KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
KOMPIUTERINIŲ TINKLŲ KATEDRA

Olegas Strašnovas

Plagijavimo nustatymas taikant semantinės pranešimų analizės metodą

Magistro darbas

Darbo vadovas
doc. dr. B.Tamulynas

Kaunas, 2006

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
KOMPIUTERINIŲ TINKLŲ KATEDRA

Olegas Strašnovas

Plagijavimo nustatymas taikant semantinės pranešimų analizės metodą

Magistro darbas

Kalbos konsultantas
lekt. J. Jonušas
2006-05

Vadovas
doc. dr. B. Tamulynas
2006-05

Recenzentas
doc. dr. E. Karčiauskas
2006-05

Atliko
IFM-3 gr. stud.
Olegas Strašnovas
2006-05-07

Kaunas, 2006

Turinys

1. ĮVADAS

Šiuolaikinėje mokslo sistemoje labai akcentuojamas savarankiškas besimokančiųjų darbas. Labiausiai paplitęs mokslo žinių lygio patikrinimo metodas yra rašomieji darbai: referatai, kontroliniai darbai, kursiniai ir diplominiai darbai. Be to, pagal statistiką, iki 80% visų studentų nors vieną kartą atiduodavo dėstytojui svetimą darbą kaip savo. Yra labai daug baigtų darbų bet kokiomis temomis, kurias galima nemokamai parsisiųsti internetų. Dėstytojas dažnai net nenutuokia, kad jam duotas darbas yra plagiatas. Tokiu būdu, dėstytojai neturi galimybės objektyviai įvertinti savo studentų žinių lygio, kita vertus studentai patys negauna jiems būtinų žinių. Mokymosi procese plagiatas tampa jiems įprasta norma. Kitais žodžiais sakant, plagiatas pakerta visą akademinę mokslo sistemą [2]. Auga karta, nesugebanti mąstyti savarankiškai, kuriai plagiatas – įprasta norma. Paskutiniais statistikos duomenimis kas antras studentas atsiskaitydamas bent kartą yra pasinaudojęs svetimais darbais. Todėl kova su plagiatu tampa vis aktualesnė problema.

Paprasčiausi būdai, kuriais naudojasi dėstytojai, siekdami atskleisti anksčiau minėtus atvejus yra šie:

- **Teksto rašymo stiliaus pasikeitimas** – kai tekste akivaizdžiai matomi atskirų teksto dalių rašytinio stiliaus nesutapimai.
- **Skyryba** – mažai tikėtina, kad skirtingi autoriai naudotų sintaksės žinias vienodai.
- **Bendras tekstų panašumas**– kai lyginami darbai, parašyti viena tema, jie visad bus šiek tiek panašūs: vardai, įvykiai, ir t.t..
- **Gramatinės klaidos** – irgi vienas iš paprastų būdų įtarti plagijavimą. Labai retai du skirtingi autoriai daro panašias gramatines klaidas.
- **Žodžių panaudojimas** – kitas būdas. Kartais plagiatą galima aptikti atkreipus demesį į tam tikrų žodžių dažną pasikartojimą skirtinguose tekstuose.
- **Sintaksinė teksto struktūra** – nustatyti plagijavimą galima ir atsižvelgiant į sintaksinę tekstų struktūrą.
- **Teksto aiškumas** – plagijavimo nustatymui galima panaudoti tokią metrinę charakteristiką, kaip Flesch'o skaitymo lengvumo formulė arba SMOG indeksą. Ja nustatomas bendras teksto įskaitomumo indeksas. Vienodi skirtingų autorių tekstų indeksai gali būti laikomi plagiato įrodymu.

Bet ar įmanoma būtų supaprastinti užduotį, automatizuojant plagiato nustatymo procedūrą? Šiame darbe pabandysime panagrinėti kaip ir kokias būdais būtų galima sukurti

tokią programinę įrangą, kuria, dėstytojai galėtų analizuoti pateiktus studentų rašto darbus, siekdami nustatyti plagijavimo atvejus.

2. ANALITINĖ DALIS

Visų pirma apžvelgime jau egzistuojančias automatizuoto tekstų palyginimo sistemas [8].

2.1. UŽDAVINIO APLINKOS ANALIZĖ

Plagiarism.org, Turnitin.com

Nuoroda

http://plagiarism.org, http://turnitin.com

Aprašymas

Naudojimosi būdas: rašto darbai analizei sistemai pateikiami per vartotojo sąsają Turnitin.com portale. Panaudojant daugelę unikalių algoritmų, sistema palygina pateiktus darbus su šimtais tūkstančių dokumentų, esančių Turnitin.com portalo duomenų bazėse bei kitose partnerių saugyklose. Ataskaita el.paštu nusiunčiama vartotojui. Pranešime grąžinami atgal pateikti dokumentai, kuriuose "sutampančios" vietos pažymėtos kita spalva ir pridėtos nuorodos į dokumentus, kuriuose šie sutapimai buvo rasti. Kartu pateiktiems dokumentams sistema po analizes generuoja tam tikrą sutapimo indeksą, kuris yra įtraukiamas į ataskaitą.

Kaina

Paslauga mokama, bet bandymo metu pirmosios 10 ataskaitų nemokamos.

Neigiamos charakteristikos

- Sistema mokama.
- Ataskaitos paruošiamos tik po 12 valandų nuo pateikimo.
- Būtina turėti priėjimą prie interneto tinklo.

EVE2

Nuoroda

http://www.canexus.com/eve/index.shtml

Aprašymas

Šią programą būtina įdiegti į vartotojo kompiuterį. Dokumentai įkraunami iš kietojo disko elektroninių laiškų pavidalu. Programa palaiko šiuos dokumentų tekstinius formatus – Paprastas tekstas, Microsoft Word, Word Perfect. Programoje numatyti keli analizės lygiai, dėl to skiriasi analizės sudėtingumas bei keičiasi nuskaitymo laikas. Jei plagijavimo atvejis nustatytas, įrašas apie tai pateikiamas ataskaitoje su nuoroda į sutapusią vietą.

