse loomiseks.

### **2.2.1 Tekstitöötlus**

Veebilehtedelt kogutud informatsioon tuleb peale kogumist töödelda, et tuvastada kasutajale kõige olulisemad andmed. Kõige intuiitsem viis on lasta kasutajal informatsiooni töödelda seda lugedes ning sealt inimesena kõige olulisemat välja filtreerides. See on aga ajamahukas ning toetub tihtipeale ka kasutaja informatsiooni mõistmisele ning subjektiivsele hinnangule.

Kui andmed on teksti kujul, saab edasist töötlust teha kasutades tekstianalüüsi. Märkimisväärseimad tekstianalüüsi rakendused, eriti inimloodud ja struktureerimata tekstidel, on sentimentaalanalüüs ja arvamuste kaevandamine (*sentiment analysis and opinion mining*) ning nimeolemite tuvastamine (*named entity recognition*). Mõlema tulemusi saab kasutada hilisemate otsuste vastuvõtmisel.

Sentimentaalanalüüsi kasutatakse näiteks tagasisidevormide tekstidest tonaalsuse määramiseks. Loomuliku keele tõlgendamine on keeruline, kuna inimesed kirjeldavad oma tundeid, arvamusi ja hinnanguid tihti läbi kõnekujundite ja väljendite, mis pole nii üheselt mõistetavad. Sellise analüüsi lihtsustamiseks loodi raamistik [5], mis hotelli tagasiside vormidest klassifitseerib vastuseid positiivseteks, negatiivseteks ja neutraalseteks. Esialgu märgistatakse käsitsi osad tagasisided vastavalt lugeja

subjektiivsele arvamusele tonaalsuse osas, seejärel treenitakse närvivõrku nendel andmetel. Valideerimiseks kasutati OpinRank [6] andmehulka, milles olevad andmed olid eelnevalt märgistamata. Klassifitseerimise algoritmidest andis kõige parema tulemuse Naive Bayes polünoom.

Sentimentaalanalüüsi teostamiseks on loodud ka mudeleid, mis kasutavad paika pandud matemaatilisi reeglistikke. Üheks selliseks on VADER (*Valence Aware Dictionary for sEntiment Reasoning*) – kokkuhoidlik (*parsimonious*) reeglipõhine sentimentaalanalüüsi mudel sotsiaalmeedia tekstide analüüsimiseks inglise keeles. VADER tarbib vähem ressursi võrreldes teiste masinõppe mudelitega, kuna see ei vaja suurt kogust treeningandmeid [7].

VADER on parem sotsiaalmeedia tekstide klassifitseerimises positiivseteks, negatiivseteks ja neutraalseteks, kui seda on inimesed. VADER'i täpsusklass valideerimisandmete peal on 0.96, inimesel 0.84, st 96% juhtudest määratleb VADER teksti õigesti positiivseks või negatiivseks, samal ajal, kui inimene teeb seda õigesti vaid 84% juhtudel. Seda seetõttu, et VADER võtab arvesse nii sõna polaarsust (positiivne või negatiivne) kui ka intensiivsust (kui positiivne või negatiivne). Nende kahe koosmõjul arvutatakse valentsimäär, mida mõõdetakse skaalal -4 kuni 4, kusjuures -4 on kõige negatiivsem meelsus ja 4 kõige positiivsem meelsus. Skaala keskpunkt on 0, mis tähistab neutraalset hinnangut [7].

Valentsimäära arvutamisel võetakse arvesse viit erinevat heuristilist reeglit:

1. Kirjavahemärgid – näiteks hüüumärk lause lõpus tähendab üldiselt kõrgemat lause intensiivsust kui punkt. „*The weather is hot!!!*“ versus „*The weather is hot.*“.
2. Suur- ja väiketähed – suurtähtedes sõnad on üldiselt suurema rõhu all. „*The weather is HOT.*“ versus „*The weather is hot.*“.
3. Võimendavad sõnad – muudavad lause intensiivsust. „*The weather is extremely hot.*“ versus „*The weather is slightly hot.*“.
4. Sidesõnade kasutus – näiteks sõna „aga“ näitab tõenäoliselt lauseosade polaarsuse vahetust, kusjuures sõnale järgnev lause on dominantsema polaarsusega. „*The weather is hot, but it is bearable.*“ omab kahte polaarsust, kuid on määratletud teise lausepoole meelsusega.



5. Polaarsuse eitus – sõnad „ei“, „pole“ jms muudavad lause polaarsust. „*The weather isn't really that hot.*“ viitab pigem mitte väga kuumale ilmale [7].

Kokkuvõttev skoor arvutatakse lause sõnade polaarsushinnangute summana ning normaliseeritakse vahemikku -1 (kõige negatiivsem) kuni 1 (kõige positiivsem). Tegemist on kõige kasulikuma meetrikaga hinnangu väljendamiseks. Uuringutes kasutatud normaliseerimise valem on järgnev:

$$x' = \frac{x}{\sqrt{x^2 + \alpha}}$$

Valemis tähistab  $x$  lauses olevate sõnade valentsimäärade summat ning  $\alpha$  normaliseerimise konstanti, mille vaikimisi väärtus on 15.  $x'$  tähistab normaliseeritud polaarsushinnangut [7].

Sentimentaalanalüüsi saab kasutada ka erinevate sotsiaalsete sündmuste ennustamiseks. Näiteks analüüsiti Twitteri kasutajate postitusi eesmärgiga mõõta nende arvamust 2016 aasta USA presidendikandidaatide üle [8]. Selleks kasutati sõnaraamatu-läheneid, kus võrreldi saadud andmeid eeldefineeritud arvamussõnadega, et määrata andmete polaarsus. Kasutuses oli nii Natural Language Toolkit [9] kui TextBlob [10], mis on Pythoni teegid tekstandmete töötlemiseks. Selgus, et Hillary Clinton sai Twitteris rohkem positiivset vastukaja kui Donald Trump või Bernie Sanders, Donald Trump sai vastupidiselt hoopis kõige rohkem negatiivset kajastust.

Sentimentaalanalüüs pakub võimalusi arendada uusi rakendusi, tuginedes eeskätt blogide ja sotsiaalvõrgustike üha kasvavale informatsioonihulgale. Näiteks saab soovitusüsteem võtta arvesse soovitatavate toodete või teenuste eelnevaid positiivseid või negatiivseid hinnanguid. Hindamissüsteemid saavad informatsiooni koguda mitmetelt veebilehtedelt ning nende põhjal arvutada keskmise hinnangu näiteks mõne kandidaadi, toote või teenuse kohta [11].

Nimeolemite tuvastamine (NER, *Named Entity Recognition*) on kasutusel inimeste, kohtade, aegade, organisatsiooninimede jms tuvastamiseks tekstandmetest. NER süsteemi loomiseks saab kasutada raamistikke ja teke nii Javas, Pythonis kui Cythonis. Ühest läbiviidud võrdlusuuringust selgub, et samade treening- ja valideerimisandmete peal osutub Pythoni spaCy [12] kõige täpsemaks [13]. Andmestik pärines IBM'i

otsingumootorist, mis sisaldab võimalikke otsingusõnu ja -fraase, millega ennustada kasutaja päringut. Valideerimisel arvestati nii täpsust, kiirust kui protsessi kvaliteeti.

Nimeolemite tuvastamine on valdkond, mis liigub üha enam närvivõrkude kasutamise poole. Ühe uuringu kohaselt on üleminek närvivõrkudele andnud märgatavalt paremaid tulemusi kui varasemalt kasutatud *feature-engineering* (tunnusjoonte põhine õpe) ja muud juhendatud õppega algoritmid, mis keskenduvad ühele keelele või domeenile [14].

Kõiki veebilehel sisalduvaid andmeid ei pruugi vaja minna. Näiteks osad uudiste artiklid või Twitteri postitused sisaldavad sõnu ja lauseid, mis pole vajalikud ja tekitavad teksti töötlemisel n-õ müra. Informatsioonist relevantsema välja filtramiseks saab kasutada ka abistavaid eeltreenitud mudeleid nagu Word2Vec [15], mis paigutab sõnu või fraase vektorruumi sõltuvalt selle tähendusest või muust sidususest teiste sõnadega. Erinevate uudisvoogude ja Twitteri postituste sarnasuse kohta tehtud uuring tõestas, et Word2Vec mudelite kasutamine närvivõrkude treenimisel viib täpsema klassifitseerimiseni [16]. Selgus ka, et uudiste klassifitseerimine on edukam kui postituste oma, kuna sisaldab ühtlasemalt struktureeritud andmeid.

Saamaks sisendinformatsiooni õppekavade täiustamiseks ja sobitamiseks tööturul otsitavaga, uuriti veebikraapimise (*web-scraping*) kasutamise võimalikkust tööpakkumiste portaalides [17]. Leiti, et sealt saadud CSV (*comma-separated values*) kujul tekstandmeid on võimalik töödelda vastavalt võtmesõnadele, mis annaksid õppekavade koostajatele sisendi tööturul toimuva kohta ning aitaksid neil teha õppekavalisi otsuseid lähtuvalt tööturu vajadustest.

### **2.2.2 Chrome laiendus**

Tänapäevases infotehnoloogia ajastus on üha suurem fookus tavarakenduste kõrval ka veebibrauseri laienduste kasutamisel. Laiendused on üldiselt väikesemahulised tarkvararakendused, mis kohandavad veebikasutamist ning muudavad selle mugavamaks ja meeldivamaks.

Laiendus luuakse tavaliselt ühe kindla probleemi või ülesande lahendamiseks. See võib endas siduda mitmeid komponente ja suurt hulka funktsionaalsust, kuid need peaks toetama algse ülesande lahendamist [18].

Üks kõige tuntumaid ja populaarsemaid on Google Chrome veebibrauseri laienduste kogu *Chrome Extensions*, mis pakub suurt hulka erinevate kategooriate rakendusi. Võrdluses teiste tuntud brauseritega [19] on Chrome laienduste hulk suurem, mis on tõestus antud brauserilaienduste populaarsusest.

Võrreldes tavapärase rakenduse loomisega on brauserilaienduse loomine mugavam, lihtsam [20] ja tihtilugu ka kiirem. Esimese lihtsakoelise laienduse saab luua vaid minutitega [21] ning selle arendamine ja arendustulemuste valideerimine on lühikese tagasisidestamise perioodiga.

Chrome laiendusi luuakse veebiarenduse tehnoloogiatega: HTML, CSS ja JavaScript [22]. Laienduse aluspõhi on manifest.json fail, mis paneb paika kogu laienduse toimimise.

JavaScriptil on aga mitmeid raamistikke, mida saab kasutada ka brauserilaiendust luues. Populaarseimate JavaScripti raamistike, Vue.js, Angular ja ReactJS üldises võrdluses selgus, et ühte parimat raamistikku ei eksisteeri, kuid välja võib tuua ReactJS'i paindlikkust ja Vue.js'i fookust kasutada kõige uuemaid tehnoloogiaid [23].

ReactJS on deklarativne, tõhus ja paindlik JavaScripti raamistik, millega saab luua kasutajaliideseid. ReactJS võimaldab keerulisi kasutajaliidese elemente panna kokku väiksematest osadest, mida kutsutakse komponentideks [24].

ReactJS'ile tuginedes on loodud näiteks Chrome laiendus võõrkeele õppimise eesmärgil. Lisaks toodi samas töös välja laienduse lihtsuse olulisus ja sellest intuiitse arusaamise vajalikkus. Keerulise kasutajaliidese laiendusi ei kasutata. Olemasolevate Chrome laienduste nimekirjas on rohkem kui 87% selliseid, mida on alla laetud vähem kui 1000 korda. Laiendused, mida on alla laetud vähem kui 16 korda, moodustavad 50% kõikidest pakutavatest laiendustest [25].

Tempo on ettevõtte, mis keskendub Atlassiani tooteperekonnast pärit JIRA süsteemi kasutamise efektiivsemaks muutmisele. Neil oli soov luua brauserilaiendus, mille eesmärk on kasutajate töötundide logimise ja haldamise lihtsustamine ja mugandamine. Ka see laiendus loodi kasutades ReactJS raamistikku [26].

## 2.3 Protsessi kirjeldus

Käesolevas töös on probleemi osaliseks lahendamiseks kasutatud *Chrome Extension* lähenemist.

### 2.3.1 Ettevõtte kohta informatsiooni analüüsimine

Selleks, et mugandada ettevõtete kohta informatsiooni otsimise protsessi ja pakkuda kasutajale võimalust koguda informatsiooni võimalikult erinevatest allikatest, kasutatakse käesolevas töös *Chrome Extension* lahendust. See võimaldab luua lahenduse, mis ei ole tugevas sõltuvuses eelnevalt kokkulepitud veebilehtedest ning nendel lehekülgedel sisalduvast informatsioonist.

*Chrome Extension* luuakse kasutades ReactJS raamistikku, kuna see on kaasaegne, mugav ja hiljem on loodud lahendust lihtsam hallata ning laiendusele lisa-arendust teha.

Chrome laiendus lubab määrata, millistel veebilehtedel laiendust kasutada saab, mis teeb mugavamaks hilisema kasutamise, st kasutaja saab osadel veebilehtedel laienduse töö peatada.

Sentimentaalanalüüs teostatakse kasutades vabavaralist NPM paketti, mida on Node.js rakendusse mugav integreerida ja sealt paketti kasutada. Välja valitud paketiks on VADER sentimentaalanalüüs, mis on kohandatud just JavaScriptis kasutamiseks [27].

Laienduse kaudu saadava informatsiooni alusel on kasutajal võimalik võtta vastu otsus potentsiaalsele ettevõttele pakkumise koostamise osas.

## 3 Lahenduse tehniline dokumentatsioon

Käesoleva magistritöö lahenduse loomine jagunes kahte etappi: Concise Systems OÜ-ga lahenduse tingimuste määratlemine ning teksti analüüsiva ReactJS'i-põhise Chrome laienduse loomine.

### 3.1 Eeltöö

Esimene etapp oli Concise Systems OÜ vajaduste ja soovide väljaselgitamine. Ettevõttel oli mitmeid mõtteid, mida soovitud lahendus võiks teha, kuid puudus selge visioon, kuidas selle lahenduseni jõuda. Seega anti käesoleva magistritöö autorile üsna vabad käed lahenduse tehnoloogilise valiku osas.

