

VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
INFORMATIKOS INSTITUTAS  
PROGRAMŲ SISTEMŲ KATEDRA

Praktikos ataskaita

**Santraukų sudarymo instrumento projektavimas ir kūrimas**  
(Design and Development of Tool for Summarization)

Atliko: 4 kurso 3 grupės studentas

Naglis Kontautas

(parašas)

Darbo vadovas:

prof. dr. Virginijus Marcinkevičius

(parašas)

Vilnius  
2025

## Turinys

1. Įvadas .....	3
1.1. Sąvokos .....	3
1.2. Tikslas .....	5
1.3. Praktikos užduotys .....	5
1.4. Praktiko spręsti uždaviniai .....	5
2. Praktikos vieta .....	7
2.1. Pagrindinės instituto veiklos kryptys: .....	7
2.2. Darbo sąlygos .....	7
3. Praktikos veikla .....	8
3.1. Tekstų santraukų literatūrinė analizė .....	8
3.2. Darbas su MIF HPC .....	8
3.3. Sistemos projektavimas .....	9
3.4. Programos kūrimas .....	9
4. Praktikos rezultatai .....	11
4.1. Bendroji programos informacija .....	11
4.1.1. Programos funkcionalumas .....	11
4.2. Techniniai reikalavimai .....	12
4.2.1. Funkciniai reikalavimai .....	12
4.2.2. Nefunkciniai reikalavimai .....	12
4.2.3. Sistemos ribojimai .....	12
4.3. Programos funkcionalumų aprašas .....	13
4.3.1. Panaudos atvejis „Sudaryti teksto santrauką“ .....	13
4.3.2. Panaudos atvejis „Sudaryti abstraktyvią santrauką“ .....	16
4.3.3. Panaudos atvejis „Sudaryti ekstraktyvią santrauką“ .....	18
4.3.4. Panaudos atvejis „Peržiūrėti santraukų istoriją“ .....	20
4.3.5. Panaudos atvejis „Atsisiųsti santrauką (-as) iš istorijos“ .....	22
4.4. Programos sistemos komponentų aprašas .....	24
4.5. Duomenų bazė .....	25
4.5.1. Stulpelių aprašymai ir paskirtis .....	26
4.5.2. Duomenų bazės technologija .....	26
4.6. Programos kodo komponentai .....	27
4.6.1. Šakninis lygmuo .....	27
4.6.2. controllers/ .....	27
4.6.3. models/ .....	27
4.6.4. services/ .....	27
4.6.5. static/ .....	27
4.6.6. templates/ .....	28
4.6.7. instance/ .....	28
4.6.8. summarizer/ .....	28
4.7. Naudojimasis programa .....	28
4.7.1. Santraukos tipo pasirinkimas .....	29
4.7.2. Pasirenkamas santraukos ilgis .....	30
4.7.3. Teksto santraukos generavimas .....	31
4.7.4. Istorijos peržiūra .....	32
4.7.5. Sugeneruotos santraukos atsisiuntimas .....	33
5. Išvados .....	35

Literatūra .....	36
Priedas Nr.1	

# 1. Įvadas

## 1.1. Sąvokos

**Ekstraktyvi santrauka.** Susideda iš originalaus teksto sakinių, atrinktų pagal svarbą. Šių sakinių struktūros nėra keičiamos. Kitaip tariant, ekstraktyvi santrauka yra teksto pagrindinių sakinių atrinkimas.

**Abstraktyvi santrauka.** Tekstas, kuris perteikia pagrindines originalaus teksto mintis ir idėjas, tačiau savo žodynu ir sakinių struktūra gali ženkliai skirtis nuo pradinio šaltinio. Savo pobūdžiu abstrakti santrauka yra artima žmogaus sukurtai santraukai.

**Dirbtinis intelektas.** Kompiuterinių sistemų gebėjimas atlikti užduotis, kurioms reikia žmogaus intelekto: mokytis, samprotauti, spręsti problemas, suvokti aplinką ir priimti sprendimus. Praktikos ataskaitoje ši sąvoka trumpinama kaip **DI**.

**Mašininis mokymasis.** Tai DI šaka, kurioje naudojami algoritmai, „išmokstantys“ iš duomenų. Šiame projekte ML modeliai naudojami teksto abstraktyvių santraukų generavimui.

**HPC.** Aukšto našumo kompiuterijos (angl. High Performance Computing) aplinka, klasteris (angl. Cluster) valdomas SLURM planuotojo, naudojamas didelės apimties skaičiavimams. [Vil25b]

**SLURM.** Atvirojo kodo užduočių planavimo sistema, valdoma komandų rinkiniu ir paskirstanti darbo krūvį klasterio mazgams. [YJG03]

**Vertinimo metrikos.** Kiekybiniai rodikliai, skirti įvertinti tekstų kokybę, lyginant kompiuteriu sugeneruotą tekstą su etaloniniu žmogaus kūriniu. Programa naudoja *ROUGE* [Lin04] metriką, taip pat vertinimui galima taikyti *BERTScore* [ZKW<sup>+</sup>19], *BLEU* [PRW<sup>+</sup>02] ir *METEOR* [BL05].

**Žetonizavimas.** (angl. Tokenization). Teksto apdorojimo žingsnis, kuriuo tekstas skaidomas į atskirus vienetų: žodžius, sakinius arba jų dalis.

**JupyterHub.** Serverinė sistema, leidžianti daugeliui naudotojų naudotis Jupyter užrašų knygtėmis per naršyklę, platinant skaičiavimo resursus per klasterį. [Jup24]

**Kaggle.** Internetinė platforma duomenų mokslo eksperimentams, siūlanti Python ir R užrašų knygtes, debesies skaičiavimo išteklius ir atvirus duomenų rinkinius. [Kag24]

**LabelStudio.** Atviro kodo duomenų anotacijų ir žymėjimo įrankis, palaikantis tekstų, vaizdų ir kitų duomenų tipų žymėjimą interaktyvioje sąsajoje. [Hea24]

**TextRank.** Ekstraktyvacijos algoritmas, paremta PageRank metodu, kuris sudaro teksto sakinių grafą ir atrenka svarbiausius sakinius pagal ryšius tarp jų. [MT04]

- BERTSum.** Modelių rinkinys, pagrįstas BERT architektūra, skirtas ekstraktyvios santraukos generavimui per specializuotą antraščių sluoksnį. [LL19]
- NewsRoom.** Didelis viešai prieinamas teksto–santraukos rinkinių duomenų paketas, skirtas santraukų modelių mokymui ir vertinimui.
- NLTK.** Python biblioteka natūralios kalbos apdorojimui, siūlanti žetonizavimo, POS žymėjimo ir kitus įrankius. [BKL09]
- spaCy.** Greita ir modernaus dizaino Python bibliotekėlė natūralios kalbos apdorojimui, siūlanti efektyvias žetonizavimo ir NER (vardo atpažinimo) funkcijas. [HM17]
- stanza.** Stanfordo universitetų NLP įrankių rinkinys Python kalba, teikiantis žetonizavimo, lematizavimo ir sintaksinės analizės galimybes. [QZZ<sup>+</sup>20]
- ProphetNet.** Iš anksto apmokytas kalbos modelis abstraktyvioms santraukų sudarymui, pasižymintis specialiu nurodymų previzijos mechanizmu. [YZL<sup>+</sup>21]
- SSH.** Saugus nuotolinio prisijungimo protokolas, leidžiantis vykdyti komandas nuotoliniame serveryje per užšifruotą kanalą. [YL06]
- Qluster.** Vilniaus universiteto MIF HPC klasterio vardas, kuriame veikia skaičiavimo mazgai ir SLURM planuotojas. [Vil25a]
- Figma.** Debesies pagrindu veikianti dizaino ir prototipų kūrimo platforma, leidžianti vizualizuoti vartotojo sąsajos maketus. [Fig24]
- MVC architektūra.** Programavimo modelis (Model–View–Controller), atskiriantis vartotojo sąsajos sluoksnį nuo verslo logikos ir duomenų valdymo.
- Flask.** Minimalistinis Python žiniatinklio karkasas, suteikiantis lankstesnę alternatyvą didesniems framework’ams. [Pro24]
- Django.** Pilnai aprūpintas Python web framework’as, teikiantis ORM, administravimo sąsają ir standartines programų struktūras. [Fou24]
- ORM.** Objektinio–relinio atvaizdavimo principas (Object–Relational Mapping), kuris supaprastina duomenų bazių valdymą per programos klases.
- SQLite.** Įterpiamoji (angl. embeddable) SQL duomenų bazė, sauganti visus duomenis viename faile be atskiro serverio. [Con24]
- SQLAlchemy.** Python ORM biblioteka, leidžianti dirbti su SQL duomenų bazėmis objektiniu būdu.
- CSV.** Tekstinis duomenų formatas, kuriame laukų reikšmės atskiriamos kableliais (angl. Comma-Separated Values).

## 1.2. Tikslas

Yra keli praktikos tikslai, dėl to buvo atliekami keli skirtingų sričių uždaviniai. Pirmasis tikslas buvo susipažinti su kalbos tekstų santraukų sudarymais naudojant algoritmus ir dirbtinį intelektą: kaip jie veikia, kokius algoritmus bei procesus naudoja. Antrasis tikslas - suprojektuoti ir realizuoti savo programą, kuri veiktų kaip anglų kalbos tekstų ekstraktyvių ir abstraktyvių (žr. 1.1) santraukų sudarymo įrankis.

## 1.3. Praktikos užduotys

1. Atlikti literatūros analizę, ekstraktyvaus ir abstraktyvaus santraukų sudarymui.
2. Susipažinti su santraukų sudarymo tikslumo matais.
3. Susipažinti su JupyterHub (žr. 1.1) ir Kaggle (žr. 1.1) eksperimentavimo aplinka.
4. Išanalizuoti LabelStudio (žr. 1.1) programinę įrangą ir jos projektą santraukų sudarymui.
5. Susiprojektuoti savo sistemą, kuri turėtų tokių funkcionalumą:
  - Naudotojas galėtų kurti projektus, santraukų sudarymui.
  - Naudotojas galėtų įkelti tekstą ir jį padaryti jo santrauką.
  - Naudotojas turėtų galimybę eksportuoti sugeneruotus tekstus į Excel ir Json formatą.
  - Sistema turi turėti galimybę automatiškai sudaryti abstraktyvias santraukas naudojant pasirinktą didelį kalbos modelį.
  - Sistema gebėti palyginti kalbos modelio rezultatą su žmogaus sudaryta santrauka.
6. Aprašyti šios funkcinius ir nefunkcinius.
7. Realizuoti sukurtą sistemą (naudojant Django (žr. 1.1), Flask (žr. 1.1)).