Kaina

Asmeninė licenzija kainuoja 20$, licenzija įmonei kainuoja 400$. Bandymams yra nemokama 15 dienų programos versija.

Neigiamos charakteristikos

- Programa yra mokama
- Kadangi reikalaujama įdiegimo – kai kurie vartotojai negalės ja naudotis
- Programa yra sukurta Kanados kompanijos, dėl to programa kai kuriuos žodžius gali interpretuoti netaisyklingai.

CopyCatchGold

Nuoroda

http://www.copycatch.freeserve.co.uk/index.html

Aprašymas

Tai dar viena internetinė splagijavimo nustatymo Interneto sistema. Užsakovui atsiunčiama ataskaita. Pasak kūrėjų, sistema gali atpažinti atvejus, kai studentai pakeičia žodžių sekas sakinyje arba vartoja sinonimus.

Kaina

250 £ kasmet vienam vartotojui.

Neigiamos charakteristikos

- Labai aukšta kaina
- Ilgas ataskaitų pateikimo laikas

WordCheck

Nuoroda

http://www.wordchecksystems.com

Aprašymas

Ši programa analizuoja žodžių panaudojimą tekste bei žodžių panašumus. Analizei naudoja asmeninę vartotojo „biblioteką“. Vartotojas naudodamas šią programą gali nustatyti, ar vieno autoriaus rašyti darbai yra panašūs tarpusavyje ar turi panašumų su kitų autorių darbais.

Kaina

Paprasta versija - $295 (bibliotekoje gali būti iki 1000 dokumentų)
Profesionali versija - $1295 (bibliotekoje gali būti iki 5000 dokumentų)

Neigiamos charakteristikos

- Labai aukšta programos kaina
- Tik anglų kalba

WCopyfind

Nuoroda

http://www.plagiarism.phys.virginia.edu

Aprašymas

Šią programą sukūrė Virginijos universiteto fizikos profesorius Lou Bloomfield. Programa analizuoja grupę vartotojo nurodytų dokumentų ir išspausdina sutampančius teksto fragmentus.

Kaina

Nemokama, tačiau naudojimasis ja yra ribojamas atsižvelgiant į vietą.

Neigiamos charakteristikos

- Programa gali surasti tik visiškai sutampančius frazės ar teksto fragmentus
- Licenzijos apribojimai

Glatt

Nuoroda

http://plagiarism.com

Aprašymas

Programa remiasi Wilson Taylor procedūra. Iš įtartino dokumento išimamas kas 5 žodis ir pakeičiamas tuščiais simboliais, paliekant tą patį žodžio ilgumą. Sutapimas įvertinamas atsižvelgiant į paieškos tikslumą ir laiką, per kurį tušti tarpai buvo užpildyti trūkstamais žodžiais.

Kaina

300$, jei perkama tik analizės programa, ir 250$, kai perkama kartu su plagijavimo nustatymo apmokymo programa.

Neigiamos charakteristikos

- Mokama programa
- Ši programa ne visai tinka plagijavimo nustatymui, ji daugiau skirta plagijavimo nustatymo apmokymams.

Urkund

Nuoroda

http://www.urkund.com/UK/

Aprašymas

Urkund programa turi galimybę atlikti analizę ir automatiškai bei visiškai kontroliuoti procesą.

Visiškai kontroliuojama – kai dėstytojas pats ieško su tam tikrais įrankiais pagalba arba pats ieško sutapimų Urkund asmeninėje duomenų bazėje arba automatiškai kontroliuojama – kai studentai patys siunčia savo darbus nurodytu el. pašto adresu, o dėstytojas gauna tik patikrinimo rezultatus.

Kaina

Mokama. Dėl kainos reikia derėtis su Urkund atstovais.

Neigiamos charakteristikos

- Galimi intelektinės nuosavybės praradimai.
- Mokama.
- Bazėje ne visi šaltiniai originalūs.

2.2. UŽDUOTIES FORMULAVIMAS

Išanalizavus esamus programinius produktus, prieita prie išvados, kad sukurta nemažai programinių produktų,padedančių nustatyti plagiatą, bet šių produktų vartojimas susijęs su tam tikrais nepatogumais.

Sistemos, skirtos naudojimuisi interneto tinklų pasižymi tuo, kad sugeba analizuoti milžiniškus duomenų kiekius, bet tuo pačiu tokios analizės pasekmė - jų analizės laikas gana ilgas, t.y. nuo pusės dienos iki keletos parų ar netgi savaičių.

Sistemos, skirtos naudojimuisi vartotojų darbo stotyse (asmeniniuose kompiuteriuose), apsiriboja tik vartotojo duomenų šaltiniais (analizuoja tik tuos dokumentus, kuriuos nurodo pats vartotojas), bet visos šios sistemos yra komercinės, didžioji dauguma turi daug papildomų funkcijų, ne visai susietų su plagijavimo nustatymo užduotimi dėl to šių programinių produktų įsigijimo kainos gana didelės ir nevienkartinės, o pasibaigus licenzijos laikotarpiui mokestis už naudojimąsi jomis vėl imamas.

Be to, visos šios sistemos daugiausiai analizuoja tik galutinio (visiškas teksto sutapymas) arba dalinio (sutampa tik atskiri teksto dalys) plagiato atvejus. Bet tai gana "paprasti" plagijavimo atvejai. Tačiau yra dar kitas plagijavimo būdas, kai tekstas yra perrašomas taip, kad panašus į originalą jis būna tik savo prasme (stiliaus plagiatas).

Šio darbo tikslas – išbandyti Prasminio palyginimo (Semantinės analizės) metodą, palyginti jo veikimą su kitais analizės metodais, pateikti išvadas dėl šio metodo taikymo racionalumą.