Esialgne idee hõlmas endas ka eeldefineeritud veebilehtedelt informatsiooni kogumist kasutades veebikraapimist. Seetõttu said koostöös Concise Systems OÜ-ga paika pandud veebileheküljed, mida nende töötajaskond informatsiooni otsimisel tavaliselt kasutab. Samuti määratleti ka osad otsingukriteeriumid, mille alusel üldiselt potentsiaalse koostöö osas otsuseid tehakse.

Lõplik veebilehekülgede nimekiri, mis hõlmas nii Concise Systems OÜ soovitusi kui autori enda otsitud võimalusi, oli järgmine:

- Finding Startups (<https://findingstartups.com>)
- WebSummit (<https://websummit.com>)
- EU-startups (<https://www.eu-startups.com>)
- Startuptracker (<https://startuptracker.io>)
- Crunchbase (<https://www.crunchbase.com>)
- Scandinavian Startup Week (<https://scandinavianstartupweek.com>)
- Angel CO (<https://angel.co>)

- The Hub (<https://thehub.io>)

Concise Systems OÜ poolt eelnevalt defineeritud peamised otsingukriteeriumid olid järgmised:

- Ettevõtte peab tegutsema mõnes Skandinaavia riigis või Saksamaa-Austria-Šveits (nn DACH (*Germany (D) – Austria (A) – Switzerland (CH)*)) regioonis.
- Ettevõtte võiks olla oma toote või teenuse planeerimisel jõudnud idee tasandist kaugemale, st ollakse jõudnud reaalse teostamise faasi.
- Ettevõttel võiks olla olemasolev rahastus toote või teenuse realiseerimiseks.
- Ettevõtte võiks otsida endale toote/teenuse realiseerimiseks tööjõudu.

Autoripoolse lahenduse loomise käigus muutus ülesande skoop ja peamine fookus. Loodava lahenduse eripärade tõttu selgus, et eelnevalt kokkulepitud veebilehtedega ja kriteeriumitega lahenduse loomise jaoks pole otseselt vaja arvestada.

### **3.2 Chrome laiendus**

Teise etapi eesmärk oli luua ReactJS'i-põhine Chrome laiendus, mis suudaks tuvastada leheküljelt olulist tekstilist informatsiooni ning seda meelsuse osas analüüsida.

Autoril puudus eelnev Chrome laienduste loomise kogemus, seega kasutas autor käesoleva magistritöö teise etapi lahenduse loomiseks Joe Martella poolt loodud alusprojekti Chrome laienduse loomiseks ReactJS'i ja TypeScripti baasil [28]. Antud alusprojekt sai valitud, sest see sisaldas endas lahenduse loomiseks vajalikke komponente ning oli eelnevalt struktureeritud loogiliselt ja seadistatud mugavaks arenduseks. Selle projekti põhjal hakkas käesoleva magistritöö autor looma enda kohandatud rakendust.

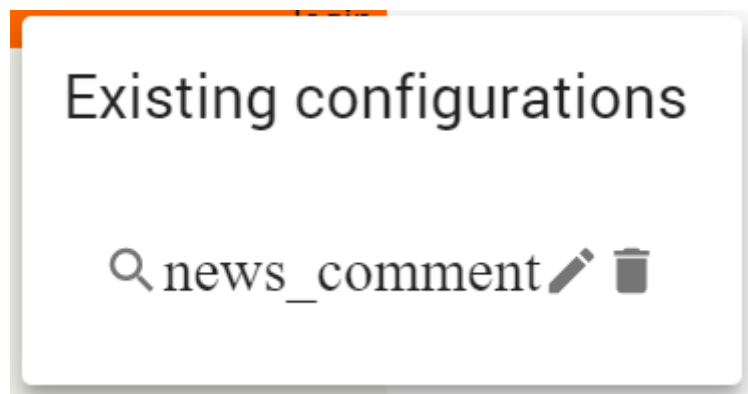
Programmikoodi kirjutamiseks kasutas autor JetBrainsi IDE (*Integrated Development Environment*) nimega IntelliJ. Programmikoodi versioonihaldus teostati kasutades GIT'i ning repositoorium loodi GitLab keskkonda.

TypeScripti mugavamaks kasutamiseks ning koodibaasis ühtsema ilme säilitamiseks kasutas autor projektis ka programmikoodi vormistamise tööriista Prettier [29].

Lisaks on lahenduse visuaalsete komponentide stiilimiseks kasutatud Material-UI raamistikku [30], mis mugandab komponentidevahelist ühtse väljanägemise hoidmist. Samuti on töös kasutatud Ramda teeki [31], mis võimaldab käidelda andmeid kasutades funktsionaalse programmeerimise võtteid, suurendades käitlemise jõudlust, lühendades programmikoodi ridade arvu ja lihtsustades programmikoodist arusaamist.

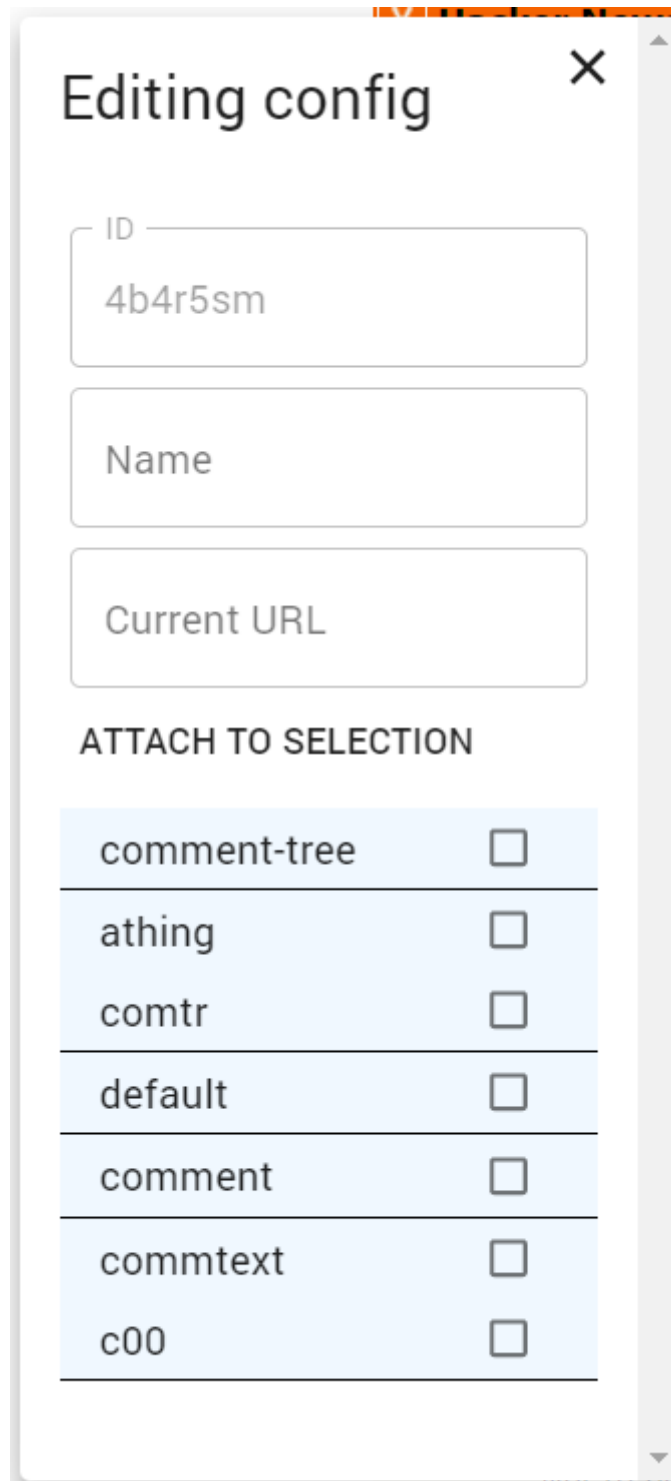
Antud kogulahenduse osa jagati planeerides kahte töövoogu: veebiallikapõhise seadistuse loomine ning seadistuse põhjal tekstide meelsusanalüüside kuvamine. Seetõttu toimus arendus vastavalt nendele töövoogudele.

Veebiallikapõhise seadistuse haldamiseks kuvatakse kasutajale kahte veebileheküljel paiknevat tööriistariba. Parempoolne element võimaldab veebileheküljele juba eelnevalt loodud seadistuste haldamist – muutmise alustamist (pliiatsiga ikoon) ning seadistuse kustutamist (prügikasti ikoon). Samuti toimub selle tööriistariba kaudu meelsusanalüüsi teostamiseks vajalik olemasoleva seadistuse järgne skaneerimine (luubi ikoon) (Joonis 1).



Joonis 1. Parempoolse tööriistariba kuvatõmmis.

Vasakpoolne tööriistariba (Joonis 2) võimaldab muuta olemasolevat seadistust või luua uut. Antud ribal kuvatakse seadistuse nimi ja veebilehekülg, kus antud seadistus peab töötama. Samuti kuvatakse seal seadistuse jaoks võimalike HTML klasside nimekiri, mille kaudu saab kasutaja endale sobiva seadistuse luua.



Joonis 2. Vasakpoolse tööriistariba kuvatõmmis.

Tehnilises teostuses on antud elemendid loodud ReactJS elementidena ning asuvad content.js failis. Chrome laienduses käivitatakse *content* failid otse veebileheküljel, st need on kasutamise hetkel veebilehekülje osad.



Seadistused talletatakse Chrome *local storage*'s. See on Chrome enda poolt pakutav informatsiooni hoiustamise võimalus. Antud lahenduses hoitakse seadistuste informatsiooni mälus JSON kujul (Joonis 3).

```
{
  "scannerConfig": {
    "url_1": {
      "id_1": {
        "classNames": [
          "className_1"
        ],
        "id": "id_1",
        "name": "config_name_1"
      },
      "id_2": {
        "classNames": [
          "className_2", "className_3"
        ],
        "id": "id_2",
        "name": "config_name_2"
      }
    },
    "url_2": {
      "id_3": {
        "classNames": [
          "className_3"
        ],
        "id": "id_3",
        "name": "config_name_3"
      }
    }
  }
}
```

Joonis 3. Seadistuste JSON kuju programmikoodi näide.

TypeScriptis kasutatakse programmikoodis seadistuste mugavamaks haldamiseks tüüpi ScannerConfig (Joonis 4).

```
export type ScannerConfig = {
  [domain: string]: {
    [id: string]: {
      id: string;
      name: string;
      classNames: string[];
    };
  };
};
```

Joonis 4. Seadistuste TypeScripti tüübi programmikoodi näide.

Andmete töötlemise ja mälu töötlamise eest vastutab antud Chrome laienduses background.js fail. Sedalaadi fail ei ole veebilehekülje osa, olles nõ tagaprogramm, kuid

on *content* failidele kättesaadav. Suhtlus *content*'i ja *background*'i vahel on Chrome enda poolt defineeritud *listener*'idena. Mõlemad pooled kuulavad sõnumeid teiselt poolelt ning saavad sama kanali kaudu vastata (Joonis 5, Joonis 6, Joonis 7) [32].

```
chrome.runtime.sendMessage({greeting: "hello"}, function(response) {
  console.log(response.farewell);
});
```

Joonis 5. Sõnumi *content*'ilt *background*'ile saatmise programmikoodi näide.

```
chrome.tabs.query({active: true, currentWindow: true},
function(tabs) {
  chrome.tabs.sendMessage(tabs[0].id, {greeting: "hello"},
function(response) {
  console.log(response.farewell);
});
});
```

Joonis 6. Sõnumi *background*'ilt *content*'ile saatmise programmikoodi näide.

```
chrome.runtime.onMessage.addListener(
function(request, sender, sendResponse) {
  console.log(sender.tab ?
    "from a content script:" + sender.tab.url :
    "from the extension");
  if (request.greeting == "hello")
    sendResponse({farewell: "goodbye"});
}
);
```

Joonis 7. Sõnumi *content*'is ja *background*'is vastuvõtmise programmikoodi näide.

### 3.2.1 Seadistuse loomine

Kasutaja poolt uue seadistuse loomine toimub järgmiselt:

1. Kasutaja navigeerib huvipakkuvale veebileheküljele. Laienduse sisselülitamisel kuvatakse kasutajale kaks tööriistariba (Joonis 8).
2. Kasutaja valib sobiva tekstiobjekti ning tõstab selle esile. Tekstiobjekt võib koosneda kuitahes mitmest osast, sh pildid, lingid jms (Joonis 9). Tekstiobjekt tuleb esile tõsta enne „*Attach to selection*“ nupuvajutust vasakpoolsel tööriistaribal (Joonis 2).

3. Kasutaja klikib vasakul tööriistaribal „*Attach to selection*“ nuppu. Seejärel leiab laiendus märgitud elemendi lehekülje HTML struktuurist ning asendab selle elemendi HTMLis uue elemendiga, lisades olemasolevale elemendile *span tag*'i, mis kuvab kasutajale elemendi ümber punast märgistust (Joonis 10). Inspiratsiooniks sai siinkohal Jérôme Parent-Lévesque'i Highlighter'i projekt [33].

Samuti otsib laiendus lehekülje HTML struktuuripuust üles kõik valitud elemendi *parent* ja *children* üksused. Leitud üksuste klassid kuvatakse grupeeritult vasakule tööriistaribale (Joonis 10).

Uuele seadistusele luuakse suvaline identifikaator, mida kasutaja muuta ei saa. Selle identifikaatori alusel toimub seadistuse hilisem muutmine ja kustutamine.

4. Kasutaja saab märkeruutude kaudu panna paika endale sobiva seadistuse. Igast grupeeringust saab hilisema seadistuse kasutamise eesmärgil valida vaid ühe klassi. Selline piirang kehtestati seetõttu, et valitud klassinimed salvestatakse andmebaasi ühemõõtmelise listina, mistõttu pole võimalik hiljem tuvastada, kas valitud klassid peavad olema ühel elemendil korraga, või mitme elemendi peale jaotatult. Seega saab hetkel valida iga elemendi ehk grupeeringu kohta vaid ühe klassi.