## 1.4. Praktiko spręsti uždaviniai

1. Atlikta literatūros analizė apie tekstų ekstraktyvios santraukos algoritmus (pvz. TextRank (žr. 1.1), BERTSum (žr. 1.1)).
2. Susipažinta su santraukų sudarymo kokybės vertinimo metrikomis (žr. 1.1).
3. Susipažinta su teksto žetonizavimu (žr. 1.1) ir žetonizavimo įrankiais pvz. spaCy (žr. 1.1).
4. Prisijungta prie MIF HPC (žr. 1.1) sistemos naudojant SSH protokolą (žr. 1.1).
5. Prisijungta prie MIF JupyterHub aplinkos.

6. Parašytas kodas, kuriame realizuota ekstraktyvios santraukos sudarymas naudojant skirtingus teksto žetonizavimo, santraukos sudarymo metodus bei vertinimo metrikas - spaCy žetonizavimas ir TextRank algoritmas.
7. Atlikta literatūros analizė apie tekstų abstraktyvios santraukos modelius.
8. Parašytas kodas, kuriame realizuota abstraktyvios santraukos sudarymas naudojant Prophet-Net modelį (žr. 1.1).
9. Imituotas pokalbis su programos užsakovu.
10. Susipažinta su programos sistemos projektavimu.
11. Aprašyta užsakovo pateiktą sistema ir sudaryti funkciniai bei ne funkciniai sistemos reikalavimai.
12. Aprašytos sistemos funkcijos ir sukurtos architektūros diagramos.
13. Sukurtas sistemos dizainas bei jos veikimo simuliacija naudojant Figma (žr. 1.1).
14. Realizuota sistema naudojant Python kalbą, Flask programinį karkasą.
15. Realizuotas prisijungimas prie MIF HPC sistemos ir abstraktyvios santraukos sudarymas naudojant dirbtinio intelekto modelį.

## 2. Praktikos vieta

VU Duomenų mokslo ir skaitmeninių technologijų institutas yra Vilniaus universiteto Matematikos ir informatikos fakulteto padalinys.

Institutas įsteigtas Lietuvos ūkiui ir tarptautiniam bendradarbiavimui svarbiems ilgalaikiams tarptautinio lygio moksliniams tyrimams vykdyti. Pagrindinė Instituto veikla yra moksliniai tyrimai ir eksperimentinė plėtra (MTEP).

Be mokslinių tyrimų ir eksperimentinės plėtros, Instituto uždaviniai yra kartu su aukštosiomis mokyklomis rengti mokslininkus, padėti joms rengti specialistus bei teikti informatikos ir matematikos mokslinę bei metodinę paramą švietimo įstaigoms.

**Misija** – vykdyti ir vystyti informatikos, informatikos inžinerijos ir matematikos mokslinius tyrimus, eksperimentinę plėtrą bei doktorantūros studijas.

**Vizija** – tapti efektyviai veikiančiu ir pasaulyje žinomu mokslo centru, teikiančiu pažangias mokslo paslaugas Lietuvai.

### 2.1. Pagrindinės instituto veiklos kryptys:

- Tikimybių teorijos ir matematinės statistikos, finansų ir draudos matematikos, diferencialinių lygčių ir jų skaitinių sprendimo metodų bei matematinės logikos, algoritmų sudėtingumo ir diskrečiosios matematikos problemų tyrimai;
- Atpažinimo procesų, duomenų analizės, daugiaekstremalių optimizavimo uždavinių bei multimedijos technologijų ir interaktyviųjų sistemų tyrimai;
- Informatikos teorinių ir metodinių pagrindų, mokyklinės informatikos problemų, kompiuterizuotų sistemų ir kompiuterių tinklų inžinerijos metodų tyrimai bei taikymai naujoms informacinėms technologijoms kurti; kultūros paveldas skaitmeninėje erdvėje.
- Leidžiami tarptautiniai (ISI Web of Science) žurnalai „Informatica“, „Lithuanian Mathematical Journal“, „Nonlinear Analysis. Modelling and Control“ ir kt.

### 2.2. Darbo sąlygos

- Robotikos dirbtuvių kabinetas, kuriame dirba praktikantai.
- Nuosavas nešiojamas kompiuteris.
- Nemokamas universiteto internetas.
- Yra virtuvė, tualetas ir valgykla.



### 3. Praktikos veikla

#### 3.1. Tekstų santraukų literatūrinė analizė

Vyko teorinis susipažinimas su tekstų santraukų generavimu: kokių tipų yra santraukos, kokie yra santraukos generavimo žingsniai ir kokie yra santraukos generavimo įrankiai.

Santraukų tipai yra du: abstraktyvus ir ekstraktyvus. Kadangi ekstraktyvaus tipo generavimas yra paprastesnis suprasti bei įgyvendinti nei abstraktyvaus, todėl pradėjau nuo ekstraktyvios santraukos analizės. Išsiaiškinau, jog ekstraktyvios santraukos sudarymo žingsniai yra tokie: duomenų pateikimas, duomenų žetonizavimas, santraukos sudarymas ir sudarytos santraukos įvertinimas.

Susipažinau, išanalizavau ir įgyvendinau algoritmus, skirtus sudaryti ekstraktyvias santraukas, tokiais kaip TextRank ir BERTSum. Duomenų paketui pasirinkau naudoti NewsRoom duomenis (žr. 1.1). Išsiaiškinau kas yra žetonizavimas ir kokių yra žetonizavimo įrankių. Susipažinau, išanalizavau ir įgyvendinau skirtingus žetonizavimo įrankius: NLTK (žr. 1.1), spaCy bei stanza (žr. 1.1). Taip pat, Susipažinau, išanalizavau ir įgyvendinau su vertinimo metrikomis: ROUGE, BERTScore, BLEU, METEOR (žr. 1.1).

Tada, susipažinau, išanalizavau ir įgyvendinau algoritmą ProphetNET, skirtą sudaryti abstraktyvias santraukas. ProphetNET yra kalbos modelis, treniruotas abstraktyvių santraukų sudarymui, jis atlieka sudėtingus skaičiavimus ir teksto analizavimus. Modeliai, skirti abstraktyvioms santraukoms sudaryti, tarpusavyje veikia skirtingai, todėl jie turi specialiai savo modeliui pritaikytą žetonizavimą, tad, kitaip sakant, mano pasirinktas žetonizavimo įrankis yra ProphetNET. Duomenų paketą naudoju tą patį.

#### 3.2. Darbas su MIF HPC

Kadangi darbas vyksta su modeliais, kuriems reikia greitų kompiuterinių resursų, tam, kad procesas vyktų greitai. Dėl to yra poreikis prisijungti prie MIF HPC sistemos ir pasinaudoti aukšto pajėgumo skaičiavimo resursais. Išsianalizavau MIF paruoštą instrukciją, kaip reikia prisijungti prie MIF HPC sistemos naudojant SSH protokolą. Prisijungiama yra prie Qluster sistemos (žr. 1.1), kurioje su SLURM planuotoju (žr. 1.1) galima prisijungti prie MIF HPC skaičiavimo resursų ir, naudojant SLURM, sudaryti užduotis. Dėl to, teko susipažinti su Qluster ir SLURM bei išsiaiškinti, kaip visa šita grupė sistemų veikia.

Susipažinau su JupyterHub sistema, kuri paleidžiama prisijungus prie MIF HPC sistemos ir resursų. Ji reikalinga pačiam santraukų modelių ir algoritmų kodo rašymui. Joje buvo padaryti užduotys su tekstų santraukų generavimu bei teorija, minėtais anksčiau.

Atskiruose failuose buvo parašyti kodai:

- **Ekstraktyvi santrauka.** Realizuota ekstraktyvios santraukos sudarymas, kuriame galima dinamiškai naudoti skirtingus teksto žetonizavimo įrankius, santraukos sudarymo metodus bei vertinimo metrikas. Išanalizuoti ir realizuoti žetonizavimo įrankiai: NLTK, spaCy bei stanza.

Išanalizuoti ir realizuoti ekstraktyvios santraukos algoritmai: TextRank ir BERTSum. Išanalizuotos ir realizuotos sugeneruotos santraukos vertinimo metrikos: ROUGE, BERTScore.

- **Abstraktyvi santrauka.** Realizuota abstraktyvios santraukos sudarymas naudojant ProphetNet modelį bei žetonizatorių.

### 3.3. Sistemos projektavimas

Susipažinus su dirbtinio intelekto skirtu santraukų generavimui teorija, buvo imituotas pokalbis su santraukų generavimo programos užsakovu. Užsakovui išsakius lūkesčius dėl programos ir jos funkcionalumų, sudariau programos aprašymą bei funkcinis ir nefunkcinis reikalavimus.

Susipažinau apie programų projektavimą: kokie yra projektavimo žingsniai, kokių diagramų reikia ir ką tos diagramos atskleidžia. Remdamasis reikalavimais, kūriau programos projektą ir sudariau panaudos atvejų diagramą (angl. Use Case diagram, žr. pav. 1). Kiekvienam panaudos atvejui sudariau ir aprašiau scenarijus, kaip ir kokie žingsniai yra atliekami, kokios yra atvejo pasekmės ir reikalavimai, taip pat, aprašiau alternatyvius scenarijus. Kadangi aiškiai aprašyti scenarijus kartais yra sudėtinga, nubraižiau veiklos diagramas, kurios iliustruoja, veikimo scenarijus.

Sudariau aukšto lygio architektūrinę diagramą, kurioje pavaizduoti pagrindiniai programos sistemos komponentai.

Pagal diagramas, sukūriau programos simuliaciją naudodamasis Figma įrankiu. Simuliacijoje yra parodomas programos naudotojo sąsajos dizainas bei pagrindiniai funkcionalumai, tokie kaip santraukos užklausos sudarymo ar santraukų istorijos funkcionalumai.

Visos šios diagramos yra reikalingos pagerinti komunikaciją tarp programuotojų komandos ir užsakovo - programuotojai žino, ko tiksliai nori užsakovas, o užsakovas gali pakeisti funkcionalumus, kol jie dar nėra pradėti programuojami. Todėl, programos architektūros projektavimas užtikrina programos sistemos aiškumą, sumažina klaidų rizikas ir padeda išvengti pridėtinių kainos ir laiko išlaidų.

### 3.4. Programos kūrimas

Remdamasis architektūros diagramomis, funkciniais bei nefunkciniais reikalavimais ir sudarytu sistemos dizainu, sukūriau ir pačią programą. Programa yra internetinės svetainės tipo, rašyta Python kalba ir naudojant Flask programinę rėmą bei MVC programavimo architektūra (žr. 1.1). Vartotojo sąsaja parašyta HTML bei CSS kalbomis, o vaizdinis sąsajos dizainas darytas pagal jau sudarytą simuliacijos dizainą.