Tam, kad tai būtų galima atlikti, turi būti sukurta automatinio tekstų palyginimo programa, kurioje bus realizuoti analizės metodai. Papildomai reikėtų apsispręsti, pagal kokius kriterijus atlikinėti palyginimą, t.y. kokius plagijavimo atvejus bus siekiama aptikti.

Nagrinėsime darbų plagijavimą toliau išvardintais būdais:

Yra kelios plagiato rūšys [1]. Dauguma studentų atpažįsta vieną formą vadinamą ją "Copy ir Paste plagiatas", bet tai yra viena iš akivaizdžiausių rūšių.

I rūšis: **Copy ir Paste**
II rūšis: **Žodžių sukeitimas**
III rūšis: **Stilius**
IV rūšis: **Idėja**

I rūšis: **Copy ir Paste** plagiatas

Aprašymas: kai iškeliamas sakinys arba reikšminga baigtinė frazė iš straipsnio.

Pavyzdys:

Originalas	Plagiatas
Especially since the launch of HST and the unprecedented clarity of the images satellites have given us, you've all seen on the news or in books, beautiful color pictures of various sights in the cosmos. But is this the way you would see these objects if you went there? Well, to tackle that question, first we have to consider the nature of light and color. Light is made of waves of electromagnetic radiation. We perceive different wavelengths of visible light as different colors.	Everyone is interested in astronomical images, especially since the launch of HST and the unprecedented clarity of the images satellites have given us. But is this the way you would see these objects if you went there?

II rūšis: Žodžių sukeitimo plagiatas

Aprašymas: kai paimamas sakinys iš straipsnio ir pakeičiami keli žodžiai, tai taip pat plagiatas.

Pavyzdys:

Originalas	Plagiatas
All solid bodies emit light: stars, rocks and people included. The temperature of the star, rock or person determines which wavelength of light will be most strongly radiated. In the constellation Orion, the upper left star is Betelgeuse (Armpit of the giant), 520 l-y distant. Betelgeuse is a supergiant star, 14,000 times brighter than our sun, and so big, if you were to put Betelgeuse in place of our sun, its surface would reach all the way out to Jupiter. Betelgeuse's color is bright red. On the other hand, another supergiant star, Rigel, with a luminosity 57,000 times that of the sun, appears whitish-blue. The reason that Betelgeuse is red and Rigel is blue is that their surface temperatures are different. Hot stars at 30,000 degrees emit a lot more blue light than red light, and so hot stars look blue or bluish-white. Cool stars at 3,000 degrees give off more red light than blue, and so these stars look red.	Stars, rocks and people all emit light, and which wavelength of light will be most strongly radiated depends on the temperature of the star, rock or person. For example, the star Betelgeuse in the constellation Orion, Armpit of the Giant, is a supergiant star, 14,000 times brighter than our own sun.

III rūšis: **Stiliaus plagiatas**

Aprašymas: kada „Originalo tekstas" naudojamas nuo sakinio iki sakinio, arba nuo paragrafo iki paragrafo, tai yra plagiatas, netgi kai nė vienas iš sakinių nesutampa su „Originalo tekstu".

Pavyzdys:

Originalas	Plagiatas
Especially since the launch of HST and the unprecedented clarity of the images satellites have given us, You've all seen on the news or in books, beautiful color pictures of various sights in the cosmos. But is this the way you would see these objects if you went there? Well, to tackle that question, first we have to talk about the nature of light and color. Light is made of waves of electromagnetic radiation. We perceive different wavelengths as different colors. All solid bodies emit light: stars, rocks and people included. The temperature of the star, rock or person determines which wavelength of light will be most strongly radiated. In the constellation Orion, the upper left star is Betelgeuse (Armpit of the giant), 520 l-y distant. Betelgeuse is a supergiant star, 14,000 times brighter than our sun. and so big, if you were to put Betelgeuse in place of our sun, its surface would reach all the way out to Jupiter. Betelgeuse's color is bright red. On the other hand, another supergiant star, Rigel, with a luminosity 57,000 times that of the sun, appears whitish-blue. The reason that Betelgeuse is red and Rigel is blue is that their surface temperatures are different. Hot stars at 30,000 degrees emit a lot more blue light than red light, and so hot stars look blue or bluish-white. Cool stars at 3,000 degrees give off more red light than blue, and so these stars look red.	The beautiful pictures that the space telescope has given us show spectacular color. But is the color real? First, we have to consider what light and color are. Different wavelengths of light correspond to different colors, and light is called electromagnetic radiation. The temperature of an object determines the color of light emitted, and all things, including people, emit light. In the constellation Orion, the star Betelgeuse is a huge, giant star, as big as the orbit of Jupiter. Betelgeuse is red. Another star in Orion, Rigel, is blue. The reason that they are different colors is that they each have a different surface temperature. Cold stars are at about 3,000 degrees and emit more red than blue light and very hot stars emit blue light since they have temperatures of about 30,000 degrees.

IV rūšis: **Idėjos plagijavimas**

Aprašymas: jei „Pirminio teksto" autorius išreiškia kūrybinę idėją arba nurodo problemos sprendimą, idėja ar sprendimas turi būti būtinai priskirti autoriui. Studentams atrodo sunkios skiriamosios autoriaus idėjos ir/arba sprendimai iš *public domain information* (viešos srities informacija). Viešos srities informacija yra kiekviena idėja ar sprendimas apie žinomą išvadą srities, kurioje jis stipriausias. Pavyzdžiui, kas per juodoji skylė ir kaip ji apibūdinta yra pagrindinis žinojimas. Tau nereikia nurodyti pagrindinio juodosios skylės apibrėžimo. Antrasis žemės kosminis greitis taip pat yra pagrindinės žinios ir nereikalauja apibrėžimo. Atstumas iki

Galaktikos centro taip pat yra pagrindinės žinios. Vis dėlto nauja idėja, kaip ieškoti juodųjų skylių arba naujų fizikos sprendimų turi būti priskiama autoriams.