Kui grupeeringute klassinimedes esineb korduvaid klasse (st sama klassinime on kasutatud veebilehel mitmel elemendil), tuleks seadistust luues võimalusel märkida vähemalt üks unikaalne klassinimi, mida teistel elementidel ei esine. Samuti peaks unikaalne klassinimi olema märgitud esimesena, et järgnevaid valikuid tehes oleks võimalik korduvaid klassinimesid valides säilitada klasside hilisem õige struktuuriline järjekord (st oleks teada, millised valitud korduvad klassinimed ja nende elemendid esinesid HTML struktuuris enne ja pärast unikaalse klassinimega elementi).

Korduvaid klasse valides tähistatakse alati unikaalse klassinimega elemendile lähim valitud klassinimega element. Selline tähistamine toimub ka siis, kui algselt valitud korduva klassinimega element asus unikaalse klassinimega elemendist mitu elementi ees- või tagapool. Seega võib juhtuda, et soovitud elemendi klassinime märkimiseks tuleb valikut teha mitmeid kordi. Selline laienduse

käitumine aitab taaskord tagada, et andmebaasi salvestatakse klassinimed struktuuriliselt õiges järjekorras ning hilisemal analüüsil leitakse üles õiged, kasutaja poolt soovitud elemendid.

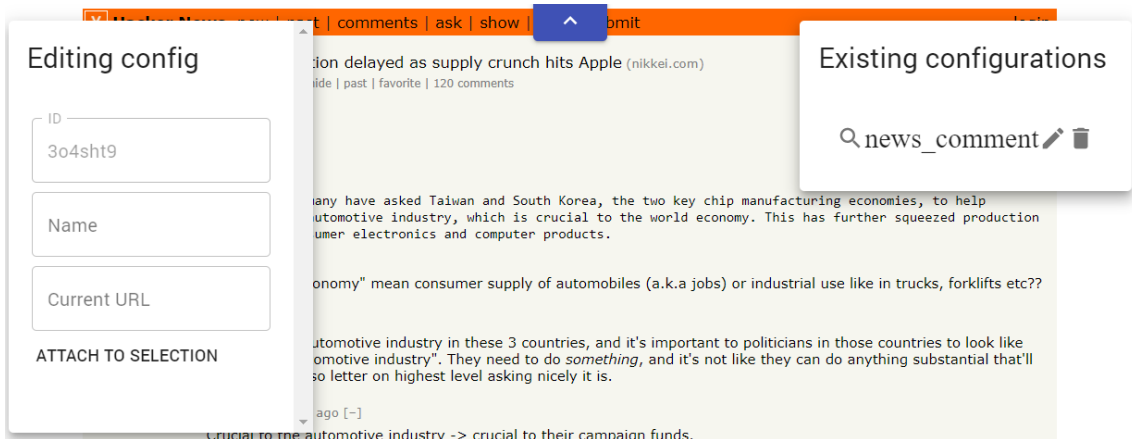
Selline erisus kehtib ainult seadistuse loomisel või muutmisel ning puudutab kasutaja vaatest vaid märkeruutude valimise ja kuvamise loogikat.

Iga märkeruudu valiku korral kuvatakse antud klassinimega elementide ümber oranž märgend. See märgend luuakse sama loogikaga mis algse elemendi punane märgend (Joonis 11).

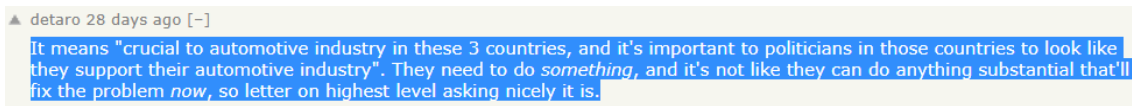
5. Kui kasutaja on jõudnud endale sobiva märgendite valikuni, on kasutajal vajalik antud seadistus nimetada ning määrata veebilehekülje URL (*Uniform Resource Locator*), kus seadistus töötab. URL on võimalik sisestada täpse viitena, mis tagab, et loodavat seadistust saab kasutada vaid täpselt antud veebileheküljel; või domeeni täpsusega, mis tagab, et loodavat seadistust saab kasutada kõigil sama domeeniga veebilehekülgedel.

Seadistuse salvestamiseks on vaja kõigi kolme komponendi olemasolu, st seadistusele peab olema määratud nimi, veebilehekülje aadress ja vähemalt üks sobiv struktuuriklass (Joonis 12).

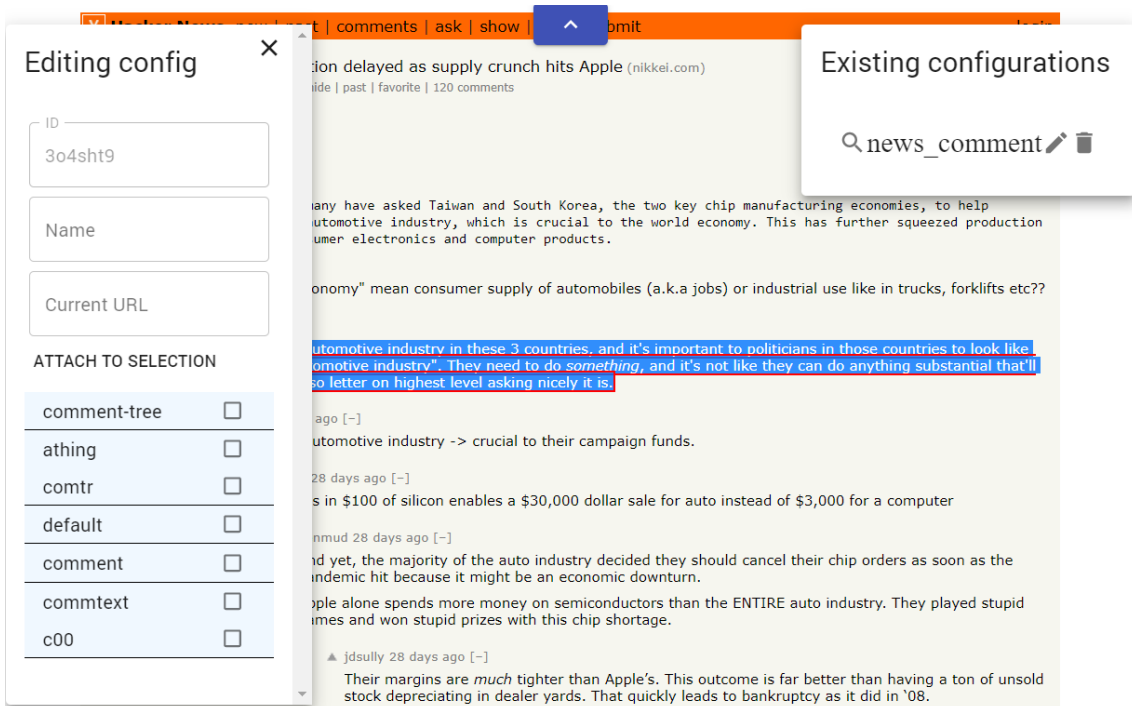
6. Tingimuste täitmisel kuvatakse kasutajale vasakpoolse tööriistariba jaluses „OK“ nuppu (Joonis 12). Sellele klikkides saadetakse *content*'i poolt *background*'i salvestamise sõnum, mis sisaldab seadistuse atribuute JSON kujul. *Background* salvestab selle Chrome *local storage*'sse.
7. Salvestatud seadistuse nimi kuvatakse kasutajale parempoolses tööriistaribas (Joonis 13).



Joonis 8. Laienduse sisselülitamisel kuvatavate tööriistaribade kuvatõmmis.



Joonis 9. Teksti selekteerimise kuvatõmmis.



Joonis 10. Selekteeritud teksti märgendamise ja HTML klassistruktuuri kuvamise kuvatõmmis.

The image shows a web application interface with two floating dialog boxes over a news article. The news article discusses Apple's supply crunch and its impact on the automotive industry.

**Editing config dialog:**

- ID: 3o4sht9
- Name: (empty)
- Current URL: (empty)
- ATTACH TO SELECTION:
  - comment-tree
  - athing
  - comtr
  - default
  - comment
  - commtext
  - c00

**Existing configurations dialog:**

- Search: news\_comment
- Icons: edit, delete

**News Article Content (highlighted):**

...ion delayed as supply crunch hits Apple (nikkel.com)  
 ...ide | past | favorite | 120 comments

...many have asked Taiwan and South Korea, the two key chip manufacturing economies, to help  
 automotive industry, which is crucial to the world economy. This has further squeezed production  
 consumer electronics and computer products.

...onomy" mean consumer supply of automobiles (a.k.a jobs) or industrial use like in trucks, forklifts etc??

...automotive industry in these 3 countries, and it's important to politicians in those countries to look like  
 automotive industry". They need to do something, and it's not like they can do anything substantial that'll  
 so letter on highest level asking nicely it is.

...ago [-]

...automotive industry -> crucial to their campaign funds.

...28 days ago [-]

...is in \$100 of silicon enables a \$30,000 dollar sale for auto instead of \$3,000 for a computer

...nmud 28 days ago [-]

...nd yet, the majority of the auto industry decided they should cancel their chip orders as soon as the  
 pandemic hit because it might be an economic downturn.

...ople alone spends more money on semiconductors than the ENTIRE auto industry. They played stupid  
 games and won stupid prizes with this chip shortage.

▲ jdsully 28 days ago [-]

...Their margins are *much* tighter than Apple's. This outcome is far better than having a ton of unsold  
 stock depreciating in dealer yards. That quickly leads to bankruptcy as it did in '08.  
 Limiting their downside risk was the right call given the information.

▲ tonyedgcombe 28 days ago [-]

...So now the auto manufacturers know they don't need to worry the next time it looks like  
 there will be a downturn. They can cancel orders and if they turn out to be wrong just get on  
 the phone to their favourite politician.

Joonis 11. Sobiva seadistuse märgendamise kuvatõmmis.

## Editing config ✕

ID  
3o4sht9

Name  
news\_comment\_2

Current URL  
https://news.ycombinat

**ATTACH TO SELECTION**

|              |                                     |
|--------------|-------------------------------------|
| comment-tree | <input type="checkbox"/>            |
| athing       | <input type="checkbox"/>            |
| comtr        | <input type="checkbox"/>            |
| default      | <input type="checkbox"/>            |
| comment      | <input checked="" type="checkbox"/> |
| commtext     | <input type="checkbox"/>            |
| c00          | <input type="checkbox"/>            |

**OK**

Joonis 12. Valiidse seadistuse kuvatõmmis.



Joonis 13. Loodud seadistuse parempoolsel tööriistaribal kuvamise kuvatõmmis.

Olemasolevat seadistust on võimalik muuta. Selle jaoks tuleb parempoolsel tööriistaribal sobiva seadistuse juures vajutada pliiatsi ikoonile, mis kuvab kasutajale seadistusest lähtuvalt eeltäidetud vasakpoolse tööriistariba. Muutmise saab lõpetada klikkides vasakpoolse tööriistariba jaluses asuvat „OK“ nuppu, mis salvestab muudatused, või klikkides vasakpoolse tööriistariba päises asuvat „X“ nuppu, mis muudatused tühistab.

Seadistuse kustutamisel, klikkides parempoolsel tööriistaribal seadistuse nime juures olevat prügikasti ikooni, saadetakse *background*'i kustutamise sõnum, misjärel *background* selle seadistuse Chrome *local storage*'st kustutab.

Tööriistaribasid saab peita ja uuesti kuvada kasutades lehekülje ülaseras asuvat *expand/hide* nuppu (Joonis 14).



Joonis 14. Laienduse komponentide kuvamise ja peitmise *expand/hide* nupu kuvatõmmis.

### 3.2.2 Meelsusanalüüsi teostamine

Sentimentaalanalüüsi teostamiseks tuleb kasutajal klikkida parempoolsel tööriistaribal sobiva seadistuse nime juures olevat skaneerimise nuppu (luubi ikoon) (Joonis 13). Seejärel leitakse *content*'is sarnaselt vasakpoolse menüü toimimisele üles kõik seadistuses valitud klassidega HTML elemendid ning nende tekstiline sisu. See saadetakse tagaprogrammi, kus saadud tekstandmetel teostatakse meelsusanalüüsi.

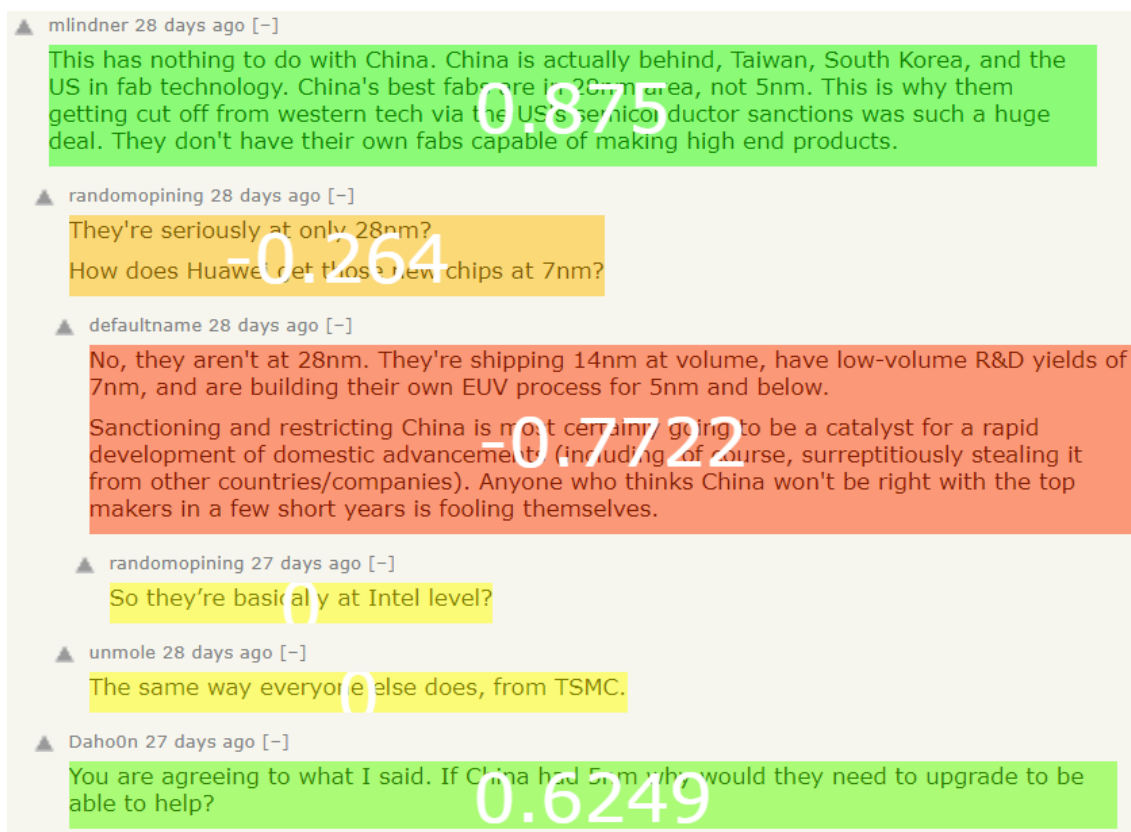


Teksti meelsuse arvutamiseks on lahenduses kasutatud NodeJS raamistikule mugandatud VADER sentimentaalanalüüsi [27]. See leiab iga sõna polaarsuse ning summeerib kõikide tekstis olevate sõnade polaarsused kokku. Seejärel tulemus ehk üldhinnang normaliseeritakse, et see jääks vahemikku -1 kuni 1.