Kadangi ekstraktyviai santraukai generuoti nėra būtina turėti greitų skaičiavimo resursų, todėl šis funkcionalumas parašytas kaip programos kodas. Tuo tarpu, abstraktyvios santraukos generavimo modeliai reikalauja aukšto lygio skaičiavimo resursų, tad, kad juos gautume, yra prisijungiama prie MIF HPC sistemos SSH protokolu su mano asmenine MIF paskyra. Prisijungus paleidžiama

Python programa, kurioje realizuotas abstraktyvios santraukos generavimas naudojant ProphetNET modelį.

## 4. Praktikos rezultatai

Atliktoje praktikoje gauti rezultatai yra programos architektūrinis aprašymas bei pati programa.

### 4.1. Bendroji programos informacija

Teksto santraukų generatorius – žiniatinklio tipo programa (angl. Web), kuri suteikia naudotojams efektyvią ir automatizuotą priemonę glaustoms santraukoms generuoti anglų kalba. Programa naudoja mašininio mokymosi (ML) modelius, kurie analizuoja ir sutrumpina įvestą tekstą, išlaikydama jo esmę bei svarbiausius aspektus.

Pagrindinis sistemos tikslas – palengvinti informacijos santraukų kūrimo procesą tyrėjams, studentams, turinio kūrėjams ir profesionalams, kuriems reikia greitai gauti esminę informaciją iš didelių tekstinių duomenų anglų kalba kiekių. Naudotojai gali įvesti tekstą, pateikti užklausą ir per kelias sekundes gauti santrauką. Programa užtikrina tikslumą, minties esmę ir nuoseklumą sugeneruotuose tekstuose.

Ši programa gali būti naudinga studentams, akademikams, žurnalistams ir specialistams, kurie dirba su dideliais tekstų kiekiais ir ieško greito, automatizuoto bei kokybiško informacijos apibendrinimo be rankinio darbo.

Teksto santraukų generatorius siekia padidinti produktyvumą, pagerinti prieigą prie informacijos ir suteikti patikimą, dirbtiniu intelektu pagrįstą sprendimą teksto santraukų generavimui.

#### 4.1.1. Programos funkcionalumas

Programa leis naudotojams atlikti šiuos veiksmus:

- Pasirinkti santraukos tipą (ekstraktyvi ar abstraktyvi santrauka, žr. 1.1).
- Įvesti arba nukopijuoti tekstą.
- Nustatyti pageidaujamą santraukos ilgį (nuo 10 % iki 50 % teksto).
- Sugeneruoti santrauką paspaudus mygtuką.
- Matyti pradinio ir santraukos tekstų ilgius.
- Išsaugoti sugeneruotas santraukas (pvz., CSV formatu, žr. 1.1).
- Peržiūrėti santraukų generavimo istoriją.
- Gauti klaidų pranešimus, jei tekstas per ilgas ar yra kitų nesklandumų.

## **4.2. Techniniai reikalavimai**

### **4.2.1. Funkciniai reikalavimai**

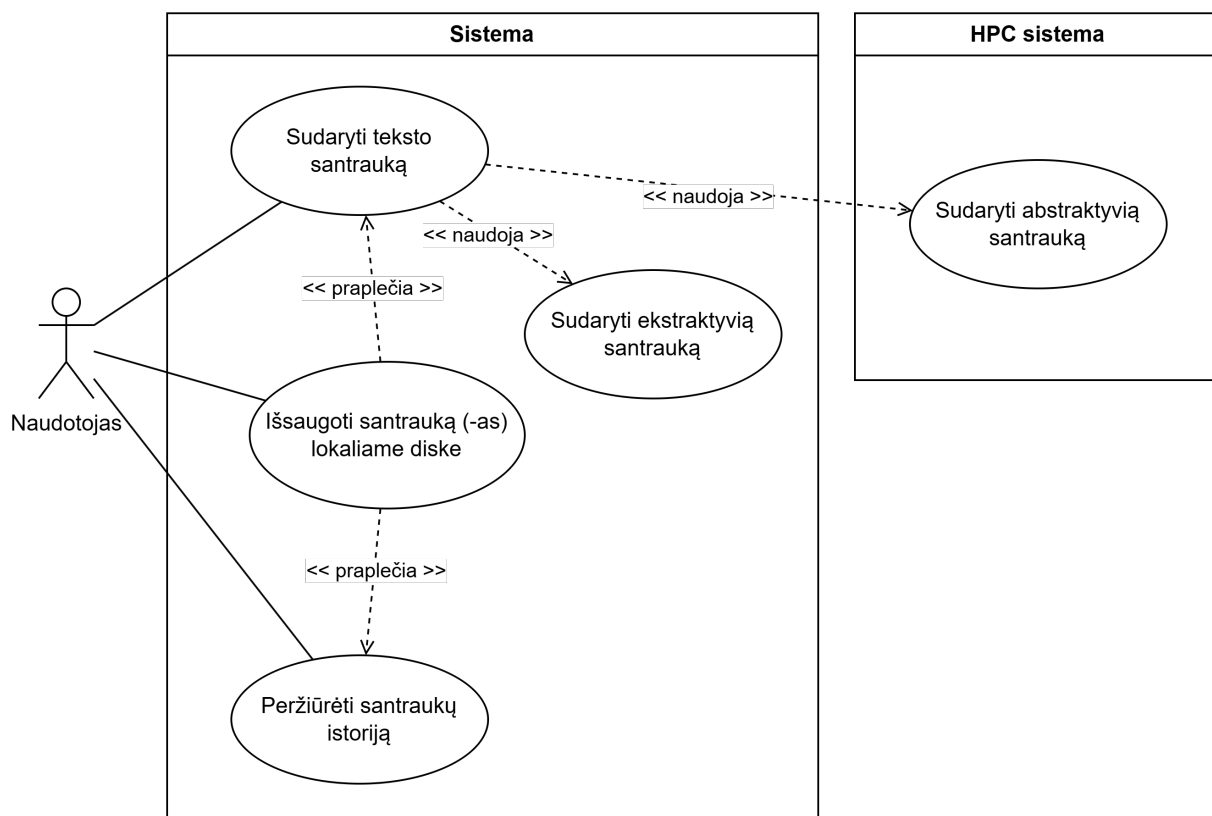
- Naudotojas gali pasirinkti santraukos tipą.
- Naudotojas gali įvesti arba įkelti tekstą.
- Naudotojas gali pasirinkti santraukos ilgį (nuo 10% iki 50%).
- Sistema turi sugeneruoti santrauką pagal naudotojo nurodytus parametrus.
- Sistema turi rodyti originalaus teksto ir sugeneruotos santraukos ilgį.
- Naudotojas gali atsisiųsti sugeneruotą santrauką CSV formatu.
- Naudotojas gali atsisiųsti sugeneruotas santraukas iš istorijos arba visą istoriją CSV formatu.
- Sistema turi informuoti apie klaidas (pvz., per didelį tekstą, tuščią įvestį).

### **4.2.2. Nefunkciniai reikalavimai**

- Santraukos generavimo trukmė neturi viršyti 5 sekundžių.
- Sistema turi veikti Linux (Ubuntu) operacinėje sistemoje.
- Backend'as turi būti sukurtas naudojant Python (Flask arba Django).
- Frontend'o dalis gali būti sukurta naudojant HTML ir JavaScript.
- Programa turi veikti Matematikos ir Informatikos HPC klasteryje (žr. 1.1).
- Santraukų tikslumo rodikliai turi siekti ROUGE (F1 ir tikslumas) > 0,3.
- Programa turi turėti lietuvišką naudotojo instrukciją.
- Turi būti pasiūlyti dviejus dizaino variantus.
- Sistemos naudotojo sąsaja turi būti pateikta lietuvių kalba.

### **4.2.3. Sistemos ribojimai**

- Įvedamo teksto santraukos sudarymui ilgis turi būti iki 300 žodžių.
- Sistema naudosis tik neregistruoti naudotojai. T.y. sistema neturės administratorių bei prisijungimo prie naudotojo paskyros.



1 pav. Visos sistemos panaudos atvejų diagrama

## 4.3. Programos funkcionalumų aprašas

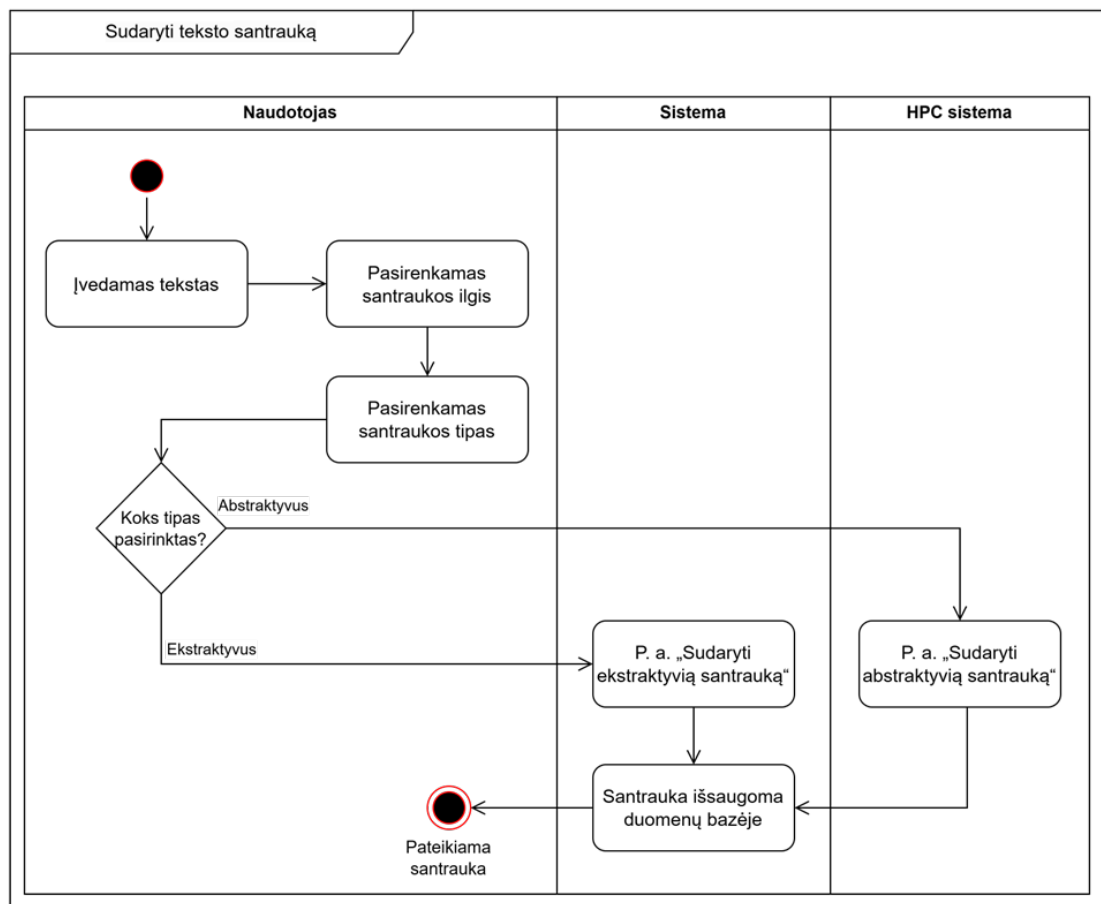
### 4.3.1. Panaudos atvejis „Sudaryti teksto santrauką“

Naudotojas įdeda savo tekstą, pasirenka santraukos parametrus ir pateikia sistemai sudaryti santrauką. Sistema sugeneruoja teksto santrauką. Tai yra pagrindinis sistemos funkcionalumas.