Pavyzdys:

Originalas	Plagiatas
Until now, infrared carbon stars have been classified as such due to either the presence of carbon-rich dust or to their presence in region VII of the Habing diagram. Our visible spectra show conclusively that these stars are true carbon stars and do not have any O-rich molecules in their atmospheres. Their weak Ba lines might indicate an under-abundance of *s*-process elements. This important result, if true, would certainly separate infrared carbon stars from the optical population.	Infrared carbon stars show weak Ba lines and this might mean that they do not have the normal amount of s-process elements in their atmospheres, making them decidedly a different type of star.

Idėjos plagijavimas nebus nagrinėjamas, nes tai reikalauja visiškai kitokios tyrimo metodikos. Lyginimui bus panaudoti 2 tekstų palyginimo metodai: Netikslaus palyginimo ir Tapatingo palyginimo metodai. Visų 3 metodų aprašymai pateikiami toliau.

2.2.1. NETIKSLAUS PALYGINIMO METODAS

[4] Netikslaus palyginimo funkcija, kaip argumentą naudoja dvi eilutes ir palyginimo parametrą – maksimalų eilučių lyginimo ilgį. Funkcijos darbo rezultatas yra skaičius, kuris yra ribose nuo 0 iki 1. Kur 0 atitinka pilną dviejų eilučių neatitikimą, o 1 – pilną eilučių atitikimą.

Eilučių palyginimas atliekamas pagal sekančią schemą. Pavyzdžiui, kaip argumentai užduotos dvi eilutės - “test” ir “text” bei maksimalus eilučių ilgis, sakykim 4. Palyginimo funkcija sudaro visas galimas eilučių kombinacijas, kurių ilgis yra iki nustatyto (4), ir paskaičiuoja jų sutapimus, dviejose lyginamose eilutėse. Sutapimų skaičius, padalinamas iš variantų skaičiaus, skelbiamas eilučių panašumo koeficientas ir išvedamas kaip funkcijos darbo rezultatas.

Lentelė 1 „Netikslaus palyginimo metodo analizės pavyzdys“

Pirma eilutė	Antra eilutė	Sutapimai	Sutapimų skaičius	Variantų skaičius
Lyginama eilutė *test* su eilute *text* ilgiu 1.				
T	t, e, x, t	taip	3	4
E	t, e, x, t	taip		
S	t, e, x, t	ne		
T	t, e, x, t	taip		
Lyginama eilutė *text* su eilute *test* ilgiu 1.				
T	t, e, s, t	taip	3	4
E	t, e, s, t	taip		
X	t, e, s, t	ne		
T	t, e, s, t	taip		
Lyginama eilutė *test* su eilute *text* ilgiu 2.				
Te	te, ex, xt	taip	1	3
Es	te, ex, xt	ne		
St	te, ex, xt	ne		

Lyginama eilutė *text* su eilute *test* ilgiu 2.				
Te	te, es, st	taip	1	3
Ex	te, es, st	ne		
Xt	te, es, st	ne		
Lyginama eilutė *test* su eilute *text* ilgiu 3.				
Tes	tex, ext	ne	0	2
Est	tex, ext	ne		
Lyginama eilutė *text* su eilute *test* ilgiu 3.				
Tex	tes, est	ne	0	2
Ext	tes, est	ne		
Lyginama eilutė *test* su eilute *text* ilgiu 4.				
Test	text	ne	0	1
Lyginama eilutė *text* su eilute *test* ilgiu 4.				
Text	test	ne	0	1
Bendras			8	20

Pateikiama eilutė iliustruoja algoritmą paskaičiuojantį dviejų eilučių panašumo koeficientą. Eilutėms “test” ir “text”, kurių maksimalus ilgis 4, mes gavome koeficentą lygų 8/20, t.y. 0,4. Jeigu apsiriboti eilučių trumpesniais ilgiais, tai mes gautumėme kitus koeficientus: pavyzdžiui, vienetinės eilutės ilgio koeficientas būtų 6/8 arba 0,75. Eilučių ilgio didinimas didina funkcijos darbo laiką. Iš kitos pusės, paieška tampa aiškesnė. Optimalaus eilučių ilgio nėra, bet rekomenduojamas 3-5.

Tokiu būdu kiekvienam teksto žodžiui išskaičiuojamas jo sutapimo koeficientas. Šie koeficientai yra sumuojami paskui dalinami iš žodžių skaičiaus tekste ir dauginami iš 100. Taip gaunamas bendras teksto sutapimo koeficientas.

2.2.2. TAPATINGO PALYGINIMO METODAS

[3] Šis metodas skiriasi nuo Netikslaus palyginimo metodo tuo kad, analizuojant x, $|x| = m$, ir y, $|y| = n$, koeficientas k lygus analizuojamo žodžio ilgiui.

Analizė šiuo metodu atliekama tokia tvarka:

Pavyzdys x modifikuojamas taip, kad jame liktų tik žodžiai (išimami tarpai, skyrybos ženklai, skaičiai). Iš teksto y paeiliui pasirenkamas vienas žodis su tarpu ir lyginamas su kiekvienu pavyzdžio x žodžiu. Jei sutapimas aptinkamas pavyzdyje x, sutapęs žodis ištrinamas, o sutapimo koeficientas padidinamas vienetu. Pasibaigus skenavimui, sutapimo koeficientas dauginamas iš x pavyzdžio žodžių kiekio ir dalinamas iš 100 – taip gaunamas sutapimas procentais.

Kai lyginami vienodo dydžio dokumentai, palyginimas atliekamas į vieną pusę. Kai dydis skiriasi, palyginimas atliekamas į abi puses, x palyginamas su y ir atvirkščiai y palyginamas su x.

2.2.3. PRASMINIO PALYGINIMO (SEMANTINĖS PRANEŠIMŲ ANALIZĖS) METODAS

[7] Šiuo metodu tekstai analizuojami ne tiesiogiai lyginant žodžius, o lyginamos jų prasmės (prasminis turinys).