VADER'i järgi on polaarsuse klassifikaatorid järgmised:

- Positiivne sentiment – üldhinnang suurem kui 0.05 (k.a.)
- Neutraalne sentiment – üldhinnang vahemikus -0.05 kuni 0.05
- Negatiivne sentiment – üldhinnang väiksem kui -0.05 (k.a.) [27]

Kasutajale kuvatakse teksti üldhinnangut värviliste kastidena tekstide peal (Joonis 15). Igal kastil on näha ka üldhinnangu kirje. Kastide värvid on sõltuvalt hinnangust – mida negatiivsem hinnang, seda punasem kast ning mida positiivsem, seda rohelisem. Neutraalsed hinnangud on toodud kollasega.

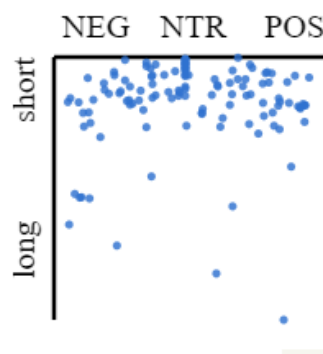


Joonis 15. Meelsusanalüüsi üldhinnangu tekstilõikudel kuvamise kuvatõmmis.

Lisaks kuvatakse kasutajale veebilehe ülaservas kogu antud klassidega HTML elementide tekstide keskmist tulemust, mis annab kasutajale indikatsiooni kogu lehekülje meelsuse osas (Joonis 16). Samuti on kasutajal võimalik näha koondtulemusi hajuvusdiagrammil (Joonis 17). Keskmise skoori arvutamine ja hajuvusdiagrammi täitmine võtab sõltuvalt leitud tekstilõikude arvust veidi aega.

Scores for 2rr7dc0: -0.002240990954934751

Joonis 16. Lehekülje keskmise meelsusskoori kuvamise kuvatõmmis.



Joonis 17. Meelsusanalüüsi koondtulemused hajuvusdiagrammil kuvatõmmis.

Hajuvusdiagrammi horisontaaltelg näitab leheküljel tuvastatud soovitud tekstilõikude meelsust negatiivsest positiivsemani. Diagrammi vertikaalteljel on kujutatud tekstilõikude pikkused tähemärkide arvu järgi lühemast pikemani. Mida rohkem on leheküljel tuvastatud tekste, seda heledamalt kuvatakse täpid graafikul, et oleks silmaga kergem tuvastada täpikogumikke ning nende tihedust.

Hajuvusdiagramm on realiseeritud käsitsi SVG (*Scalable Vector Graphics*) koordinaate kasutades. Lahendust luues olid kaalumisel ka mitmed ReactJS'ile sobilikud diagrammide raamistikud nagu Chart.js [34], react-charts [35] või react-google-charts [36], kuid antud raamistikke ei olnud võimalik piisavalt mugavalt liidestada autori poolt loodava lahendusega, mistõttu tundus olevat antud skoobis kiirem kasutada autori poolt loodud kohandatud vektorgraafika lahendust.

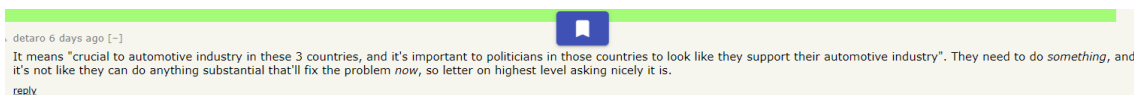
Lahenduse loomise käigus selgus, et VADER ei suuda mõningatel juhtudel tuvastada tekstides sisalduvaid sõnu ja väljendeid piisavas mahus, et anda optimaalset hinnangut koguteksti meelsusele. Seetõttu meelsushinnanguid hilisema kasutuse tarvis ei salvestata – need on kasutajale vaid üldemotsiooni indikaatoriks ning annavad informatsiooni

veebilehe üldmeelsuse kohta. Täpsema hinnangu saamiseks tuleb kasutajal siiski huvipakkuvamaid tekste endal lugeda ning sellest järeldusi teha.

Hajuvusdiagrammi saab peita ja uuesti kuvada kasutades lehe ülaservas asuvat *expand/hide* nuppu (Joonis 14). Meelsusanalüüsi tulemuste eemaldamiseks tuleb külastatav veebilehekülg uuesti laadida.

### 3.2.3 Informatsiooni salvestamine ja kuvamine

Huvipakkuvamaid tekstilõike saab hilisema inimanalüüsi tarbeks salvestada. Selle tarbeks kuvatakse kasutajale tekstilõigu meelsusskoori kuvamiseks loodud komponendi juures hiirega tekstilõigu kohale navigeerides salvestamise nuppu (Joonis 18).



Joonis 18. Teksti hilisemaks töötlemiseks salvestamise nupu kuvatõmmis.

Nupule klikkides avaneb hüpikaken, kus kasutaja saab valitud tekstilõigule lisada omalt poolt hilisemaks filtreerimiseks ja äratundmiseks lisainformatsiooni (Joonis 19). Kasutaja saab sisestada, millise ettevõtte otsingu tulemusel antud tekstilõiku soovitakse talletada ning lisada juurde võtmesõnu ja vabatekstilise kommentaari. Võtmesõnu saab valida nii varem kasutatud võtmesõnade hulgast kui luua uusi võtmesõnu. Automaatselt tuvastatakse ka veebiaadress, kus antud tekstilõik asub, et kasutaja saaks originaalse tekstilõigu hiljem kiirelt taas üles leida. Kasutajal on kohustuslik lisada ettevõtte märgend; võtmesõnad ja/või vabatekstilise kommentaari võib jätta täitmata.

A screenshot of a "You are about to save" dialog box. It contains the following fields: "Url: https://news.ycombinator.com/item?id=26743519", a "Company" dropdown menu with "Apple" selected, a "Comment:" field with the text "It means 'crucial to automotive industry in these 3 countries, and it's important to politicians in those countries to look like they support their automotive industry'. They need to do something, and it's not like they can do anything substantial that'll fix the problem now, so letter on highest level asking nicely it is. reply", a "Tags" field with a "politics" tag, and a "Your own comment" field with the text "Automotive industry problems". A blue "SAVE" button is at the bottom.

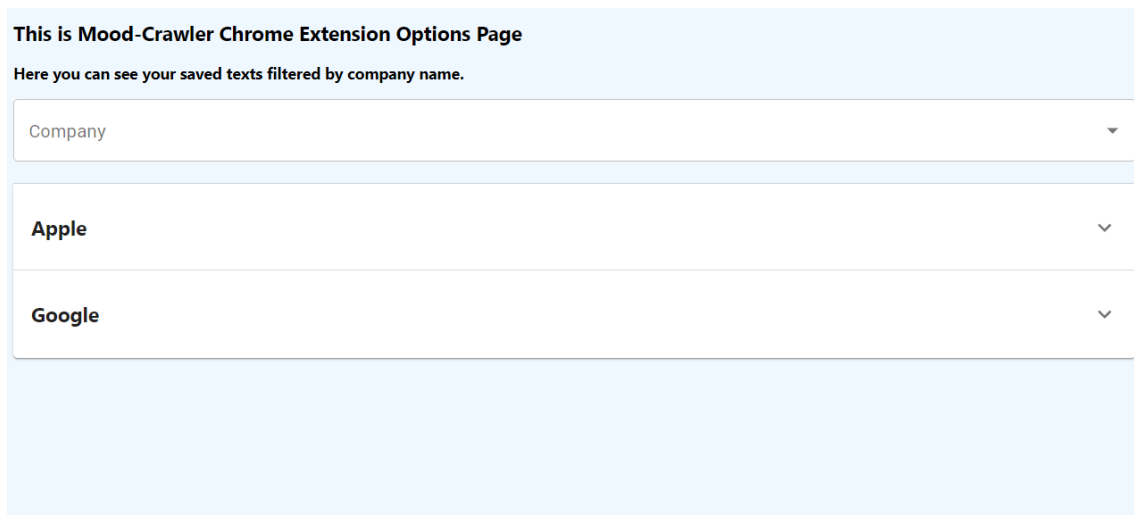
Joonis 19. Teksti salvestamise hüpikakna kuvatõmmis.

Informatsioon salvestatakse klikkides hüpikaknas „Save“ nuppu, mispeale salvestatakse informatsioon Chrome *local storage*'sse sarnaselt nagu seda tehakse seadistuste salvestamisel. Teksti salvestamisel lisatakse kasutaja andmetele ka unikaalne identifikaator UUID (*Universally Unique Identifier*) [37] kujul (Joonis 20). Identifikaatorit kasutajale ei kuvata, kuid selle kaudu toimub hilisem salvestatud tekstilõikude muutmine ning kustutamine.

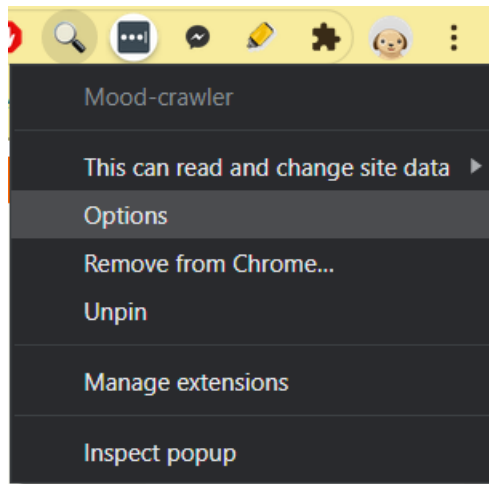
```
export type SavedTexts = {
  id: string;
  company: string;
  tags: string[];
  text: string;
  notes: string;
  source: string;
}[];
```

Joonis 20. Huvipakkuva teksti TypeScripti tüübi programmikoodi näide.

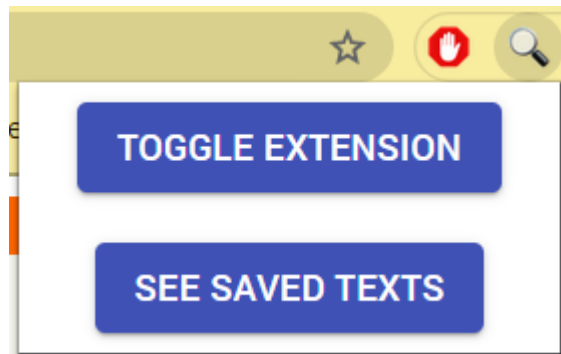
Kasutaja poolt salvestatud tekstilõike saab hiljem lugeda Chrome laienduse eraldi leheküljelt (Joonis 21), kuhu pääseb tehes paremklõpsu tegumiribal laienduse ikoonil ning valides *Options* (Joonis 22) või valides „See saved texts“ laienduse hüpikmenüüst (Joonis 23).



Joonis 21. Salvestatud tekstilõikude üldvaate lehekülje kuvatõmmis.