<b>Scenarijai:</b>	
<b>Pagrindinis</b>	<p><b>Teksto ekstraktyvios santraukos užklauso scenarijus</b></p> <ol style="list-style-type: none"> <li>1. Naudotojas įveda arba įkelia tekstą į teksto lauką. Tai jis gali padaryti tekstą įrašydamas arba nukopijuodamas ir įklijuodamas į teksto lauką. Sistema neleidžia įkelti daugiau nei 300 žodžių (žr. 4.2.3).</li> <li>2. Naudotojas pasirinka santraukos tipą „Ekstraktyvi“ (žr. 1.1). Pagal nutylėjimą, nepasirinkus santraukos tipo, jis yra abstraktyvus.</li> <li>3. Naudotojas nusistato pageidaujamą santraukos ilgį. Pagal nutylėjimą, santraukos ilgis yra trumpiausias - 10% įvesto teksto.</li> <li>4. Naudotojas paspaudžia mygtuką „Sutrumpinti“.</li> <li>5. Vykdomas panaudos atvejis „Sudaryti ekstraktyvią santrauką“.</li> <li>6. Rezultatas ir jo parametrai išsaugomi duomenų bazėje.</li> <li>7. Sistema pateikia naudotojui sugeneruotą santrauką.</li> </ol>
<b>Alternatyvūs</b>	
<b>1 alt.</b>	<p><b>Naudotojas pasirinko abstraktyvų santraukos tipą</b></p> <ol style="list-style-type: none"> <li>1. Naudotojas pasirenka santraukos tipą „Abstraktyvi“.</li> <li>2. Naudotojas paspaudžia mygtuką „Sutrumpinti“.</li> <li>3. Vykdomas panaudos atvejis „Sudaryti ekstraktyvią santrauką“.</li> <li>4. Rezultatas ir jo parametrai išsaugomi duomenų bazėje.</li> <li>5. Sistema pateikia naudotojui sugeneruotą santrauką.</li> </ol>
<b>2 alt.</b>	<p><b>Teksto įvesties ilgis pasiekia 300 žodžių:</b></p> <ol style="list-style-type: none"> <li>1. Sistema sustabdo teksto įkėlimą ir simbolių įvedimą.</li> <li>2. Simbolių skaičiaus rodiklis pakeičia spalvą - taip naudotojas gauna klaidos pranešimą apie viršytą teksto ilgį.</li> </ol>
<b>3 alt.</b>	<p><b>Pateiktas tuščias teksto laukas</b></p> <ol style="list-style-type: none"> <li>1. Sistema informuoja naudotoją apie tuščio lauko pateikimą ir prašo įvesti tekstą.</li> </ol>

<b>4 alt.</b>	<b>HPC serverio klaida:</b> <ol style="list-style-type: none"> <li>1. Sistema informuoja naudotoją apie serverio problemą.</li> <li>2. Naudotojas gali bandyti pakartoti užklausą vėliau.</li> </ol>
<b>Sąlygos:</b>	
<b>Prieš</b>	Naudotojas atsidaro santraukos sudarymo langą.
<b>Po</b>	Naudotojas gauna teksto santrauką; Sudaryta teksto santrauka yra išsaugoma duomenų bazėje.
<b>Reikalavimai:</b>	<ul style="list-style-type: none"> <li>• Naudotojas įveda arba įkelia tekstą, kurio ilgis neviršija 300 žodžių.</li> <li>• Naudotojas gali pasirinkti santraukos tipą: ekstraktyvi arba abstraktyvi.</li> <li>• Naudotojas gali nustatyti pageidaujamą santraukos ilgį – nuo 10% iki 50% originalaus teksto ilgio.</li> <li>• Sistema privalo perduoti tekstą ir pasirinktus parametrus HPC sistemai, gauti generuotą santrauką bei grąžinti rezultatą per ne daugiau kaip 5 sekundes.</li> <li>• Sistema privalo saugoti sugeneruotą santrauką, jos tipą bei pasirinktą ilgį duomenų bazėje.</li> <li>• Sistema turi aiškiai informuoti naudotoją apie galimus klaidų atvejus (pvz., per ilgas tekstas, tuščia įvestis, serverio klaida).</li> <li>• Sistema privalo parodyti pradinio teksto ir sugeneruotos santraukos ilgį.</li> <li>• Duomenų bazėje turi būti išsaugomi: originalus tekstas, santrauka, santraukos tipas.</li> </ul>



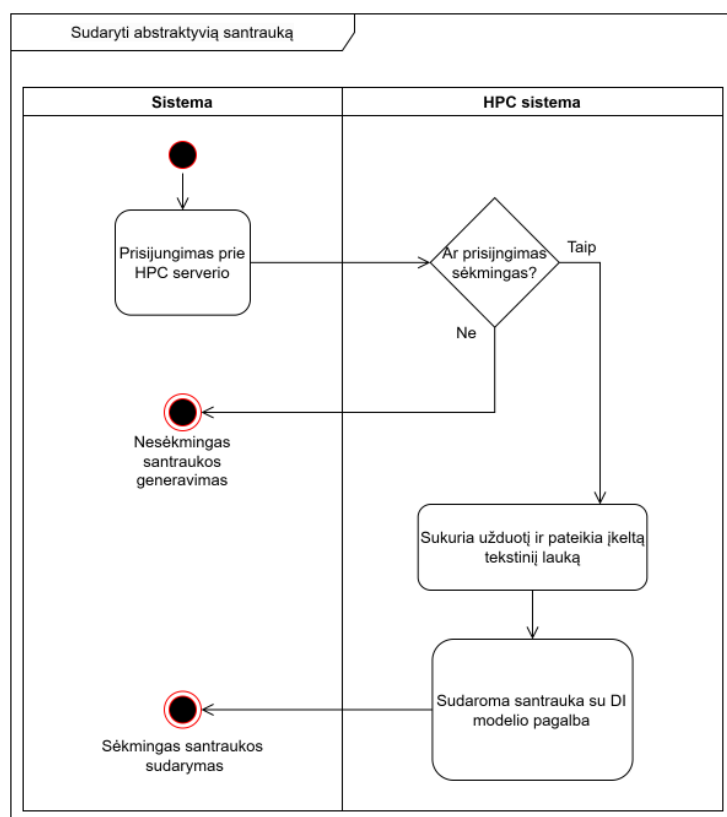


2 pav. Teksto santraukos sudarymo veiklos diagrama

#### 4.3.2. Panaudos atvejis „Sudaryti abstraktyvią santrauką“

<i>Scenarijai:</i>	
<b>Pagrindinis</b>	<b>Sėkmingas abstraktyvios santraukos scenarijus</b> <ol style="list-style-type: none"> <li>1. Sistema SSH protokolu inicijuoja prisijungimą prie HPC sistemos.</li> <li>2. Jei prisijungimas yra sėkmingas, HPC aplinkoje su SLURM sukuriamą naują užduotį su įkeltu tekstu.</li> <li>3. Užduotis perduodama DI modeliui (žr. 1.1), kuris apdoroja tekstą ir sugeneruoja abstraktyvią santrauką.</li> <li>4. HPC sistema grąžina sugeneruotą santrauką programos sistemai.</li> </ol>
<b>Alternatyvūs</b>	

<b>1 alt.</b>	<b>HPC serverio klaida:</b> <ol style="list-style-type: none"> <li>1. Sistema gauna klaidą iš HPC serverio.</li> <li>2. Naudotojas informuojamas, kad reikia bandyti pakartoti užklausą vėliau arba susisiekti su sistemos administratoriumi.</li> </ol>
<b><i>Sąlygos:</i></b>	
<b>Prieš</b>	Naudotojas pasirinko „abstraktyvios“ santraukos tipą, įvedė tekstą ir paspaudė mygtuką „Sutrumpinti“.
<b>Po</b>	Naudotojas mato sugeneruotą abstraktyvią santrauką. Taip pat, sukurta užklausa išsaugoma duomenų bazėje.
<b>Reikalavimai:</b>	<ul style="list-style-type: none"> <li>• Prisijungimas prie HPC sistemos ir abstraktyvios santraukos sudarymas turi būti įvykdytas per ne ilgiau kaip 5 sekundes.</li> <li>• Sugeneruota santrauka turi būti perduota atgal į sistemą ir parodyta naudotojui realiu laiku.</li> <li>• Sistema aiškiai informuoja naudotoją apie galimas klaidas.</li> </ul>

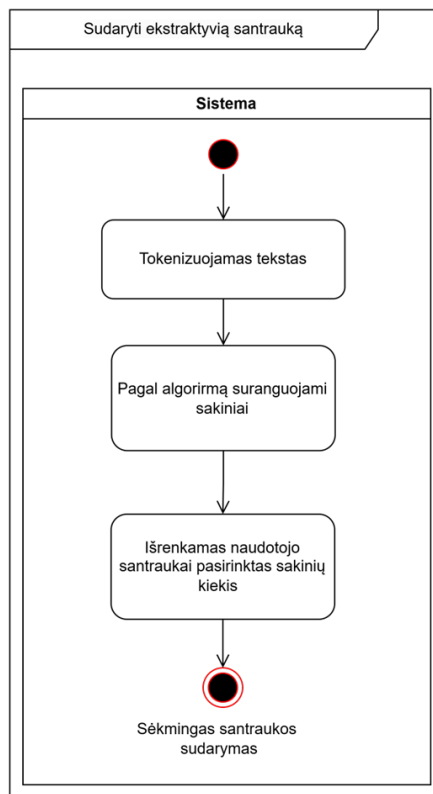


3 pav. Abstraktyvios santraukos sudarymo veiklos diagrama

#### 4.3.3. Panaudos atvejis „Sudaryti ekstraktyvią santrauką“

<i>Scenarijai:</i>	
<b>Pagrindinis</b>	<b>Sėkmingas ekstraktyvios santraukos scenarijus</b> <ol style="list-style-type: none"> <li>1. Sistema gauna naudotojo įkeltą tekstą.</li> <li>2. Tekstas žetonizuojamas – suskaidomas į sakinius ar žodžius.</li> <li>3. Naudojant teksto rangavimo (TextRank) algoritmą, sakiniai suranguojami pagal svarbą.</li> <li>4. Pagal naudotojo pasirinktą santraukos ilgį išrenkamas atitinkamas kiekis aukščiausiai suranguotų sakinių, t.y. svarbiausių sakinių (žr. Ekstraktyvi santrauka 1.1).</li> </ol>
<b>Alternatyvūs</b>	