Įvairių dokumentų turinio apibendrinimas – reziumavimas (*summarizing*), bei raktinių žodžių, koncepcijų išrinkimas – visa tai yra semantinės analizės objektas.

Paprastai prasminio apdorojimo metodai taikomi tekstiniams dokumentams. Semantinė teksto analizė sudaryta iš eilės praktiškai svarbių užduočių. Viena iš jų – kontekstiškai laisva informacijos paieška: visų tekstų natūraliąja kalba radimas ir "panašių" į užduotą tekstą –pavyzdį iš tam tikro masyvo. Svarbiausia čia išgauti informaciją iš teksto ir pateikti ją formalios žinių sistemos pavidalu.

Šiame skyriuje pateikiamas tekstų analizės metodas pagrįstas semantinės analizės dėsniais

Bus vadovaujamasi semantikos nuostata, kad, jei dokumentai panašūs pagal žodyno sudėtį, greičiausiai jie panašūs ir pagal prasminį turinį.

Atliekant analizę šiuo metodu, tekstų eilutės paverčiamos į ją sudarančių žodžių prasmių rinkinius. Toliau eilutės bus lyginamos tarpusavyje (pagal principą viena su visomis). Galiausia sutapimų koeficientai bus sumuojami, bus gautas bendras sutapimo koeficientas procentais.

Metodo aprašymas:

Tekstų paruošimas analizei

Prieš analizę abiejų lyginamų tekstų žodžiai pakeičiamos jų prasmėmis. Semantinio analizatoriaus pagalba kiekvienam tekstinės eilutės žodžiui parenkamas jo atitikmuo, kuris traktuojamas kaip šio informacijos vieneto prasmė.

Analizė

Duomenys, kurie pateikiami analizei, tai tekstinės eilutės, susidedančios iš prasminių identifikatorių, ir vieno būtino kiekvienai eilutei atributo – šių identifikatorių kiekio eilutėje.

Analizuodamas palyginimo modulis įvertina ne tik žodžių sutapimus, bet ir žodžių sekų sutapimus

Pvz.:

Eilutės – „**abc def sdr vpn ert“** ir „**vbn dph vpn ert sdr“** – kaip matome iš šio pavyzdžio, sutapo eilučių dalis „**sdr vpn ert“ ir „vpn ert sdr“.** Pilnas eilutės sutapimas (žodžių, jų sekų) įvertinamas 1 balu. Eilučių sutapimo koeficientas išskaičiuojamas sekančiai:

Žodžių eilutėje yra 5 (kai žodžių skaičius eilutėse nėra lygus, imamas didžiausias).

1 / 5 = 0,2 – vieno žodžio svoris eilutėje. Kadangi sutapo 3 žodžiai 0,2 + 0,2 + 0,2 = 0,6. Toliau seka žodžių sekos sutapimai. Kaip matome iš pavyzdžio, sutapo seka iš dviejų žodžių. Tai yra 0,2 + 0,2 = 0,4. Ir eilutės sutapimo koeficientas yra (1 + ((0,6 + 0,4) / 2))/2 = 0,75.

Kai reikia iškaičiuoti ne vienos eilutės, bet viso teksto sutapimo koeficientą. Tvarka yra tokia: kiekvienos eilutės didžiausias sutapimo su kitomis eilutėmis koeficientas skaičiuojamas kaip pavaizduota aukščiau. Toliau išskaičiuojamas bendras žodžio svoris tekste – 1 dalinamas iš visų žodžių skaičiaus tekste. Eilutės sutapimo koeficientas skaičiuojamas taip – žodžių kiekis eilutėje dauginamas iš žodžio svorio tekste ir dauginamas iš didžiausio eilutės sutapimo koeficiento. Eilutės sutapimo koeficientai yra sumuojami ir taip gaunamas bendras teksto sutapimo koeficientas.

3. PROJEKTINĖ DALIS

Pagrindiniai kuriamos programinės įrangos tikslai yra:

- Automatiškai sulyginti pateiktus elektroninius dokumentus, siekiant nustatyti jų tarpusavio panašumą.
- Programoje turi būti įdiegtos trys skirtingos palyginimo technologijos.
- Kaip papildomas funkcionalumas programoje turi būti realizuota paieškos dokumento tekste galimybė.

3.1. PROJEKTO APRIBOJIMAI

Tai apribojimai, kurie įtakoja reikalavimų specifikaciją ir sistemos kūrimo eigą bei charakteristikas.

3.1.1. APRIBOJIMAI SPRENDIMUI

Pagrindiniai apribojimai kuriamai informacinei sistemai:

- Vartotojo sąsaja: Turi būti langų tipo vartotojo sąsaja. Pagrindiniai langai turi būti neperkrauti nereikalingais elementais, o suvedama informacija neturi būti perteklinė.
- Programinės įrangos patikimumas: Įvykus nenumatytiems programinės įrangos gedimams, vartotojo duomenys turi būti apsaugoti nuo neatstatomo pažeidimo.
- Programinės įrangos saugumas. Programinės įrangos saugumas turi būti aukštas, kadangi produkto bei duomenų saugumas tiesiogiai proporcingas programinės įrangos saugumui.
- Nesankcionuotas programos bei duomenų panaudojimas: Jokia speciali apsauga nebus diegiama, nes programinė įranga nėra komercinis produktas. Programa ir jos duomenim galės naudotis kiekvienas norintis.

3.1.2. DIEGIMO APLINKA

Kuriama programinė įranga skirta naudojimuisi vartotojų darbo stotyse arba kitaip personaliniuose kompiuteriuose, kuriuose įdiegta Windows šeimos operacinė sistema.