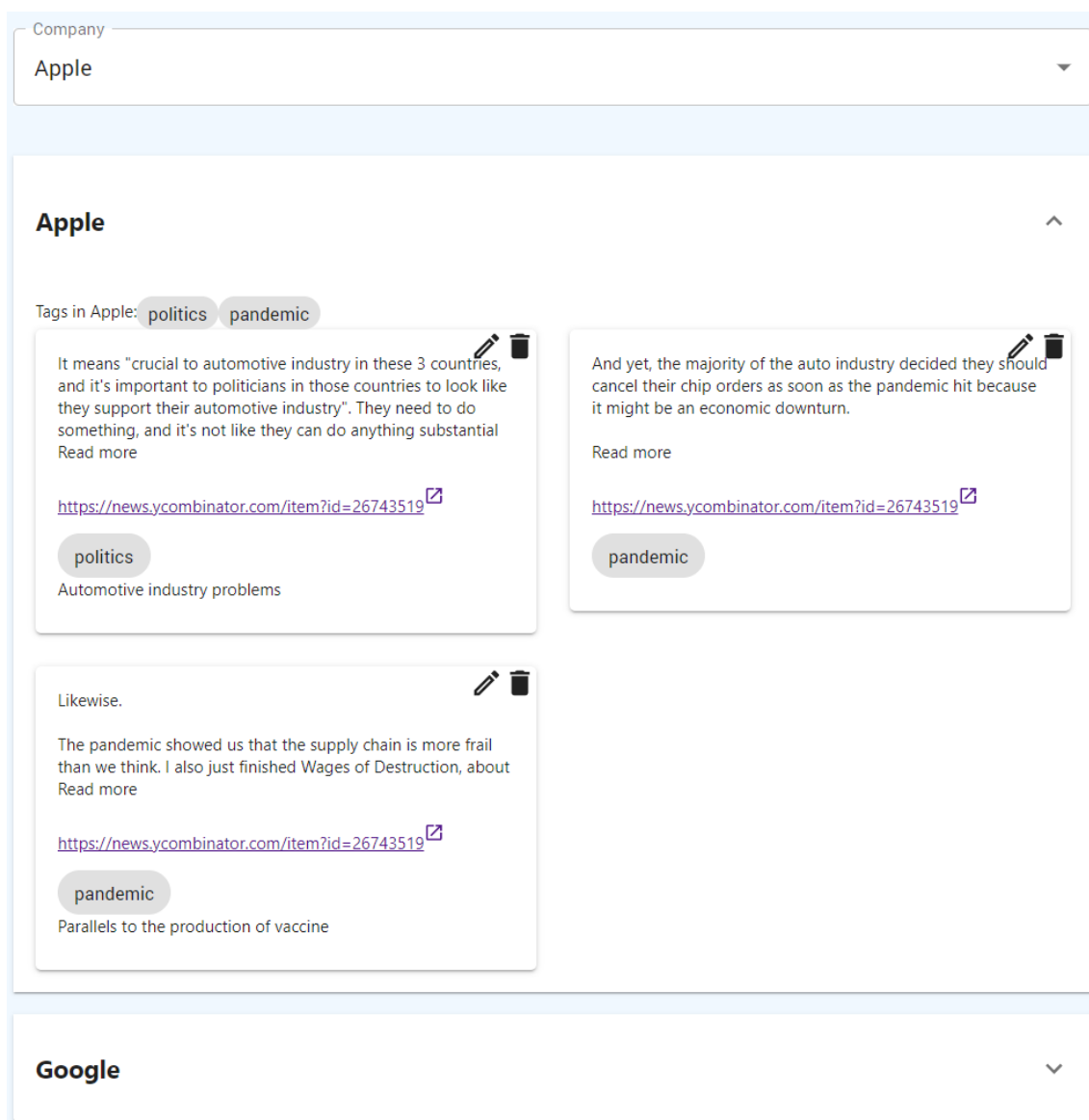


Joonis 22. Chrome laienduse eraldi leheküljele pääsemise kuvatõmmis.



Joonis 23. Chrome laienduse hüpikmenüü kuvatõmmis.

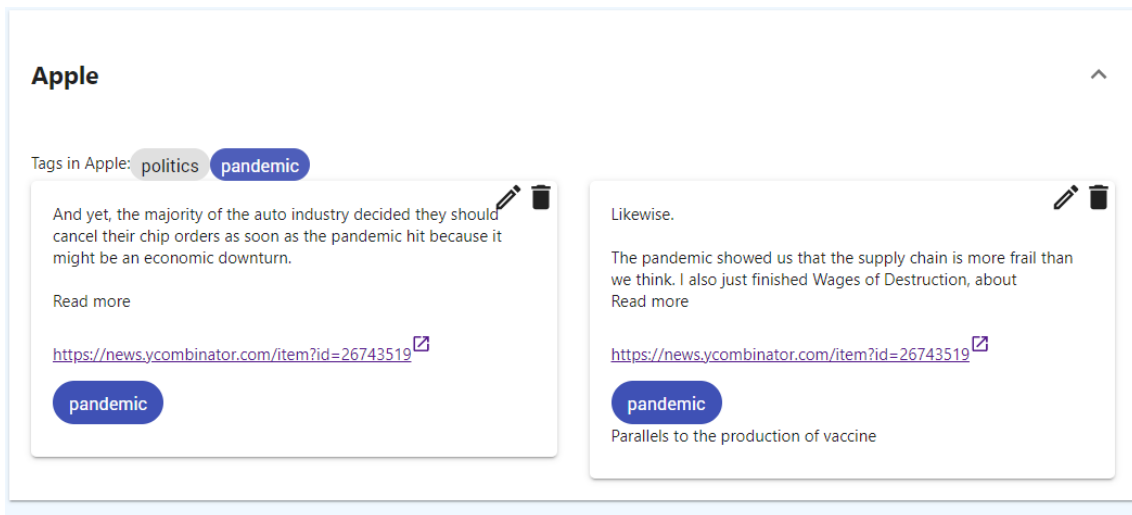
Leheküljel on kasutajal võimalus lugeda salvestatud tekstilõike. Informatsioon on leheküljele grupeeritud ettevõtte nime järgi ning kuvatakse kasutajale lehekülje lihtsuse ja ülevaatlikkuse huvides akordion-stiilis elementides. Tekstilõike näeb kasutaja avades vastava ettevõtte nimega elemendi (Joonis 24).



Joonis 24. Ettevõtte kohta käivate salvestatud tekstilõikude kuvamise kuvatõmmis.

Kasutajale kuvatakse salvestatud tekstilõigust esimesed laused, et leheküljel ruumi kokku hoida. Kasutajal on võimalus ülejäänud teksti kuvada klikkides tekstilõigu all asuvale „*Read more*“ nupule. See kuvab kogu tekstilõigu sisu. Sarnaselt on võimalus kasutajal pikemad tekstilõigud taaskord peita, klikkides selleks „*Read less*“ nupule tekstilõigu all.

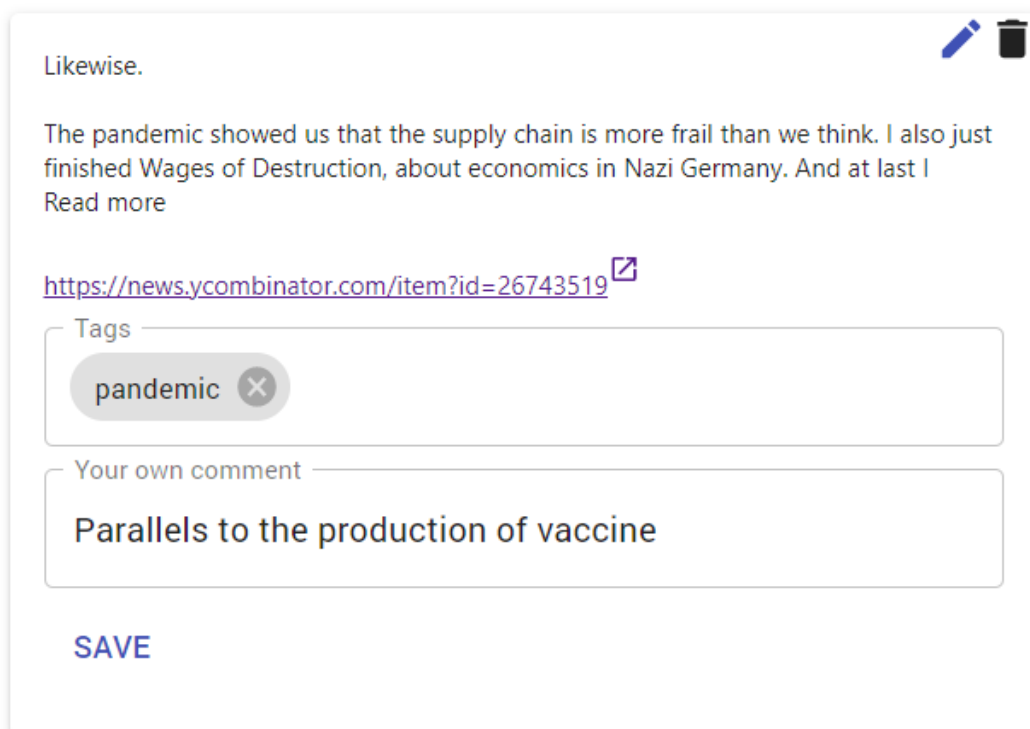
Lihtsustamaks vajaliku ettevõtte kohta käivate tekstilõikude leidmist, on võimalik andmeid ettevõtte nime järgi filtreerida. Seda saab teha lehekülje ülaserivas paikneva tekstivälja kaudu. Lisaks on võimalik ühe ettevõtte raames filtreerida tekstilõike nende võtmesõnade järgi. Seda on võimalik teha klõpsates nii avatud ettevõtte juures väljatoodud võtmesõnade loendis asuvale võtmesõnale, kui ka mõne tekstilõigu juures olevale võtmesõnale (Joonis 25).



Joonis 25. Tekstilõikude võtmesõna alusel filtreerimise kuvatõmmis.

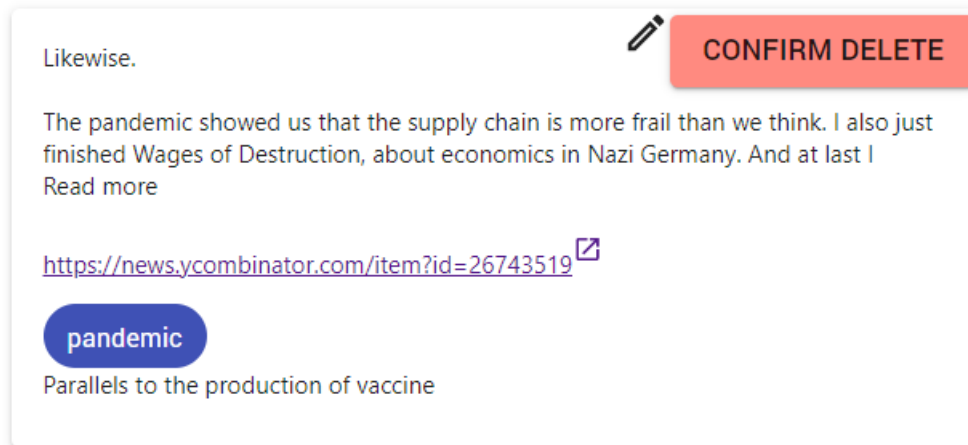
Klõpsates tekstilõigu juures asuvale veebiaadressi lingile, suunatakse kasutaja veebiaadressile, kust antud tekstilõik originaalselt salvestati.

Kasutajal on võimalik salvestatud tekstilõikude võtmesõnasid ja vabateksti muuta, klõpsates tekstilõigu juures asuvale pliiatsi ikoonile. Võtmesõnade ja vabateksti muutmise toimib sama põhimõttega nagu nende loomine tekstilõigu esmakordsel salvestamisel (Joonis 26). Muudatused salvestatakse klõpsates „Save“ nupule.



Joonis 26. Salvestatud tekstilõigu informatsiooni muutmise kuvatõmmis.

Soovi korral saab kasutaja salvestatud tekstilõigu kustutada klõpsates tekstilõigu juures asuvale prügikasti ikoonile. Kustutamisel küsitakse kasutaja kinnitust (Joonis 27). Kinnitamise nupp kaob kolme sekundi pärast automaatselt ning kui kasutaja sellele klõpsanud ei ole, tekstilõiku ei kustutata.



Joonis 27. Salvestatud tekstilõigu kustutamise kinnitamise kuvatõmmis.

Lisaks laienduse elementide peitmisele on võimalik laiendus kõikidel veebilehekülgedel peatada või taas-lubada. Seda saab teha kasutades tegumiribal laienduse ikoonile klõpsimisel avanevat hüpinkakent ning selles asuvat „*Toggle extension*“ nuppu (Joonis 23).



## 4 Analüüs ja järeldused

Analüüsi teostamiseks viidi läbi laienduse testkatsetused, simuleerides reaalseid situatsioone ja nendes esinevaid tekstandmeid. Seda tegi nii autor kui ka Concise'i poolt välja pandud meeskond, kes seni sarnaseid otsinguid manuaalselt on teinud.

### 4.1 Autori üldised tähelepanekud

Töö algne eesmärk nägi ette lahendust, mis automaatselt tuvastaks veebilehelt ettevõtete kohta käivat informatsiooni, nii faktilist (töötajate arv, asukoht, valdkond jne) kui ka meelsuslikku (kas ja kuidas suhtutakse ettevõttesse, mida nende sotsiaalmeedia postitused nende kohta räägivad, milline on kuvand meedias jne), eesmärgiga soovitada Concise'le ettevõtteid, kellega partnerlussuhtesse minna.

Selgus, et sedalaadi täislahenduse loomine on keeruline. Esimeseks takistuseks sai asjaolu, et kõik veebilehed on üles ehitatud erineva struktuuriga, mistõttu on automaatseks andmete kogumiseks tarvis esiteks tutvuda iga veebilehega eraldi ning seejärel panna paika strateegia, milliseid andmeid kust kohast automaatselt kätte võiks saada. See on ajakulukas ning loob lahenduse, mis on käsitööna loodud töötama vaid valitud hulgal veebilehtedel. Samuti oleks pidanud lähtekoodi sisse viima muudatusi igal korral, kui veebilehe struktuur mingil põhjusel muutub.

Seega keskendus autor teisele – meelsuse poolele ning püüdis luua lahendust, mis ei oleks tugevas sõltuvuses veebilehe struktuurist. Samuti pole lahendus seotud vaid valitud hulga veebilehtedega. See tähendab, et lahenduse kasutajal on suurem valikuvabadus informatsiooni allikate valikul. Samas peaks lahendus olema piisavalt informatiivne ja automaatne, et Concise Systems OÜ meeskonna töökoormust mingilgi määral vähendada.

Loodud lahendus küll ei lahenda täielikult Concise Systems OÜ probleemi, kuid on abiks väiksema osa automatiseerimisel. Samuti on loodud lahendust võimalik tulevikus täiendada, et tuvastada ka tekstilõikudes sisalduvat informatsiooni, mis on seotud ettevõtte faktiliste andmetega.

Lahenduse loomisel toetus autor statistikale [25], mille alusel keerulise ülesehitusega Chrome laiendusi laialdaselt ei kasutata, ning proovis magistr töö raames loodava lahenduse hoida küllaltki lihtsa ja intuitiivse.

Lahenduse loomisel otsustas autor just Chrome laienduse kasuks, kuna sedalaadi lahenduse implementeerimine paistis olevat mugav ja kiire. Chrome on käesoleval ajal üks kõige populaarsemaid ja kättesaadavamaid veebibrausereid [38], mis sobib enamike tänapäevaste seadmetega, seega võimaldab see autori poolset laiendust vähese vaevaga edasi jagada. Lisaks pakub Chrome laiendustele nõ andmebaasi, Chrome *storage*'t, mille kasutamine laiendustes on käepärane. Seetõttu ei pidanud autor magistr töö skoobis oleva lahenduse loomisel muretsema andmete hoiustamise pärast.