<b>1 alt.</b>	<b>Algoritmo vykdymo klaida:</b> <ol style="list-style-type: none"> <li>Įvyksta sistemos algoritmo klaida.</li> <li>Rodomas klaidos pranešimas, naudotojas gali pakartoti užklausą.</li> </ol>
<b>Sąlygos:</b>	
<b>Prieš</b>	Naudotojas pasirinko „Ekstraktyvios“ santraukos tipą, įkėlė tekstą ir paspaudė santraukos užklausos mygtuką.
<b>Po</b>	Naudotojui rodoma ekstraktyvi santrauka. Santrauka išsaugota duomenų bazėje. Naudotojui rodomas klaidos pranešimas, naudotojas gali kartoti užklausą.
Sėkmingas	
Nesėkmingas	
<b>Reikalavimai:</b>	<ul style="list-style-type: none"> <li>Ekstraktyvios santraukos sudarymas turi būti įvykdytas ne vėliau kaip per 5 sekundes.</li> <li>Sugeneruota santrauka turi būti perduota atgal į sistemą ir parodyta naudotojui realiu laiku.</li> <li>Sistema aiškiai informuoja naudotoją apie galimas klaidas</li> </ul>



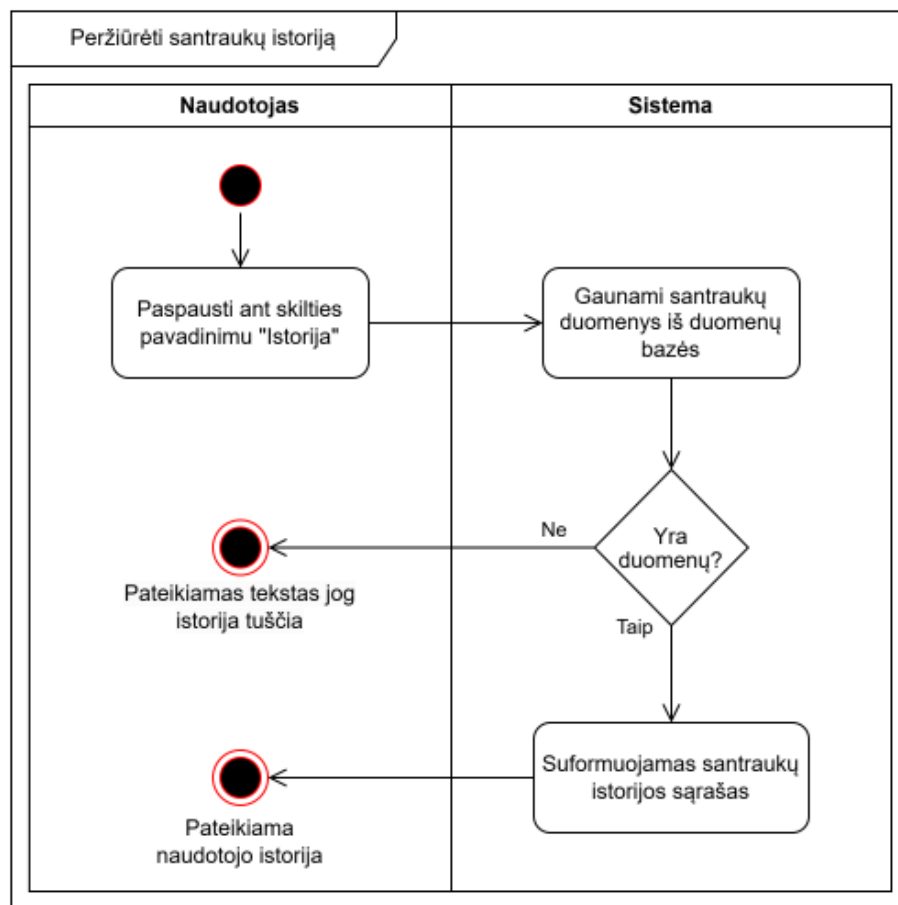
4 pav. Ekstraktyvios santraukos sudarymo veiklos diagrama

#### 4.3.4. Panaudos atvejis „Peržiūrėti santraukų istoriją“

Naudotojas gali peržiūrėti savo anksčiau generuotų santraukų ir jų parametrų istoriją. Santraukų istorijos sekcija matoma kitame lange nei teksto santraukos užklausa. Istorijos informacija yra gaunama iš duomenų bazės (HPC sistemos). Jeigu naudotojas niekada nebuvo užklauses santraukos arba yra ištrynęs istoriją, istorijos sekcija yra tuščia.

<i>Scenarijai:</i>	
<b>Pagrindinis</b>	<b>Pagrindinis istorijos peržiūros sėkmingas scenarijus</b> <ol style="list-style-type: none"> <li>1. Naudotojas atidaro istorijos meniu punktą.</li> <li>2. Sistema kreipiasi į duomenų bazę ir sudaro sąrašą iš anksčiau sugeneruotų santraukų sąrašą.</li> <li>3. Naudotojui pateikiamas sąrašas su įrašų istorija (santaraukos tekstas, santraukos tipas ir pradinis tekstas).</li> </ol>
<b>Alternatyvūs</b>	

<b>1 alt.</b>	<b>Tuščia santraukų istorija:</b> <ol style="list-style-type: none"> <li>1. Sistema santraukų istorijos sekcijoje parodo pranešimą, kad įrašų istorijos nėra.</li> </ol>
<b>2 alt.</b>	<b>Įvyko informacijos gavimo klaida:</b> <ol style="list-style-type: none"> <li>1. Sistema informuoja naudotoją apie įvykusią serverio klaidą gauti duomenis.</li> <li>2. Naudotojas gali bandyti pakartoti užklausą vėliau.</li> </ol>
<b>Sąlygos:</b>	
<b>Prieš</b>	Naudotojas yra praeityje užklausęs vieno ar daugiau teksto santraukų; Naudotojas spaudžia and skilties istorija
<b>Po</b>	Naudotojas gauna teksto santrauką; Sudaryta teksto santrauka yra išsaugoma duomenų bazėje.
<b>Reikalavimai:</b>	<ul style="list-style-type: none"> <li>• Jeigu santraukų istorijos įrašų nėra, sistema turi informuoti naudotoją apie jų nebuvimą.</li> <li>• Istorija turi būti išdėstyta pagal eiliškumą taip, kad pirmiausia matosi naujausi įrašai.</li> </ul>



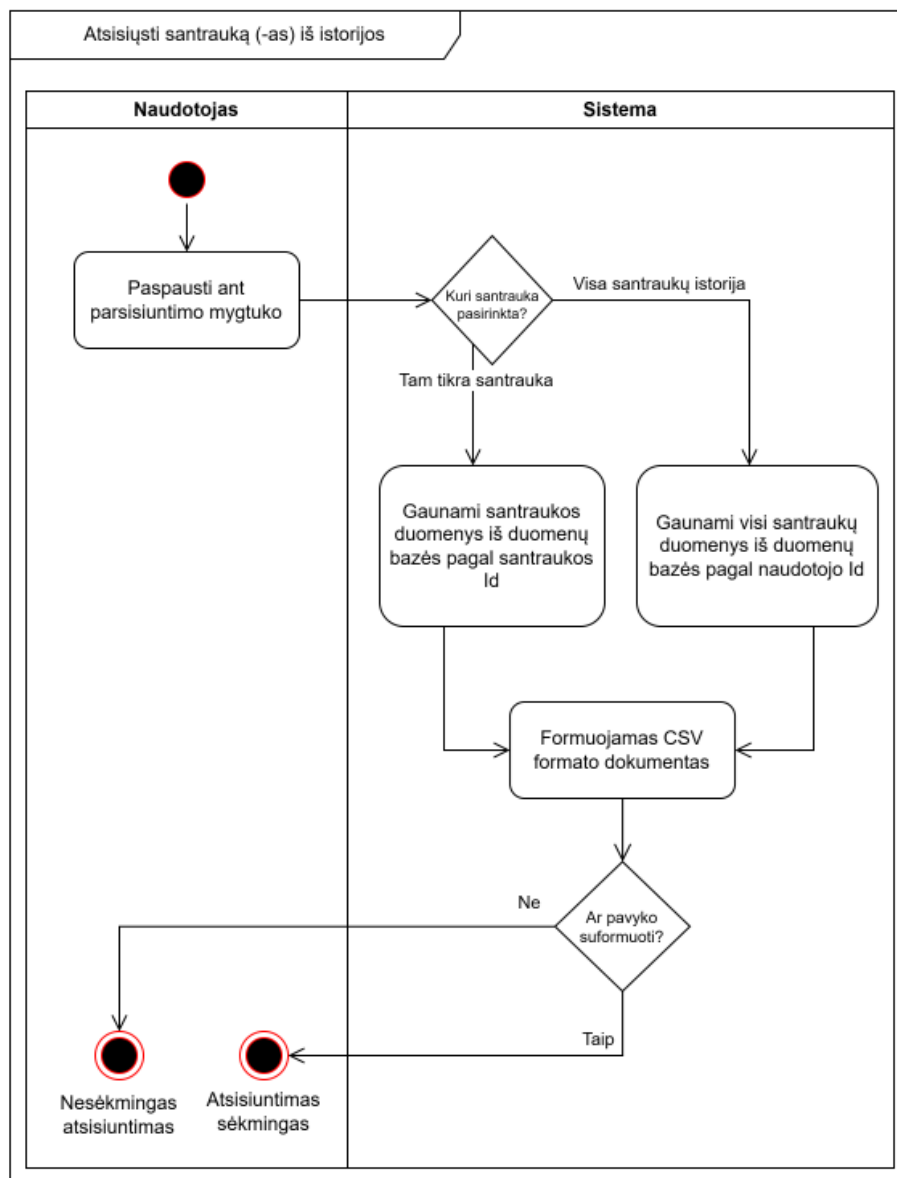
5 pav. Santraukų istorijos peržiūros veiklos diagrama

#### 4.3.5. Panaudos atvejis „Atsisiųsti santrauką (-as) iš istorijos“

Naudotojas gali išsaugoti sugeneruotą tekstą lokaliame diske CSV formatu. Išsaugoti tekstus naudotojas gali pagrindinėje, teksto santraukos, sekcijoje ir, taip pat, bet kurį tekstą iš santraukų istorijos. Išsaugojimas įvyksta paspaudus atsisiuntimo mygtuką prie norimos santraukos arba išsaugojama visa istorija paspaudus mygtuką „Atsisiųsti istoriją“.