Minimalūs reikalavimai vartotojų programinei įrangai:

Windows 9x/2000 operacinė sistema;

- CPU Intel Pentium III, 600 MHz;
- RAM 64MB;
- HDD 12 GB;
- Vaizdo plokštės (32 MB);

Windows XP operacinė sistema;

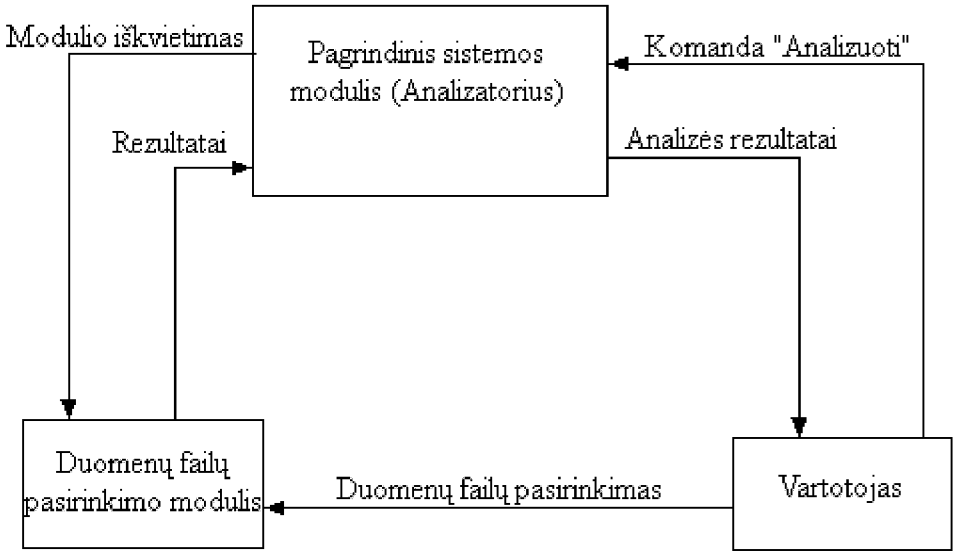
- CPU Intel Pentium 4A, 2800 MHz (5.25 x 533);
- RAM 524 MB, 400MHz DDR;
- HDD, ST340014A (40 GB, 7200 RPM, Ultra-ATA/100);
- Vaizdo plokštės – Intel(R) 82865G Graphics Controller (64 MB);
- 3.2 FUNKCINIAI REIKALAVIMAI

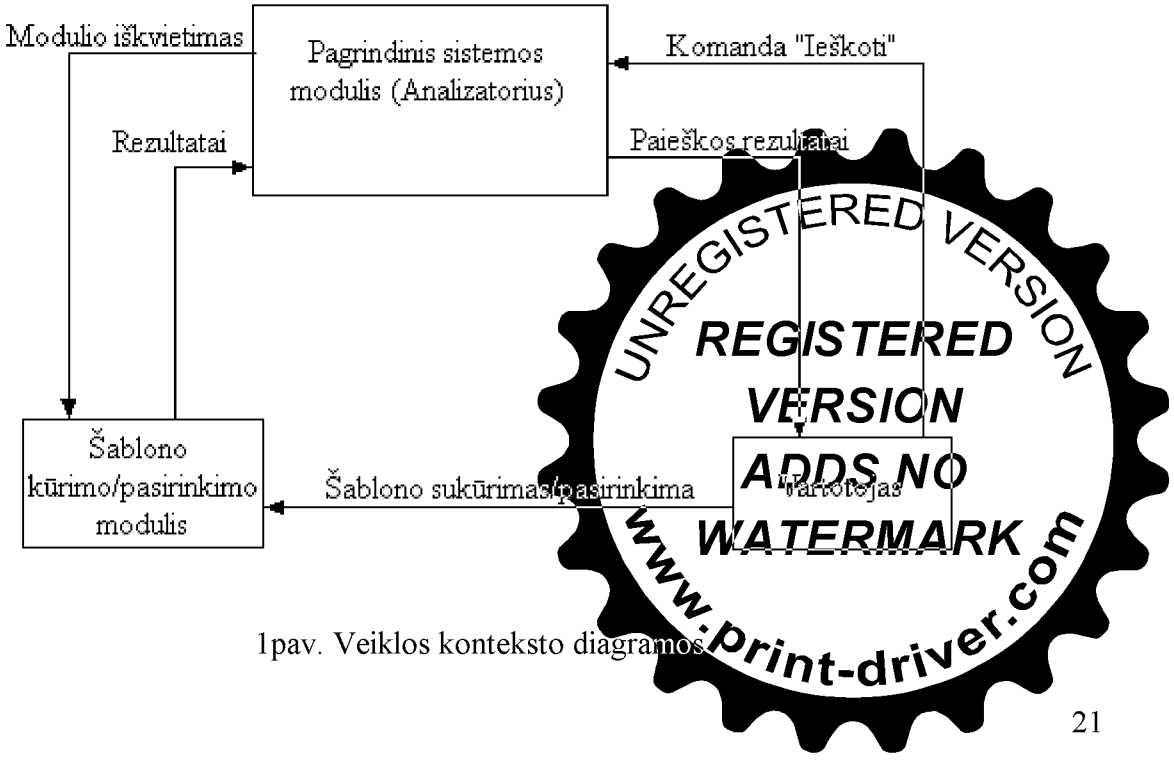
3.2. FUNKCINIAI REIKALAVIMAI

3.2.1. VEIKLOS SUDĖTIS

Veiklos kontekstas apibrėžia dominančią veiklą ir jos naudojamus bei formuojamus informacijos srautus. „Veikla“ prasideda tuomet, kai informacijos srautas įeina į sistemą ir baigiasi, kai srautas su rezultatais išeina iš sistemos.

Žemiau esančiose paveikslėliuose pateikiamos veiklos konteksto diagramos.





1pav. Veiklos konteksto diagramos

3.2.2 VEIKLOS PADALINIMAS

Šiame skyriuje pateikiamas veiklos įvykių sąrašas, kuris apima visus veiklos įvykius, už kuriuos yra atsakinga nagrinėjama veikla. Veiklos įvykiai – tai vartotojo išskiriami veiksmai, atliekami veiklos metu.