Autoril puudusid eelnevad kokkupuuted veebibrauserite laienduste loomisega, mistõttu tundus antud kontseptsiooni kasutamine autorile ka huvitav ja uudne. Esimene töötav versioon laiendusest sai valmis lühikese ajaga, kuna vajas vaid manifest.json faili ning paari lihtsamat HTML faili [21]. Iga tehtud muudatust sai koheselt testida, mis tegi tagasiside tsükli lühemaks.

Lahenduse edasiarendamise käigus selgus, et loodav lahendus võib osutuda oodatust keerulisemaks ning seetõttu otsustas autor kasutusse võtta veebiarenduse raamistikku ReactJS.

ReactJS sai valitud peamiselt seetõttu, et autoril on antud raamistikuga tööalane kokkupuude ning seetõttu ei pidanud aega kulutama uue asja õppimisele. Autori kogemusele tuginedes tundus ReactJS mõistlik valik, kuna lubab veebilehekülje elemente jagada dünaamilisteks komponentideks, mida on hiljem kergem hallata ning vajadusel taaskasutada. Samuti on ReactJS'is mugav olekute juhtimine läbi ReactJS'i sisse ehitatud *hook*'ide [39].

Soovimata alustada ReactJS rakenduse loomisega täiesti algusest, otsis autor vabavariantset malli, mida enda laienduse juures ära kasutada. Valituks osutus Joe Martella poolt loodud Chrome laienduse alusprojekt, mis kasutas ReactJS'i ja TypeScripti [28]. Selles projektis oli tehtud arendusele n-ö eelnev ettevalmistus ja seadistus, mis lühendas autoripoolset laienduse implementeerimise aega.

TypeScriptiga puudusid autoril taaskord eelnevad kogemused ning see tekitas laienduse loomise käigus kohati segadust ja põhjustas arusaamatusi. Mida aeg edasi, seda käepärasemaks antud programmeerimiskeel sai ning lõpuks oli sellest kasu mitme keerulisema funktsiooni vigade parandamisel, kuna programmeerimiskeel valideeris funktsioonide sisendite ja väljundite tüüpe. Mugav oli ka see, et IDE sai tänu TypeScripti tüüpidele autorile kasulikum olla ning sellega autoripoolsete vigade tekkimist ennetada.

Lisaks kasutas autor antud magistritöö programmikoodis ka funktsionaalse programmeerimise teeki JavaScriptile – RamdaJS. Funktsionaalse programmeerimisega oli autoril eelnev kokkupuude magistriõppeaines „Funktsionaalne programmeerimine“ ning avastades, et seda on võimalik kasutada ka JavaScripti rakendustes, soovis autor seda katsetada. Selle paradigma kasutamine nõudis eraldi õppimist ning katse-eksitus meetodi kasutamist, kuid programmikood on selle võrra lühem ja arusaadavam. Samuti osutus mugavamaks andmete käitlemine, kuna funktsionaalne programmeerimine lubab keskenduda korraga vaid ühe väikese probleemi lahendamisele [40].

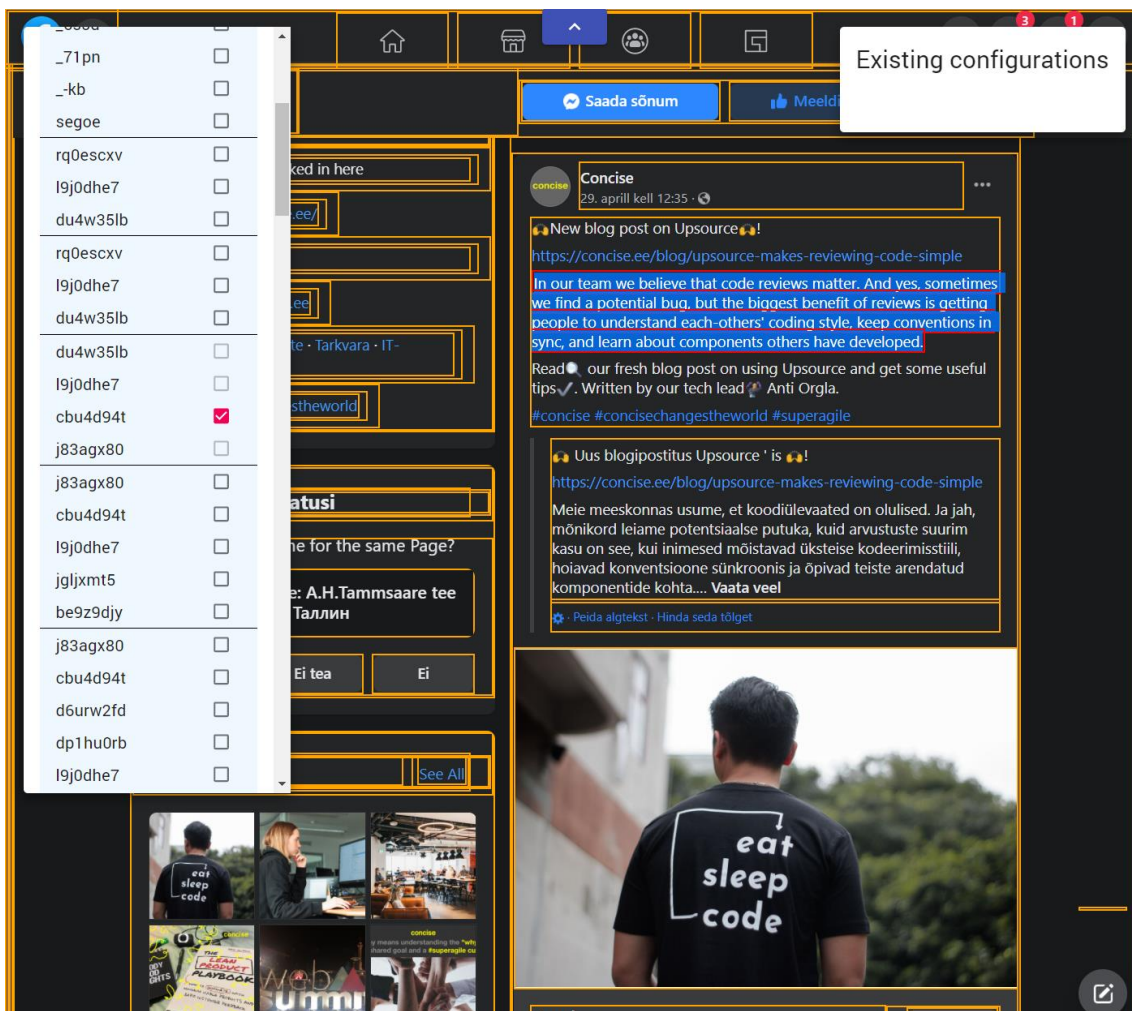
## **4.2 Autori tähelepanekud valminud lahenduse funktsionaalsuse ja käitumise osas**

Loodud lahenduse töö valideerimiseks viis autor läbi simulatsioonid erinevatel veebilehekülgedel nii lahenduse loomise ajal kui peale lahenduse valmimist. Valituteks osutusid järgmised leheküljed:

- Facebook (<https://www.facebook.com/>)
- Twitter (<https://twitter.com/>)
- BBC (<https://www.bbc.com/>)
- Hacker news (<https://news.ycombinator.com/news>)
- Reddit (<https://www.reddit.com/>)
- Gizmodo (<https://gizmodo.com/>)
- WebSummit (<https://websummit.com>)
- EU-startups (<https://www.eu-startups.com>)

- Crunchbase (<https://www.news.crunchbase.com>)
- Angel CO (<https://angel.co>)

Selgus, et osadel veebilehtedel (nt Facebook, Twitter, Crunchbase) ei tööta lahendus kõige mugavamalt. Seda seetõttu, et tänapäeval on üha rohkem veebilehti loodud sisemiselt komponendipõhiselt, st HTML DOM elementide klassinimedes on lõppkasutajale näha vaid stiiliklasside nimed, mis ei pruugi kasutajale anda informatsiooni elemendi enda olemuse kohta. See teeb klassinimepõhise selekteerimise keeruliseks. Samuti on samu stiiliklasse tihti peale kasutatud mitmetes, teineteisest sõltumatutes elementides (Joonis 28).



Joonis 28. Veebilehel facebook.com samade klassinimedega elementide hulga näite kuvatõmmis.

Samuti ei pruugi veebilehel olevad elemendid olla määratud piisava täpsusega (nt struktuuripuu võib kõige sisemine element sisaldada kogu uudise sisu koos uudise juures olevate sotsiaalmeedialinkide jms-ga, mitte ainult sisu teksti) (Joonis 29, Joonis 30).

Seetõttu võib juhtuda, et sobiva tekstilõigu omaduste piiramiseks pole võimalik sobivat klassinimedele kombinatsiooni leida.

**Editing config**

ID  
171cmil

Name  
bbc\_post

Current URL  
https://www.bbc.com/n

**ATTACH TO SELECTION**

|                                  |                                     |
|----------------------------------|-------------------------------------|
| ssrcss-1640cv5-SectionWrapper    | <input type="checkbox"/>            |
| eustbbg0                         | <input type="checkbox"/>            |
| ssrcss-1ooceyt-HeaderSection     | <input type="checkbox"/>            |
| eustbbg1                         | <input type="checkbox"/>            |
| ssrcss-u3tmht-ConsentBanner      | <input type="checkbox"/>            |
| exhqgz6                          | <input type="checkbox"/>            |
| ssrcss-6gq9s0-Wrap               | <input type="checkbox"/>            |
| e42f8510                         | <input type="checkbox"/>            |
| ssrcss-1nwia2c-ConsentBannerText | <input type="checkbox"/>            |
| exhqgz1                          | <input type="checkbox"/>            |
| ssrcss-fknekm-RichTextContainer  | <input type="checkbox"/>            |
| e5tfeyi2                         | <input checked="" type="checkbox"/> |

**Are you affected by Covid-19 in India? Please share your story by emailing haveyoursay@bbc.co.uk.**

Please include a contact number if you are willing to speak to a BBC journalist. You can also get in touch in the following ways:

- WhatsApp: +44 7756 165803
- Tweet: @BBC\_HaveYourSay
- Please read our [terms & conditions](#) and [privacy policy](#)

Are you affected by Covid-19 in India? Please share your story

0/500

**Your contact info**  
We'll be in touch if we look into your question.

Name  Email address  Contact number

Location

I am over 16 years old  
 I accept the [Terms of Service](#)

**Submit**

In some cases a selection of your comments and questions will be published, displaying your name and location as you provide it unless you state otherwise. Your contact details will never be published.

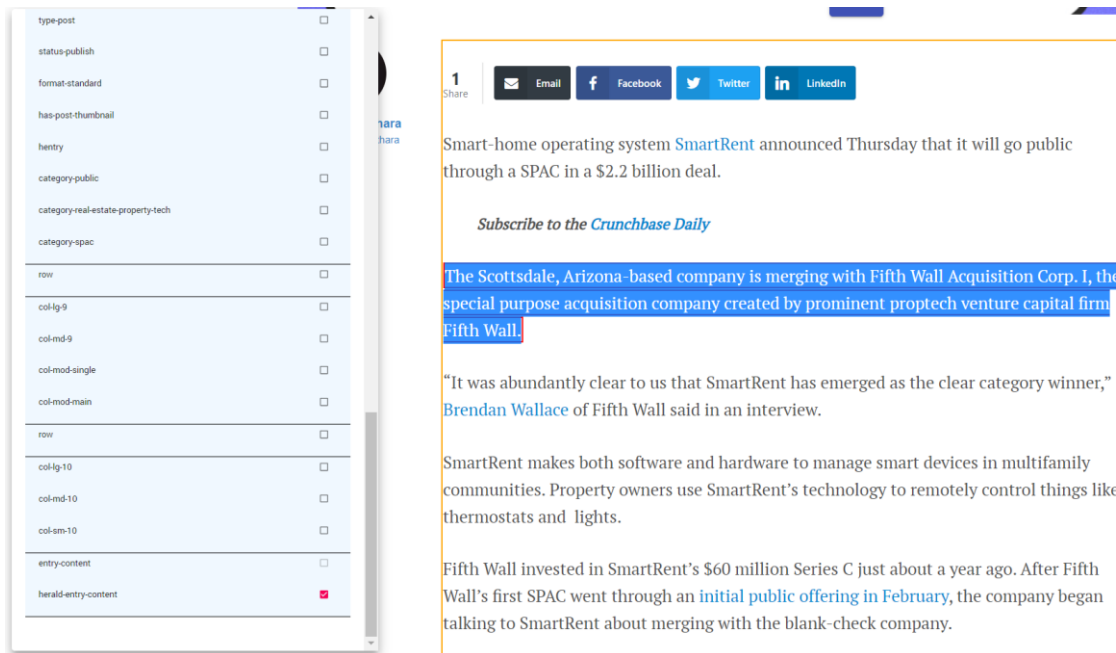
At no time should you endanger yourself or others, take any unnecessary risks or infringe any laws.

The BBC retains the right to select from these contributions based on editorial requirements and subject to [online terms and conditions](#) and [BBC editorial guidelines](#). For more information about how the BBC handles your personal data, [see here](#).

Made with Hearken | [Terms of Service](#) | [Privacy Policy](#)

If you are reading this page and can't see the form you will need to visit the mobile version of the **BBC website** to submit your question or comment or you can email us at [HaveYourSay@bbc.co.uk](mailto:HaveYourSay@bbc.co.uk). Please include your name, age and location with any submission.

Joonis 29. Veebilehel bbc.com ebapiisava täpsusega määratletud uudisartikli sisu kuvatõmmis.



Joonis 30. Veebilehel news.crunchbase.com ebapiisava täpsusega määratletud uudisartikli sisu kuvatõmmis.