<i>Scenarijai:</i>	
<b>Pagrindinis</b>	<b>Santraukų atsisuntimo sėkmingas scenarijus:</b> <ol style="list-style-type: none"> <li>1. Naudotojas paspaudžia atsisuntimo mygtuką šalia sugeneruotos santraukos arba istorijos įrašo.</li> <li>2. Sistema užklausia reikalingos informacijos iš duomenų bazės.</li> <li>3. Suformuojamas CSV formato failas, į kurį įtraukta santrauka, santraukos tipas ir pradinis tekstas.</li> <li>4. Sugeneruotas CSV failas automatiškai parsisiunčiamas į naudotojo lokalų diską.</li> </ol>
<b>Alternatyvūs</b>	
<b>1 alt.</b>	<b>Nutinka techninė problema formuojant CSV failą:</b> <ol style="list-style-type: none"> <li>1. Sistema informuoja naudotoją apie įvykusią sistemos klaidą.</li> <li>2. Naudotojas gali pabandyti iš naujo.</li> </ol>
<b>2 alt.</b>	<b>Įvyko klaida gaunant informaciją iš duomenų bazės:</b> <ol style="list-style-type: none"> <li>1. Sistema informuoja naudotoją apie įvykusią serverio klaidą.</li> <li>2. Naudotojas gali pabandyti iš naujo.</li> </ol>
<i>Sąlygos:</i>	
<b>Prieš</b>	Naudotojas nori atisiųsti sugeneruotą (-as) santraukas ir paspaudžia parsisiuntimo mygtuką.
<b>Po</b>	Santraukos atsiunčiamos ir įrašomos lokaliame naudotojo diske
<b>Reikalavimai:</b>	<ul style="list-style-type: none"> <li>• Sistema privalo surinkti ir į CSV failą įtraukti visus reikalingus duomenis: santraukos tekstą, pradinį tekstą, tipą.</li> <li>• Sugeneruotas failo tekstas turi būti UTF-8-sig formartu.</li> <li>• Suformuotas CSV formato duomenys yra išdėstyti stulpeliais: pradinis tekstas, santraukos tekstas, santraukos tipas.</li> <li>• Naudotojas atsiunčia sugeneruotą santrauką CSV formatu paspaudus atsisuntimo mygtuką.</li> </ul>





6 pav. Santraukų atsisiuntimo iš istorijos veiklos diagrama

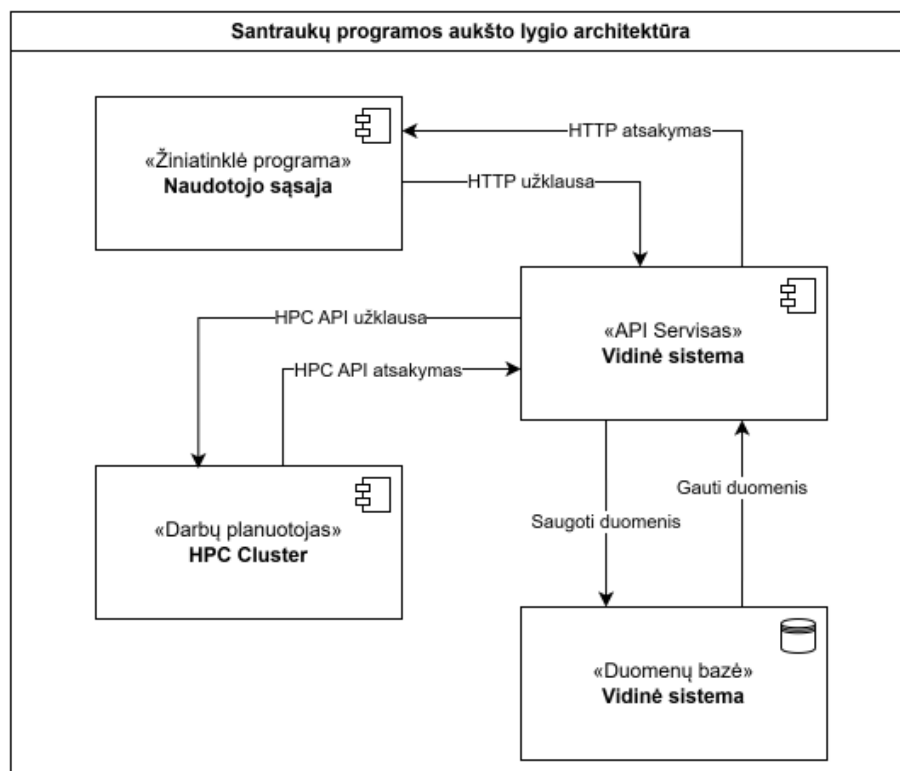
#### 4.4. Programos sistemos komponentų aprašas

Sistema susideda iš trijų pagrindinių komponentų:

- Naudotojo sąsaja (angl. Frontend): Žiniatinklio sąsaja, kurioje naudotojai įveda tekstą, pateikia užklausas santraukų generavimui ir gauna rezultatus.
- Vidinė sistema (API servisas, angl. Backend): Tarpinis sluoksnis, kuris apdoroja naudotojų užklausas, komunikuoja su aukšto našumo kompiuterijos (HPC) sistema ir grąžina sugeneruotas santraukas.
- HPC (Darbų planuotojas): Aukšto pajėgumo skaičiavimo platforma, atsakinga už santraukų generavimo skripto vykdymą naudojant mašininio mokymosi modelius didelio pajėgumo

klasteryje, valdomame SLURM užduočių planavimo sistemos. HPC priskiria reikalingus skaičiavimo išteklius, vykdo mašininio mokymo algoritmą ir siunčia apdorotus rezultatus atgal į vidinę sistemą.

- HPC (Duomenų bazė): Ta pati platforma, kuri naudojama išsaugoti ir laikyti atmintį - duomenų bazė. Kreipimasis į HPC kaip į duomenų bazę, nereikalauja planuoti darbo ir neprašo didelio skaičiavimo resursų.



7 pav. Santraukų sistemos aukšto lygio architektūra

## 4.5. Duomenų bazė

Programėlė saugo tekstų santraukų istorinius duomenis naudodama SQLite duomenų bazę (žr. 1.1), pavadintą `database.db`. Duomenų bazė naudoja vieną lentelę, `Summary_History`, kuri valdoma naudojant SQLAlchemy (žr. 1.1) objektiškai-reliacinį atvaizdavimą (angl. ORM, žr. 1.1). Žemiau pateikiama schemos struktūra ir išsamus kiekvieno stulpelio aprašymas:

Stulpelio pavadinimas	Duomenų tipas	Aprašymas
id	Integer	Pirminis raktas su automatinio didėjimo funkcija. Unikaliai identifikuoja kiekvieną santraukos įrašą.
session_id	String(256)	Universaliai unikalus identifikatorius (UUID), susietas su naudotojo sesija (saugomas kaip saugus slapukas). Tai leidžia atkurti ir rodyti naudotojo asmeninę santraukų istoriją be registracijos.
input_text	Text	Originalus naudotojo pateiktas tekstas santraukai sukurti.
summary_type	String(16)	Nurodo naudotą santraukos algoritmą; gali būti „Abstractive“ arba „Extractive“.
summary_text	Text	Sugeneruotas santraukos tekstas, gautas naudojant pasirinktą santraukos algoritmą.

6 lentelė. Summary\_History lentelės schema

#### 4.5.1. Stulpelių aprašymai ir paskirtis

- **id:** Užtikrina, kad kiekvienas santraukos įrašas būtų unikalus, leidžiant jį atskirai atsisiųsti ar ištrinti.
- **session\_id:** Veikia kaip pakaitalas įprastam naudotojo autentifikavimui, leidžiantis istorijos išsaugojimą tarp sesijų naudojant naršyklės slapuką be registracijos ar prisijungimo.
- **input\_text ir summary\_text:** Atitinkamai saugo originalų ir santraukos tekstą, siekiant palengvinti ateities peržiūras, palyginimus ir atsisiuntimus.
- **summary\_type:** Išsaugomas generuotos santraukos tipas.

#### 4.5.2. Duomenų bazės technologija

SQLite buvo pasirinkta dėl jos paprastumo ir mobilumo. Duomenų bazė nereikalauja išorinio serverio, todėl diegimas ir priežiūra tampa gerokai paprastesni. Duomenų bazės schemas sukūrimas automatiškai atliekamas naudojant SQLAlchemy ORM metodą `db.create_all()` programos inicializacijos metu. Visi duomenys saugomi viename faile, o tai supaprastina atsargines kopijas ir migravimą.

## 4.6. Programos kodo komponentai

### 4.6.1. Šakninis lygmuo

`app.py`. Pagrindinis Flask programos paleidimo ir konfigūracijos failas. Apibrėžia `create_app()` funkciją, nustato `secret_key` ir inicializuoja duomenų bazę.

`db.py`. Apibrėžia SQLAlchemy instanciją: `db = SQLAlchemy()`. Importuojamas modelių ir kontrolių bylose.

### 4.6.2. controllers/

`main_controller.py`. Valdo pagrindinius maršrutus (`/`, `/submit`, `/download`). Apdoroja vartotojo įvestį, kviečia santraukų generavimo paslaugas, saugo rezultatus sesijoje ir duomenų bazėje, atvaizduoja pagrindinį puslapį.

`history_controller.py`. Valdo istorijos puslapį (`/history`) – iš duomenų bazės paima konkretaus `session_id` įrašus, atvaizduoja juos `history.html`, bei apdoroja įrašo arba visos istorijos CSV atsisiuntimus.

### 4.6.3. models/

`SummaryModel.py`. Apibrėžia SQLAlchemy modelį `Summary`, susietą su lentele `Summary_History`. Laukai: `id`, `session_id`, `input_text`, `summary_type`, `summary_length`, `summary_text`, `timestamp`.

### 4.6.4. services/

`AbstractiveSummary.py`. Įgyvendina abstraktyvios santraukos algoritmą, vykdomą per SSH HPC serveryje.

`ExtractiveSummary.py`. Atlieka ekstraktyvios santraukos metodo logiką lokaliajame mašinoje.

`DownloadFormat.py`. Turinio formatavimo ir CSV generavimo funkcijos, naudojamos tiek vieno įrašo, tiek visos istorijos atsisiuntimui.

### 4.6.5. static/

`index.css`. Bendrieji stiliai: spalvų kintamieji, atstatymai, navigacija, pagrindinio puslapio įvesties ir santraukos sekcijos, mygtukai, slider'is.

`history.css`. Specifiniai istorijos puslapio stiliai: kortelių išdėstymas, „Atsisiųsti visą istoriją“ mygtuko bei atsisiuntimo ikonų dizainas.