Įvykių sąrašą sudaro:

- įvykio pavadinimas;
- įeinantys ir išeinantys informacijos srautai, kurie "lydi" įvykį.

Veiklos padalinimo paskirtis – identifikuoti veiklos ,,dalyvius", kurių pagrindu būtų galima nustatyti reikalavimus.

Lentelė 2 „Analizės įvykiai“

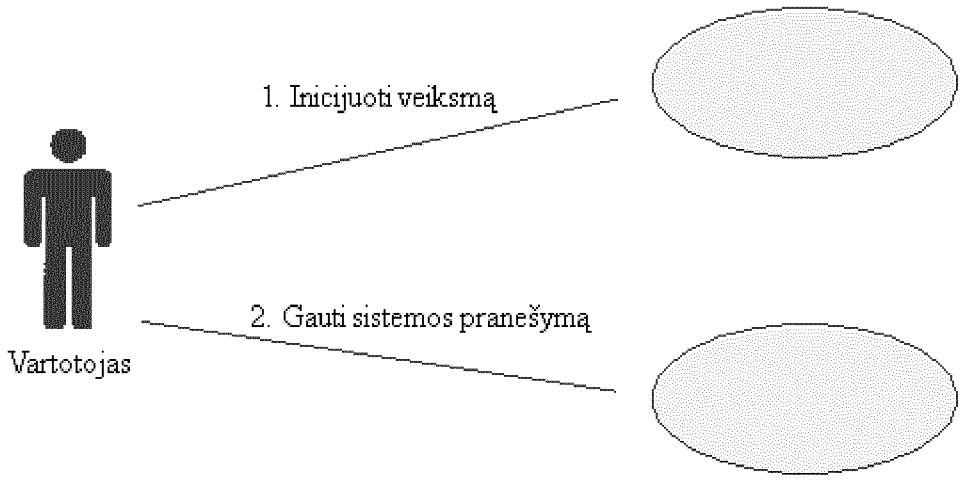
Eil. Nr	Įvykio pavadinimas	Įeinantys/išeinantys informacijos srautai
1.	Vartotojas inicijuoja analizės pradžią	Komanda „Analizuoti“ (In)
2.	Analizės modulis iškviečia duomenų failų pasirinkimo modulį	Modulio iškvietimas (In)
3.	Vartotojas pasirenka duomenų failus	Duomenų failų pasirinkimas (Out)
4	Duomenų failų pasirinkimo modulis perduoda vartotojo pasirinktus failus į analizės modulį	Rezultatai (In)
5.	Vartotojas gauna analizės rezultatus	Analizės rezultatai (Out)

Lentelė 3 „Paieškos įvykiai“

Eil. Nr	Įvykio pavadinimas	Įeinantys/išeinantys informacijos srautai
1.	Vartotojas inicijuoja paieškos pradžią	Komanda „Ieškoti“ (In)
2.	Analizės modulis iškviečia šablonų kūrimo/pasirinkimo modulį	Modulio iškvietimas (In)
3.	Vartotojas pasirenka, arba sukuria ir pasirenka šablonus paieškai	Šablonų sukūrimas/pasirinkimas (Out)
4	Šablonų kūrimo/pasirinkimo modulis perduoda vartotojo pasirinktus šablonus į analizės modulį	Rezultatai (In)
5.	Vartotojas gauna paieškos rezultatus	Paieškos rezultatai (Out)

3.2.3 SISTEMOS RIBOS

Ribas tarp sistemos ir vartotojo nusako panaudojimo atvejų diagrama.



2 pav. Sistemos ribos

3.2.4. PANAUDOJIMO ATVEJŲ SĄRAŠAS

Lentelė 4 „1 panaudojimo atvejis“

1. PANAUDOJIMO ATVEJIS:	Inicijuoti veiksmą.
Vartotojas/Aktorius:	Sistemos vartotojas.
Aprašas:	Užduoda vykdymo komandą sistemai
Prieš sąlygą:	Komandos įvykdymas programa
Sužadinimo sąlyga:	Vartotojas nuspaudžia komandos įvykdymo mygtuką
Po sąlygos:	Prasideda programos veikimas

Lentelė 5 „2 panaudojimo atvejis“

2. PANAUDOJIMO ATVEJIS:	Gauti sistemos pranešimą
Vartotojas/Aktorius:	Sistema.
Aprašas:	Išveda vartotojui pranešimą su analizės arba paieškos rezultatais
Prieš sąlygą:	Vartotojas inicijuoja veiksmą
Sužadinimo sąlyga:	Analizės arba paieškos ciklo pabaiga
Po sąlygos:	Programa nustoja veikti

3.2.5. FUNKCINIAI REIKALAVIMAI

Funkciniai reikalavimai – tai sistemos numatomų veiksmų aprašas.

1. Sistema turi leisti vartotojui inicijuoti norimą veiksmą.
2 Sistema turi leisti vartotojui pasirinkti duomenų failus iš vartotojo darbo stoties (personalinio kompiuterio) saugyklos (HDD).
3 Sistema turi įvykdyti automatinį dokumentų palyginimą pagal Netikslaus palyginimo metodą, po vartotojo pasirinkimo.
4 Sistema turi įvykdyti automatinį dokumentų palyginimą pagal Tapatingo palyginimo metodą, po vartotojo pasirinkimo.
5 Sistema turi įvykdyti automatinį dokumentų palyginimą pagal Prasminio palyginimo (semantinės analizės) metodą, po vartotojo pasirinkimo.
6 Sistema turi įvykdyti paiešką pagal šabloną dokumento tekste, po vartotojo reikalavimo.
7 Sistema turi leisti vartotojui kurti, redaguoti bei šalinti paieškos šablonus.
8 Sistema turi leisti po paieškos inicijavimo pasirinkti šabloną.
9 Sistema po analizės pabaigos turi išvesti vartotojui informacinį pranešimą – kuriame pateikiami analizės rezultatai.
10 Sistema po paieškos ciklo pabaigos turi išvesti vartotojui pranešimą su paieškos rezultatais bei pažymėti surastus žodžius dokumento tekste raudona spalva.