Kasutaja poolt loodud seadistused töötavad veebilehel üldiselt ka siis, kui veebilehte on veidi muudetud (näiteks uusi elemente juurde lisatud). Siiski ei garanteeri see täiuslikku tööd igas olukorras ning kohati on kasutajal mugavam luua uus seadistus kui kasutada vana. Kasutaja poolt seadistuses valitud klassinimed salvestatakse ühemõõtmelisse listi ning salvestatud klassinimede järjekord on oluline, et hiljem õigeid elemente üles leida. Seetõttu on seadistuse loomisel ja muutmisel vaja järgida tehtud tehniliste järeleandmiste tõttu kindlaid reegleid (unikaalse klassinimega element tuleb valida esimesena, ühel elemendil saab valida vaid ühe klassinime jne), mis kasutaja vabadust piiravad. Eriti teravalt tuleb see välja, kui veebilehel on elemente, millel on kasutatud samu klassinimesid (nt stiiliklassid). Seadistuse loomisel ja muutmisel valitud klassinimede õige järjekorra paika saamine on seetõttu lahendatud üsnagi keeruliselt. Meelsusanalüüsi teostades seadistuse järgi õigete elementide leidmine võib korduvaid klassinimesid kasutades anda mittesoovitud tulemusi.

Lisaks, kuna veebilehed kasutavad oma visuaali loomiseks mitmeid erinevaid raamistikke ja stiilimisvõimalusi, võib autori poolt loodav lahendus erinevatel veebilehtedel välja näha erinev. Selgus, et osadel heleda taustaga veebilehtedel meelsusanalüüsi teostades muutuvad originaalsed tekstilõigud valgeks ja muutuvad kasutajale peidetuks (Joonis 31).

-0.4767

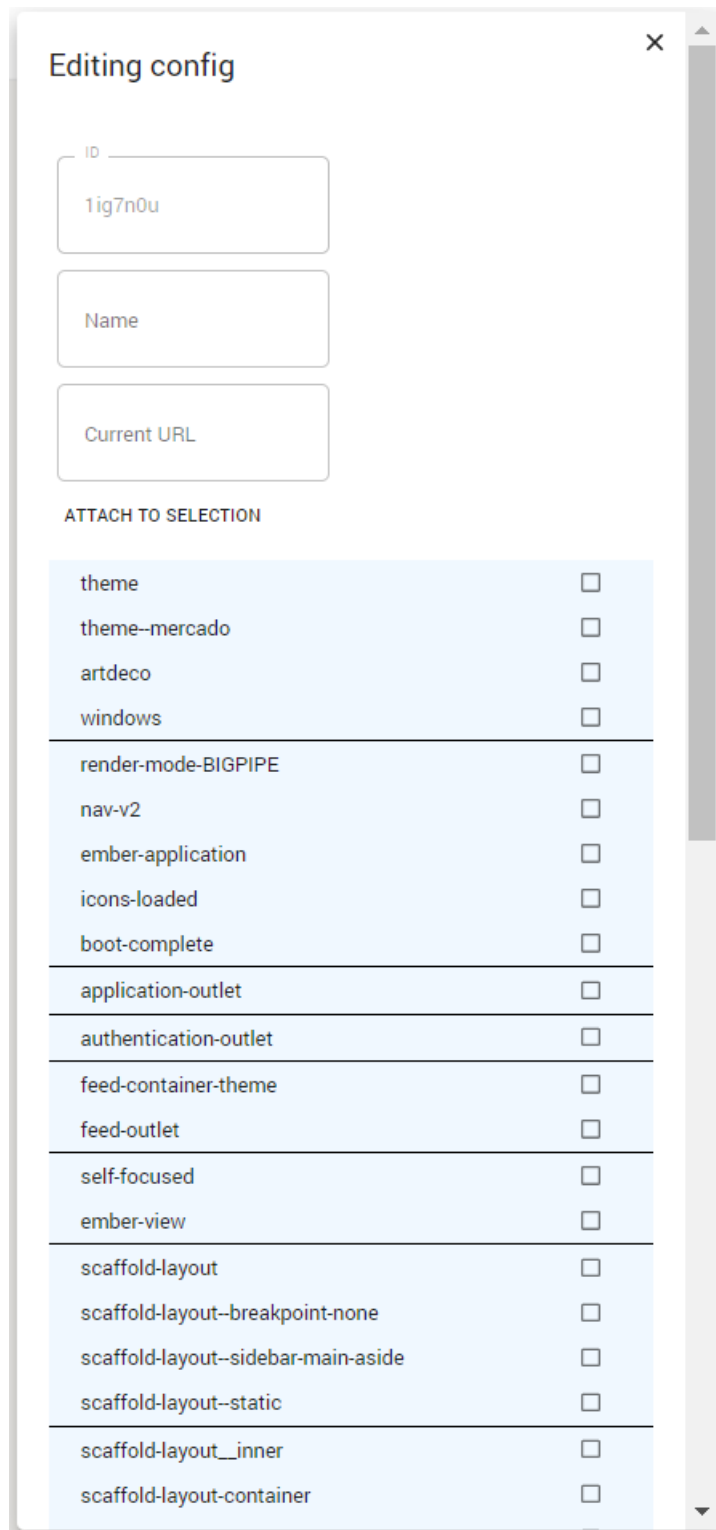
Every little thing went wrong when I tested this watch. It tracked every activity inaccurately. It said I was sleeping when I was awake. My step counts were off by more than *10,000 steps*. I changed my measurements to the imperial system, but sometimes it showed me data using the metric system anyway, just for fun. I went to test a marquee feature that the company touted during its announcement, only to find out that actually it wouldn't be available at launch.

0.4993

-0.9552

Joonis 31. Veebilehel gizmodo.com heledal taustal tekstide meelsusanalüüsi kuvatõmmis.

Funktsionaalsuse poolest lahenduses ebakõlasid ei teki, kuid ka autori lahenduse tööriistaribade ja muude elementide veidi erinev väljanägemine erinevatel veebilehekülgedel võib kasutajat häirida (Joonis 32).



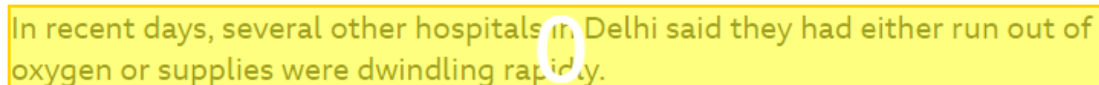
Joonis 32. Veebilehel linkedin.com vasakpoolse tööriistariba kuvatõmmis.

Samuti osutus laienduse kasutamist piiravaks asjaoluks fakt, et tekstilõikude meelsusanalüüs toimib vaid ingliskeelsetel tekstidel. See tähendab, et antud lahendust saab hetkel kasutada vaid ingliskeelsetel veebilehtedel. Kuna Concise'ile pakuvad huvi



Skandinaavia ja DACH regiooni ettevõtted, võiks meelsusanalüüs toetada ka nt rootsi või saksa keelt.

Lisaks selgus, et meelsusanalüüsi teostamiseks välja valitud VADER ei pruugi meelsuse arvutamisel olla kõige usaldusväärsem. Seda seetõttu, et VADER'i sõnavara on piiratud ning see ei tuvasta mitmeid sõnu ja fraase, millele inimene lugedes suudaks hinnangu anda. Näiteks lause „*In recent days, several other hospitals in Delhi said that they had either run out of oxygen or supplies were dwindling rapidly,*“ saab VADER'ilt neutraalse skoori, kuigi inim-lugejale jätab lause üsna negatiivse mulje (Joonis 33). Lõppkasutajale võib jääda arusaamatuks, millised sõnad meelsusskoori koondarvutuses kõige rohkem rolli mängisid ja miks.



In recent days, several other hospitals in Delhi said they had either run out of oxygen or supplies were dwindling rapidly. 0

Joonis 33. VADER'i neutraalse hinnangu kuvamine inimese jaoks negatiivsele lausele kuvatõmmis.

Meelsusarvutuste piiratud vähendamiseks saaks tulevikus antud laiendusele implementeerida mõne teise programmeerimiskeelega (nt Python) toimiva meelsusanalüüsi lahenduse, mis kasutaks suuremat sõnavara ning toetaks ka mitmeid keeli. Lisaks pole välistatud ka närvivõrkudele tugineva meelsusanalüüsi või tekstide sisulise informatsiooni analüüsiga tegelevate (nt NER) algoritmide kasutamine. Chrome laiendustes Pythoni kasutamiseks on üheks võimaluseks ilma serverita lahendus, kompileerides Pythoni programmikoodi JavaScripti programmikoodiks. Teine variant on serveripõhine, mis võimaldab Pythoni võimalusi kasutada läbi loodud teenuse [41] [42]. Antud magistr töö raames jäi selline lahendus MVP (*Minimum Viable Product*) skoobist välja.

Käesolev lahendus kasutab meelsusanalüüsi tulemustel põhinevat hajuvusdiagrammi vaid kasutajale ülevaatliku informatsiooni kuvamiseks, kuid edasiarenduste käigus saaks kuvatavat informatsiooni muuta kasutajale interaktiivsemaks. Näiteks oleks võimalus diagrammil nähtavale tekstilõiku tähistavale täpile klikkides suunata kasutaja veebilehel sama tekstilõigu juurde. See vähendaks kasutajapoolset otsimise aega – näiteks klikkides väga positiivset ja pikka tekstilõiku tähistavale täpile, saaks kasutaja suunata kohe selle tekstilõigu juurde, st kasutaja ei pea ise veebilehelt seda lõiku otsima hakkama.

Hetkeline lahendus on loodud töötama lokaalselt kasutaja arvuti veebibrauseris. See tähendab, et kasutaja loodud seadistused ja salvestatud tekstilõigud on kättesaadavad vaid ühes seadmes. Google Chrome toetab aga ka sisse logitud kasutaja põhist lähenemist, mistõttu saaks tulevikus lahendust laiendada, et kasutajal oleks võimalik oma seadistustele ja salvestatud tekstilõikudele ligi pääseda mitmest seadmest.

### 4.3 Concise Systems OÜ tähelepanekud

Autor palus Concise Systems OÜ meeskonnal anda ka omapoolne hinnang loodud lahendusele. Selle eesmärgiks oli valideerida, kas antud lahendust saaks potentsiaalsete uute koostööpartnerite leidmiseks kasutada.

Esiteks selgus, et loodud lahendus vajaks kasutusjuhendit, mille abil õppida lahendust ja selle funktsionaalsust paremini tundma. Käesoleva magistritöö raames tutvustas autor lahendust *demo* korras, kuid tulevikus vajaks lahendus kasutaja jaoks selgemaid suunised ning vihjeid, et kasutaja rakenduse tööpõhimõtetest paremini aru saaks.

Concise'i meeskond katsetas lahendust n-ö reaalsel juhitudel ja veebilehtedel, mida meeskond ise kasutab uute ettevõtete kohta hinnangu saamiseks. Katsetusi tehti järgmistel veebilehekülgedel:

- TechCrunch (<https://techcrunch.com/>)
- Reddit (<https://www.reddit.com/>)
- LinkedIn (<https://www.linkedin.com/>)
- Google Search (<https://www.google.com/search>)
- Google News (<https://news.google.com/>)
- Play Store (<https://play.google.com/store>)

Selgus, et enamikel proovitud veebilehtedel töötab lahendus korrektselt, kuid toodi välja, et lahenduse kasutamine pole kõige intuitiivsem ning võib mõnel veebileheküljel mõneti vale kasutamise korral veateadete näitamise asemel hoopis lakata täielikult töötamast. Selliste olukordadega tegeles autor juhtumipõhiselt ning üritas taasluua kasutaja olukordi, et need vead oleks hilisemas arenduses võimalik parandada.

Samuti rõhutati, et hinnangutes oli väga palju 0-väärtusega tulemeid. Samas tõdeti, et sellised 0-tulemid esinesid peamiselt lühematel tekstilõikudel. Pikematele tekstilõikudele anti pigem 0-st erinev hinnang. Samas märgati näiteks Google Play Store mõnele rakendusele hinnangu andnud inimeste kommentaaride puhul, et autori lahenduse poolt arvatud hinnang ja inimeste endi poolt pandud hinnang rakendusele ei lange kokku (Joonis 34). See viitab veelgi rohkem ka autori enda tähelepanekule, et lahenduses kasutatav VADER sentimentaalanalüüs ei pruugi anda adekvaatset hinnangut ning seetõttu ei saa seda võtta kõige usaldusväärsema tulemina.

**Hans van den Berg**  
★★★★★ 27 April 2021  
5

Okay, have been using it now for almost four months and it is easy to use, convenient, fast, with very good service even when a transfer is returned with good background info on what went wrong, cheap (it saved me in the Netherlands compared to my ING bank almost 1000 euro of fees, and ASN bank almo

0.8622

**Aamir Ognawala**  
★★★★★ 26 April 2021  
23

Transferwise was great. But since becoming Wise their mobile app has become a pain. First it logged me out on its own. And now when I try to log in, it doesn't do so and displays an error that I'm not connected to the internet. Well that's obviously not true coz I'm writing this deserving 1 star rev

0.2425

**Chamba Martinez**  
★★★★★ 3 April 2021  
125

At first, all was well. When they changed branding (Transferwise to Wise) all problems started. When I tried sending money, I initially got an internal system error. Ok, no problem. Tried again next day and this time it was asking me to clarify my identity. I wrote to CS and asked why I had to re sen

0.7184

**Andreas Kalogiros**  
★★★★★ 3 April 2021  
85

I was very happy and enthusiastic using Transferwise and my primary rating was 5 stars. However lately I had a very poor service experience due to the fact that I transferred larger amounts of money. They caused delays (I missed 4 deadlines because of that) they got very bureaucratic with very late r

0.4336

[READ ALL REVIEWS](#)

Joonis 34. Veebilehel play.google.com kasutajate hinnangu ning VADER sentimentaalhinnangu vahelise ebakõla kuvatõmmis.

Lõpetuseks toodi välja, et Concise'i meeskonnale on antud magistritöös loodud lahendusest kasu uute potentsiaalsete koostööpartnerite leidmisel vaid juhul, kui otsitava koostööpartneri nimi on teada ning otsitakse *Google Search*'i kasutades lisainformatsiooni. Antud lahendus aitaks kaasa kiire ülevaate saamisele koostööpartneri kohta kuvatavate tulemuste alusel: kas partneril on positiivne imago või mitte (Joonis 35). Samuti saaks tulemuste kaudu teada veebilehti, mida tasub lisainformatsiooni saamiseks

külastada. Samas kuvatakse *Google Search*'i tulemustes iga veebilehe kohta vaid mõned üksikud laused ning see ei pruugi alati hinnangu arvutamiseks olla piisav.