#### 4.6.6. templates/

**index.html.** Pagrindinio puslapio šablonas: toolbar (santraukos tipas, slider'is), teksto įvesties laukas, santraukos rodinys ir mygtukai.

**history.html.** Istorijos puslapio šablonas: dviejų kolonų kortelės su originaliu tekstu ir santrauka, kortelių pora mygtukų atsisiuntimui, „Atsisiųsti visą istoriją“.

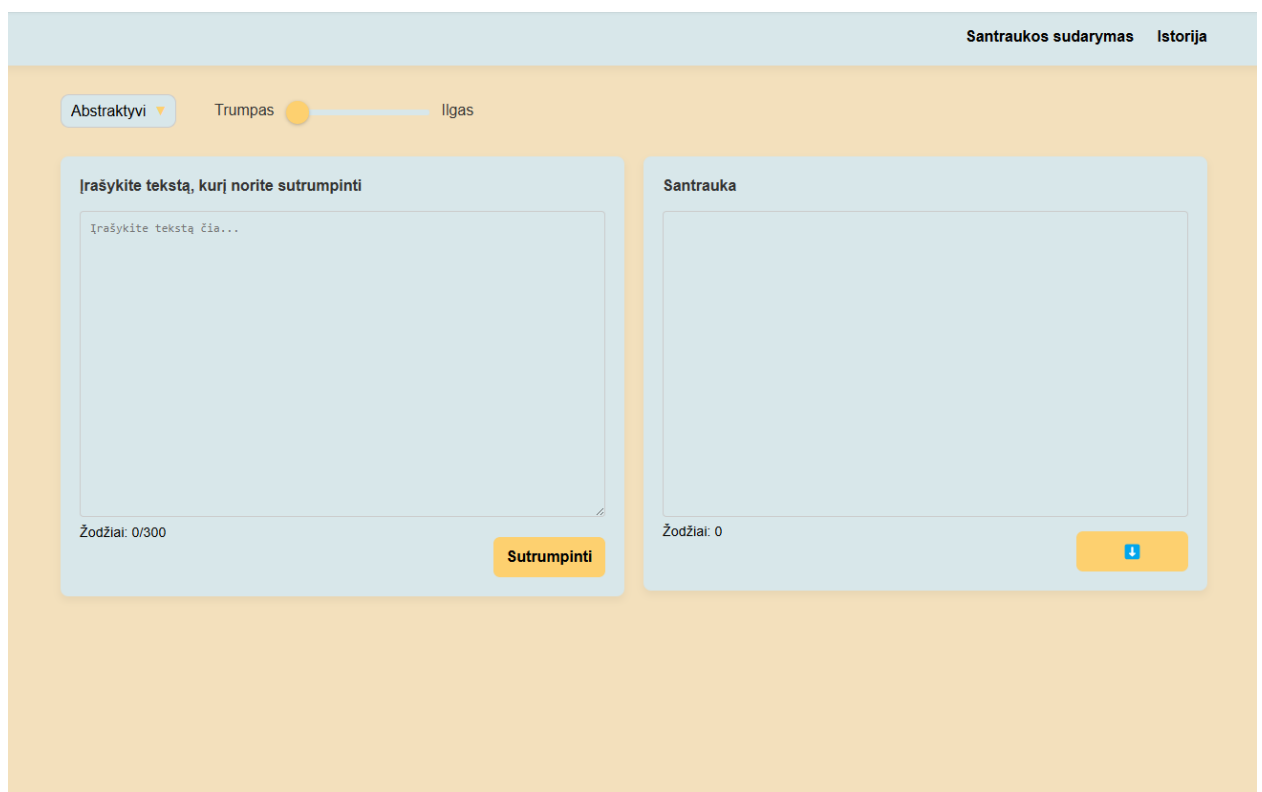
#### 4.6.7. instance/

**database.db.** SQLite duomenų bazės failas, kuriame saugomi visi Summary\_History įrašai. Automatiškai sukuriamas paleidus programą per `db.create_all()`.

#### 4.6.8. summarizer/

**(virtuali aplinka).** Projekto Python virtuali aplinka (venv), kurioje įdiegtos visos priklausomybės (Flask, SQLAlchemy, Paramiko ir kt.). Ši katalogo dalis nėra talpinama versijų valdyme.

### 4.7. Naudojimasis programa

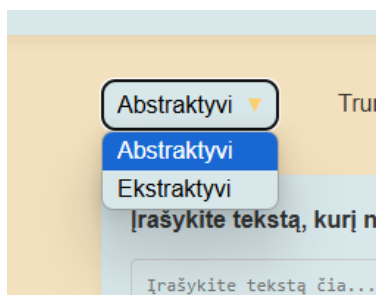


8 pav. Programos pagrindinis meniu

Naujas naudotojas, atsidaręs programos puslapį, mato pagrindinį meniu, pav. 8. Jame galimi keli veiksmai:

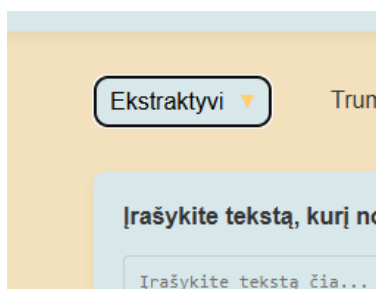
#### 4.7.1. Santraukos tipo pasirinkimas

9 pav. Santraukos tipo pasirinkimas (1)

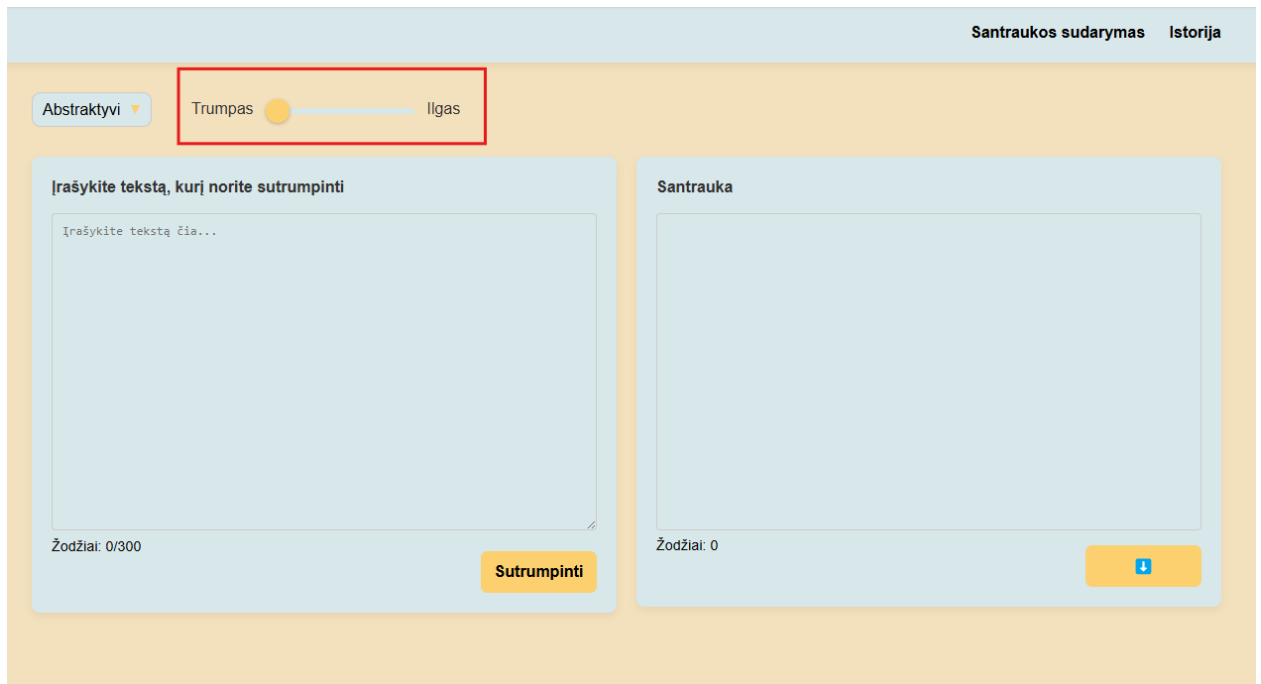


10 pav. Santraukos tipo pasirinkimas (2)

Pasirenkamas ekstraktyvus tipas:



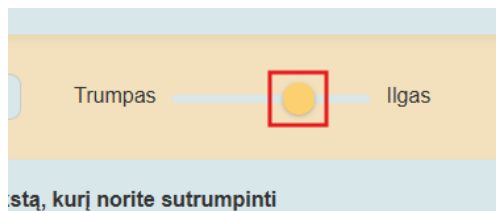
11 pav. Santraukos tipo pasirinkimas (3)



12 pav. Santraukos ilgio pasirinkimas (1)

#### 4.7.2. Pasirenkamas santraukos ilgis

Spaudžiama ant geltono apskritimo ir tempiama iki norimos vietos, pav. 12 ir pav. 13. Arba galima paprasčiausiai paspausti ant norimos vietos (pilkos spalvos plotas):



13 pav. Santraukos ilgio pasirinkimas (2)

### 4.7.3. Teksto santraukos generavimas

Įvedamas norimas tekstas ir spaudžiamas mygtukas „Sutrumpinti“, pav. 14. Sugeneruotas santraukos tekstas rodomas lauke „Santrauka“, pav. 15.

Ekstraktyvi ▼    Trumpas    Ilgas

[rašykite tekstą, kurį norite sutrumpinti]

Equality is a worthy goal—but its tough to achieve when unconscious bias so pervades the American workplace. Certainly women have made inroads in corporate America, but a Pew Research Center survey released Wednesday points at why women struggle to climb to the corporate world's highest ranks—and often tone down their ideas, hide behind an agreeable facade or leave the workplace altogether. Four out of 10 surveyed in the Pew study said that there are double standards for women seeking the highest levels of leadership in politics or business. They added that women have to outshine their male counterparts—and more than one-third of respondents believe the electorate and corporate America are not ready to put more women in top leadership positions. That conclusion comes as no surprise to Howard J. Ross, a diversity

Žodžiai: 132/300

**Sutrumpinti**

14 pav. Teksto santraukos sudarymas (1)

**Santrauka**

Certainly women have made inroads in corporate America, but a Pew Research Center survey released Wednesday points at why women struggle to climb to the corporate world's highest ranks—and often tone down their ideas, hide behind an agreeable facade or leave the workplace altogether.

Žodžiai: 44

↓

15 pav. Teksto santraukos sudarymas (2)



#### 4.7.4. Istorijos peržiūra

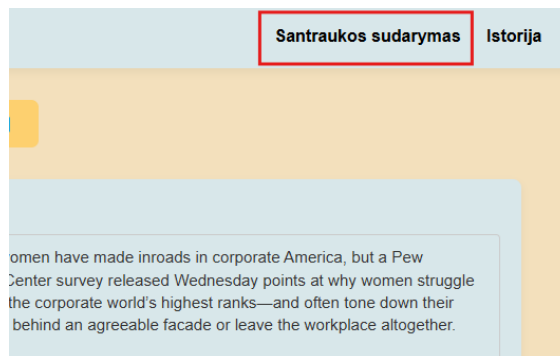
Spaudžiaumas navigacijos mygtukas „Istorija“, pav. 16.

16 pav. Istorijos peržiūra (1)

Istorijos skiltyje matoma santraukų užklausų istorija, pav. 17.

17 pav. Istorijos peržiūra (2)

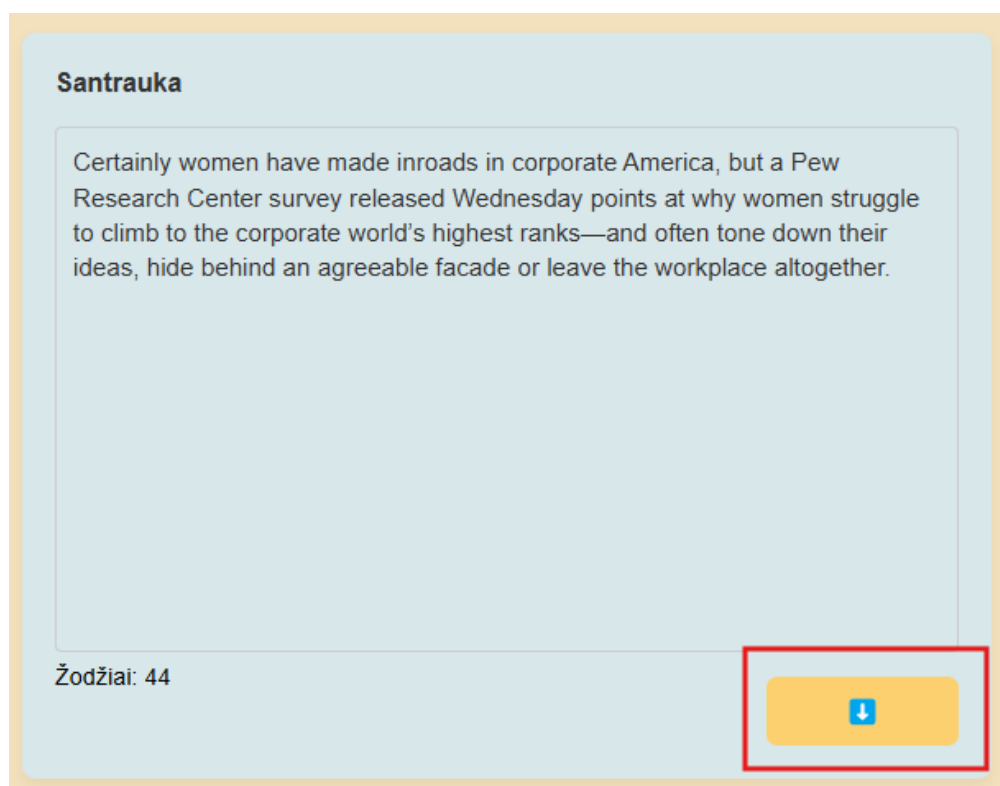
Tam, kad būtų grįžtama į tekstą santraukų sudarymo sekciją, spaudžiamas navigacijos mygtukas „Santraukos sudarymas“, pav. 18.



18 pav. Istorijos peržiūra (3)

#### 4.7.5. Sugeneruotos santraukos atsisiuntimas

Norint atsisiųsti santrauką ir jos duomenis, spaudžiami atsisiuntimo mygtukai, pav. 19 ir pav. 20. Norint atsisiųsti visą santraukų istoriją, spaudžiamas mygtukas „Atsiųsti visą istoriją“, pav. 21.



19 pav. Santraukos atsisiuntimas (pagrindinis meniu)

Santraukos sudarymas Istorija

Atsisiųsti visą istoriją


Originalus tekstas

Equality is a worthy goal—but its tough to achieve when unconscious bias so pervades the American workplace. Certainly women have made inroads in corporate America, but a Pew Research Center survey released Wednesday points at why women struggle to climb to the corporate world's highest ranks—and often tone down their ideas, hide behind an agreeable facade or leave the workplace altogether. Four out of 10 surveyed in the Pew study said that there are double standards for women seeking the highest levels of leadership in politics or business. They added that women have to outshine their male counterparts—and more than one-third of respondents believe the electorate and corporate America are not ready to put more women in top leadership positions. That conclusion comes as no surprise to Howard J. Ross, a diversity

Santrauka

Certainly women have made inroads in corporate America, but a Pew Research Center survey released Wednesday points at why women struggle to climb to the corporate world's highest ranks—and often tone down their ideas, hide behind an agreeable facade or leave the workplace altogether.

Santraukos tipas: Ekstraktyvus




Originalus tekstas

Equality is a worthy goal—but its tough to achieve when unconscious bias so pervades the American workplace. Certainly women have made inroads in corporate America, but a Pew Research Center survey released Wednesday points at why women struggle to climb to the corporate world's highest ranks—and often tone down their ideas, hide behind an agreeable facade or leave the workplace altogether. Four out of 10 surveyed in the Pew study said that there are double standards for women seeking the highest levels of leadership in politics or business. They added that women have to outshine their male counterparts—and more than one-third of respondents believe the electorate and corporate America are not ready to put more women in top leadership positions. That conclusion comes as no surprise to Howard J. Ross, a diversity

Santrauka

a pew research center survey points at why women struggle in the workplace.

Santraukos tipas: Abstraktyvus



20 pav. Santraukos atsisiuntimas (Istorijos meniu)

Atsisiųsti visą istoriją

Santrauka

achieve when unconscious bias so

Certainly women have made inroad

21 pav. Istorijos atsisiuntimas (Istorijos meniu)

## 5. Išvados

Atliekant praktika nepavyko atlikti visų užsibrėžtų užduočių, t.y. nėra išpildyti visi programos užsakovo reikalavimai. Tačiau, atsižvelgiant į praktikos laikotarpį bei uždavinių apimtį, išpildyti visas užduotis, iš tiesų, yra sudėtinga. Tad, manau, jog praktika atlikta gerai.

Kadangi praktikos apimtis yra didelė, išmokau ir gavau patirties keliose didelėse srityse, tokiose kaip: programų sistemų kūrimas ir programavimas, programų sistemų architektūros projektavimas, programų sistemų dizaino kūrimas, mašininis mokymas ir tekstų santraukų modeliai bei algoritmai. Taip pat, reikėtų įvardinti įgautą patirtį iš smulkesnės pusės: išmokau kurti internetinę svetainę naudojant Python programavimo kalbą bei Flask karkasą, įgavau patirties naudotis programų sistemų projektavimo įrankiais Figma, Draw.io bei Jupyter aplinka mašininio mokymo modelių naudojimui.

Platus sričių spektras yra didelis privalumas dėl to, kad sužinoma, kuri sritis patinka, o kuri nepatinka. Pavyzdžiui, man labai patiko rašyti programos kodą, truputi patiko sistemos projektavimas, bet mažiausiai patiko DI modelių teorijos nagrinėjimas. Supratau, jog naudoto sąsajos kodo rašymas nėra toks neįdomus ir sunkus, kaip maniau prieš atlikdamas šią praktiką.

## Literatūra

- [BKL09] Steven Bird, Ewan Klein ir Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, 2009.
- [BL05] Satanjeev Banerjee ir Alon Lavie. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, p.p. 65–72, 2005.
- [Con24] SQLite Consortium. *SQLite Documentation*. <https://www.sqlite.org/docs.html>. 2024.
- [Fig24] Inc. Figma. *Figma: Collaborative Interface Design Tool*. <https://www.figma.com>. 2024.
- [Fou24] Django Software Foundation. *Django Documentation*. <https://docs.djangoproject.com>. 2024.
- [Hea24] Heartex. *Label Studio: Data Labeling and Annotation Tool*. <https://labelstudio.io>. 2024.
- [HM17] Matthew Honnibal ir Ines Montani. spaCy: Industrial-strength Natural Language Processing in Python. *Zenodo*, 2017.
- [YJG03] Andrew B. Yoo, Matthew A. Jette ir Michael Grondona. SLURM: Simple Linux Utility for Resource Management. *Lecture Notes in Computer Science*, 2862:44–60, 2003.
- [YL06] Tatu Yl'onen ir Chris Lonvick. *SSH Security Architecture*. IETF RFC 4251. 2006.
- [YZL<sup>+</sup>21] Jing Yan, Chenguang Zhao, Ye Li, Jing Jiao, Xipeng Han, Jianfeng Gao, Bowen Zhou, Lei Li ir Yunfan Zeng. ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, p.p. 2401–2410, 2021.
- [Jup24] Project Jupyter. *JupyterHub Documentation*. <https://jupyter.org/hub>. 2024.
- [Kag24] Inc. Kaggle. *Kaggle: Your Machine Learning and Data Science Community*. <https://kaggle.com>. 2024.
- [Lin04] Chin-Yew Lin. ROUGE: A Package for Automatic Evaluation of Summaries. *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*:74–81, 2004.
- [LL19] Yang Liu ir Mirella Lapata. Fine-tune BERT for Extractive Summarization. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, p.p. 3721–3731, 2019.

- [MT04] Rada Mihalcea ir Paul Tarau. TextRank: Bringing Order into Texts. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, p.p. 404–411. Association for Computational Linguistics, 2004.
- [Pro24] Pallets Projects. *Flask Documentation*. <https://flask.palletsprojects.com>. 2024.
- [PRW<sup>+</sup>02] Kishore Papineni, Salim Roukos, Todd Ward ir Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, p.p. 311–318, 2002.
- [QZZ<sup>+</sup>20] Pengfei Qi, Yuhao Zhang, Yuhui Zhang, Joshua Bolton ir Christopher D. Manning. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *Association for Computational Linguistics*:1011–1027, 2020.
- [Vil25a] Vilniaus Universiteto MIF. *Qluster HPC klasteris dokumentacija*. Internal documentation. 2025.
- [Vil25b] Vilniaus Universiteto MIF. *Vilniaus Universiteto Matematikos ir Informatikos Fakulteto HPC dokumentacija*. Internal documentation. 2025.
- [ZKW<sup>+</sup>19] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger ir Yoav Artzi. BERTScore: Evaluating Text Generation with BERT. *International Conference on Learning Representations*, 2019.

## **Priedas Nr. 1**

### **Teksto Santraukos programos GitHub Repozitorija**

<https://github.com/Green-255/TextSummarizer.git>