Joonis 35. Veebilehel google.com Concise Systems OÜ poolt otsitava võimaliku koostööpartneri meelsusanalüüsi kuvatõmmis.

Lisaks mainiti, et tulevikus saaks lahendusele juurde arendada nn *snippet* tööriista, mis töötaks pilditöötlusprogrammide *selection* tööriista põhimõttel. See annaks kasutajale võimaluse veebilehelt selekteerida huvipakkuvat sisulist informatsiooni (nt kontaktandmed, finantsseis jms), mida seejärel salvestada. Leiti, et sellisel juhul saaks lahendust kasutada ka juba olemasolevate CRM API'dega, mis teeks informatsiooni kogumise mugavamaks.

Tulevikuarenduste juures pakuti välja ka automaatset võimekust tekstidest sisulise informatsiooni kaevandamiseks. Sel juhul annaks kasutada käesolevas magistritöös loodud lahendust, kus kasutajale kuvatavate tekstilõikude hinnangute juures arvestatakse ka eelnevalt paika pandud kriteeriume ning kuvatakse nendele vastavust näiteks erinevate värvidega. See tähendaks, et lahendus kuvaks iga otsingutulemuse kohta paralleelselt mitu värvi, kusjuures parim tulemus on defineeritud värvide omavahelise vahekorra kaudu. Kasutaja võiks nii jõuda kiiremini „õigele“ tulemusele.

#### 4.4 Järeldused tähelepanekutest

Kokkuvõtvalt võib öelda, et autori poolt käesolevas magistritöös loodud lahendus töötab rahuldavalt enamikel veebilehtedel, kuid on siiski pigem sõltuvuses veebilehtede struktuurist ja keerukusest. Mitmetel sotsiaalmeediaplattformidel on lahenduse kasutamine keeruline ja ei anna soovitud tulemust. Sotsiaalmeediaplattformid on tänapäeval aga üheks põhilisemaks brändi kuvandi hoidmise kohaks, mistõttu oleks nendelt veebilehtedelt ettevõtete enda postituste ning platvormide kasutajate arvamuste meelsuste saamine üsna oluline.

Samuti on antud lahenduses kasutatud VADER sentimentaalanalüüs kohati ebausaldusväärne, kuna selles kasutatav sõnavara on piiratud. Seetõttu oleks vaja usaldatava hinnangu saamiseks kasutada mõnda suurema sõnavaraga või muul moel keerulisemat algoritmi. Soovituslik oleks kohandatavus ka teistele keeltele peale inglise keele. Samuti võiks kasutatav algoritm tuvastada ka tekstide sisulist informatsiooni, mis aitaks kaasa ka Concise'i soovile automaatselt koguda informatsiooni võimalike koostööpartnerite faktiliste andmete kohta.

Concise Systems OÜ probleemi ja antud magistritöö eesmärki selline lahendus käesoleval kujul seega siiski ei lahenda, kuna see ei vähenda otseselt informatsiooni kogumiseks kuluvat aega. Samas leiavad nii autor kui Concise'i meeskond, et loodud lahenduse lähenemisnurk on unikaalne ja küllaltki ainulaadne ning tulevikus edasiarenduste käigus võib sellest Concise'i probleemi lahendamisel kasu olla. Seega on antud magistritöö lahendus pigem samm soovitu suunas kui lõplik probleemi lahendus.

## 5 Kokkuvõte

Käesoleva magistritöö eesmärk oli ettevõtte Concise Systems OÜ jaoks osaliselt lihtsustada koostööeelist tegevust ning automatiseerida selle jooksul toimuvaid protsesse informatsiooni kogumise raames. Hetkel kulutab Concise inimestöötunde, et manuaalselt leida uusi võimalikke koostööpartnereid ja hinnata nende sobivust.

Magistritöö raames loodi ReactJS raamistiku ja TypeScripti baasil tekstandmetel meelsusanalüüsi teostav Google Chrome laiendus. Lahendus hõlmas endas kahte töövoogu: veebilehekülje struktuurile vastava seadistuse loomine ning loodud seadistusega tekstiliste andmete analüüsimine meelsuse aspektist. Seadistuse loomine toimub veebilehekülje HTML struktuuris esinevate klassinimedega põhjal, kusjuures laienduse kasutajal on võimalik sobiv seadistus ise kokku panna. Meelsusanalüüs kasutab varasemalt loodud seadistusse kirja pandud klassinimesid ning analüüsib veebilehel leitud klassinimedega HTML elementide tekstilist sisu kasutades selleks VADER sentimentaalanalüüsi. Kasutajal on võimalik huvipakkuvad tekstilõigud hilisemaks lugemiseks salvestada.

Loodud lahendus töötab pigem lihtsamatel veebilehtedel. Kasutatud meelsusanalüüsi pakett osutus usaldusväärse hinnangu andmiseks ebapiisavaks, mistõttu vajaks lahendus võimekamat tekstianalüsaatorit. Samuti võiks teksti analüüsiv algoritm tuvastada tekstide sisulist informatsiooni ning olla võimeline tegema seda mitmes keeles.

Käesoleva magistritöö lahendus ei täitnud soovitud eesmärki vähendada võimalike koostööpartnerite kohta leiduva informatsiooni otsimiseks kuluvat ressursikulu ega lahendanud Concise Systems OÜ probleemi. Samas on loodud lahenduse lähenemismurk unikaalne ja ainulaadne ning hilisemate edasiarenduste käigus võib sellest Concise'i probleemi lahendamisel kasu olla. Käesoleva magistritöö tulem pole lõplik probleemi lahendus, kuid on samm soovitu suunas.

## Kasutatud kirjandus

- [1] R. Bose, „Customer relationship management: key components for IT success,“ *Industrial Management & Data Systems*, kd. CII, nr 2, 2002.
- [2] C. O'Leary, S. Rao ja C. Perry, „Improving customer relationship management through database/Internet marketing: A theory-building action research project,“ *European Journal of Marketing*, 2004.
- [3] G. Mitchell, „How much data is on the internet?,“ [Võrgumaterjal]. Saadaval: <https://www.sciencefocus.com/future-technology/how-much-data-is-on-the-internet/#:~:text=One%20way%20to%20answer%20this,one%20terabyte%20is%201%2C000%20gigabytes.> [Kasutatud 21 märts 2021].
- [4] „Websummit,“ [Võrgumaterjal]. Saadaval: <https://websummit.com/>. [Kasutatud 21. märts 2021].
- [5] K. Zvarevashe ja O. O. Olugbara, „A framework for sentiment analysis with opinion mining of hotel reviews,“ *Conference on Information Communications Technology and Society (ICTAS)*, Durban, South Africa, 2018.
- [6] „OpinRank Dataset,“ [Võrgumaterjal]. Saadaval: <https://github.com/kavgan/OpinRank>. [Kasutatud 20. märts 2021].
- [7] C. J. Hutto ja E. Gilbert, VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, Association for the Advancement of Artificial Intelligence ([www.aaai.org](http://www.aaai.org)), 2014.
- [8] F. Nausheen ja S. H. Begum, „Sentiment analysis to predict election results using Python,“ *2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India, 2018.
- [9] „Natural Language Toolkit,“ [Võrgumaterjal]. Saadaval: <https://www.nltk.org/>. [Kasutatud 20. märts 2021].
- [10] „TextBlob: Simplified Text Processing,“ [Võrgumaterjal]. Saadaval: <https://textblob.readthedocs.io/en/dev/>. [Kasutatud 20. märts 2021].
- [11] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero ja E. Herrera-Viedma, „Sentiment analysis: A review and comparative analysis of web services,“ *Information Sciences*, kd. CCCXI, pp. 18-38, 2015.
- [12] „SpaCy – Natural Language Processing tool,“ [Võrgumaterjal]. Saadaval: <https://spacy.io/>. [Kasutatud 20. märts 2021].
- [13] H. Shelar, G. Kaur, N. Heda ja P. Agrawal, „Named Entity Recognition Approaches and Their Comparison for Custom NER Model,“ *Science & Technology Libraries*, kd. XXXIX, nr 3, pp. 324-337, 2020.
- [14] V. Yadav ja S. Bethard, „A Survey on Recent Advances in Named Entity Recognition from Deep Learning models,“ *The 26th International Conference on Computational Linguistics*, 2018.



- [15] „Word2Vec,“ [Võrgumaterjal]. Saadaval: <https://www.tensorflow.org/tutorials/text/word2vec>. [Kasutatud 20. märts 2021].
- [16] B. Jang, I. Kim ja J. W. Kim, „Word2vec convolutional neural networks for classification of news articles and tweets,“ *Plos One*, 2019.
- [17] J. S. Norcross, J. H. Schuessler ja D. G. Jones, „Web Scraping to Inform Curriculum Decisions,“ *Journal of Higher Education Theory and Practice*, kd. 20, nr 11, 2020.
- [18] A. Arkheden ja F. Enetorp, *A Taxonomy of Browser Extensions: Researching metadata patterns of Chrome extensions related to security using Random Forest and k-modes*, Gothenburg, 2018.
- [19] A. Sjösten, S. Van Acker ja A. Sabelfeld, „Discovering Browser Extensions via Web Accessible Resources,“ *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*, 2017.
- [20] K. Murai ja V. Klyuev, „Developing Google Chrome Extensions: Case Study,“ *Proceedings of the International Workshop on Applications in Information Technology*, 2015.
- [21] S. Kumar, „How to build your first Chrome Extension in 5 minutes,“ 14 september 2020. [Võrgumaterjal]. Saadaval: <https://medium.com/dailyjs/how-to-build-your-first-chrome-extension-in-5-min-1d3eb94575>. [Kasutatud 20. märts 2021].
- [22] „Chrome Extension Development - Getting Started,“ 18. november 2020. [Võrgumaterjal]. Saadaval: <https://developer.chrome.com/docs/extensions/mv3/getstarted/>. [Kasutatud 20. märts 2021].
- [23] E. Wohlgethan, *Supporting Web Development Decisions by Comparing Three Major JavaScript Frameworks: Angular, React and Vue.js*, 2018.
- [24] „Tutorial: Intro to React,“ [Võrgumaterjal]. Saadaval: <https://reactjs.org/tutorial/tutorial.html>. [Kasutatud 21. märts 2021].
- [25] C. Corbin, D. Cetinkaya ja H. Dogan, „Design and Development of a Web Extension to Help Facilitate the Learning of a Foreign Language,“ *International Conference on Human-Computer Interaction*, 2020.
- [26] A. M. Stefánsson, H. Róbertsdóttir, S. Þorsteinsdóttir ja Þ. J. Jónsdóttir, „Tempo Chrome Extension,“ Reykjavík, 2016.
- [27] „vaderSentiment-js,“ [Võrgumaterjal]. Saadaval: <https://www.npmjs.com/package/vader-sentiment>. [Kasutatud 03. aprill 2021].
- [28] J. Martella, „Chrome Extension (built with TypeScript + React),“ [Võrgumaterjal]. Saadaval: <https://github.com/martellaj/chrome-extension-react-typescript-boilerplate>. [Kasutatud 27. märts 2021].
- [29] „Prettier,“ [Võrgumaterjal]. Saadaval: <https://prettier.io/>. [Kasutatud 19. aprill 2021].
- [30] „Material-UI,“ [Võrgumaterjal]. Saadaval: <https://material-ui.com/>. [Kasutatud 10. aprill 2021].
- [31] „Ramda,“ [Võrgumaterjal]. Saadaval: <https://ramdajs.com/>. [Kasutatud 10. aprill 2021].

- [32] „Message passing in Chrome Extension,“ [Võrgumaterjal]. Saadaval: <https://developer.chrome.com/docs/extensions/mv2/messaging/>. [Kasutatud 22. aprill 2021].
- [33] J. Parent-Lévesque, „Highlighter,“ [Võrgumaterjal]. Saadaval: <https://github.com/jeromepl/highlighter>. [Kasutatud 27. märts 2021].
- [34] „Chart.js,“ [Võrgumaterjal]. Saadaval: <https://www.chartjs.org/>. [Kasutatud 06. aprill 2021].
- [35] „react-charts,“ [Võrgumaterjal]. Saadaval: <https://github.com/tannerlinsley/react-charts#readme>. [Kasutatud 06. aprill 2021].
- [36] „react-google-charts,“ [Võrgumaterjal]. Saadaval: <https://www.npmjs.com/package/react-google-charts>. [Kasutatud 06. aprill 2021].
- [37] „Universally unique identifier,“ [Võrgumaterjal]. Saadaval: [https://en.wikipedia.org/wiki/Universally\\_unique\\_identifier](https://en.wikipedia.org/wiki/Universally_unique_identifier). [Kasutatud 18. aprill 2021].
- [38] „Browser Market Share Worldwide,“ [Võrgumaterjal]. Saadaval: <https://gs.statcounter.com/>. [Kasutatud 20. aprill 2021].
- [39] „Introducing Hooks,“ [Võrgumaterjal]. Saadaval: <https://reactjs.org/docs/hooks-intro.html>. [Kasutatud 20. aprill 2021].
- [40] M. C. Benton ja N. M. Radziwill, Improving Testability and Reuse by Transitioning to Functional Programming, 2016.
- [41] „Create a chrome plugin with Python,“ [Võrgumaterjal]. Saadaval: <https://pythonspot.com/create-a-chrome-plugin-with-python/>. [Kasutatud 07. mai 2021].
- [42] B. Tahir, „Building A ‘Serverless’ Chrome Extension,“ 30. mai 2019. [Võrgumaterjal]. Saadaval: <https://towardsdatascience.com/building-a-serverless-chrome-extension-f684740e1ffc>. [Kasutatud 07. mai 2021].

## **Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks<sup>1</sup>**

Mina, Mirjam Pajumägi

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Koostöösobilike ettevõtete kaevandamine“, mille juhendaja on Ahti Lohk
  - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

10.05.2021

---

<sup>1</sup> Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.