Lentelė 6 „1 reikalavimas“

Reikalavimas#:	**1**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA1**
Aprašymas:	Sistema turi leisti vartotojui inicijuoti norimą veiksmą				
Pagrindimas:	Sistema neturi atlikti jokių veiksmų tol, kol veiksmo neinicijuoja vartotojas				
Šaltinis:	Užsakovas				
Tikimo kriterijus:	Sistema pradeda veikti tik po veiksmo inicijavimo				
Užsakovo tenkinimas:	5			*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra			*Konfliktai:*	Nėra
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

Lentelė 7 „2 reikalavimas“

Reikalavimas#:	**2**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA1**
Aprašymas:	Sistema turi leisti vartotojui pasirinkti duomenų failus iš vartotojo darbo stoties (personalinio kompiuterio) saugyklos (HDD).				
Pagrindimas:	Sistema analizuoja pateiktus duomenų failus ir nustato jų panašumą. Duomenų failus sistemai pateikia vartotojas per specialų failų pasirinkimo modulį.				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Failų pasirinkimo modulis pasileidžia, atvaizduoja vartotojui jo darbo stoties (personalinio kompiuterio) duomenų saugyklą (HDD) bei leidžia pasirinkti norimus failus.				
Užsakovo	5			*Užsakovo netenkinimas:*	5

tenkinimas:			
Priklausomybės:	Nėra	*Konfliktai:*	Nėra
Papildoma medžiaga:			
Istorija:	Užregistruotas 2005.04.21		

Lentelė 8 „3 reikalavimas“

Reikalavimas#:	**3**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA2**
Aprašymas:	Sistema turi įvykdyti automatinį dokumentų palyginimą pagal Netikslaus palyginimo metodą, po vartotojo pasirinkimo.				
Pagrindimas:	Reikalinga, kadangi neatlikus analizės nebūtų galimybės įvertinti dokumentų panašumą				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Analizės ciklas prasideda ir pasibaigia, analizės rezultatai pateikiami				
Užsakovo tenkinimas:	5			*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra			*Konfliktai:*	Analizė negali būti įvykdyta, kai duomenų failai nepateikti sistemai
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

Lentelė 9 „4 reikalavimas“

Reikalavimas#:	**4**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA2**
Aprašymas:	Sistema turi įvykdyti automatinį dokumentų palyginimą pagal Tapatingo palyginimo metodą, po vartotojo pasirinkimo.				
Pagrindimas:	Reikalinga, kadangi neatlikus analizės nebūtų galimybės įvertinti dokumentų panašumą				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Analizės ciklas prasideda ir pasibaigia, analizės rezultatai pateikiami				
Užsakovo tenkinimas:	5			*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra			*Konfliktai:*	Analizė negali būti įvykdyta, kai duomenų failai nepateikti sistemai
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

Lentelė 10 „5 reikalavimas“

Reikalavimas#:	**5**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA2**
Aprašymas:	Sistema turi įvykdyti automatinį dokumentų palyginimą pagal Prasminio palyginimo (semantinės analizės) metodą, po vartotojo pasirinkimo.				
Pagrindimas:	Reikalinga, kadangi neatlikus analizės nebūtų galimybės įvertinti dokumentų panašumą				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Analizės ciklas prasideda ir pasibaigia, analizės rezultatai pateikiami				

Užsakovo tenkinimas:	5	*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra	*Konfliktai:*	Analizė negali būti įvykdyta, kai duomenų failai nepateikti sistemai
Papildoma medžiaga:			
Istorija:	Užregistruotas 2005.04.21		

Lentelė 11 „6 reikalavimas“

Reikalavimas#:	**6**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA1**
Aprašymas:	Sistema turi įvykdyti paiešką pagal šabloną dokumento tekste, vartotojui reikalaujant.				
Pagrindimas:	Sistemoje numatyta galimybė atlikti paiešką dokumento tekste. Kartais užtenka padaryti išvadą dėl dokumento panašumo, paprastai suradus jame tam tikrų žodžių arba žodžių junginių. Paieškos šablonai reikalingi tam, kad būtų galima nurodyti paieškai visus norimus duomenis.				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Visi žodžiai ar jų junginiai, užduoti šablone, turi būti surasti dokumento tekste, jei jie jame yra.				
Užsakovo tenkinimas:	5	*Užsakovo netenkinimas:*	5		
Priklausomybės:	Nėra	*Konfliktai:*	Nėra		
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

Lentelė 12 „7 reikalavimas“

Reikalavimas#:	**7**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA1**
Aprašymas:	Sistema turi leisti vartotojui kurti, redaguoti bei šalinti paieškos šablonus.				
Pagrindimas:	Kadangi sistemoje numatyta paieška pagal šablonus, sistemoje turi būti įdiegtas šablonų kūrimo bei redagavimo galimybė.				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Kai sukurtas šablonas atsiranda sistemoje, taip pat, kaip vaizdžiai matoma, kad po šablono redagavimo duomenys jame pasikeitė				
Užsakovo tenkinimas:	5			*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra			*Konfliktai:*	Nėra
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

Lentelė 13 „8 reikalavimas“

Reikalavimas#:	**8**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA1**
Aprašymas:	Sistema turi leisti po paieškos inicijavimo pasirinkti šabloną.				
Pagrindimas:	Sistema privalo atlikti paiešką tik po to, kai vartotojas pasirenka norimą paieškos šabloną.				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Pasirinktas paieškos šablonas atvaizduojamas paieškos informaciniame lange				
Užsakovo	5			*Užsakovo netenkinimas:*	5

tenkinimas:			
Priklausomybės:	Nėra	*Konfliktai:*	Nėra
Papildoma medžiaga:			
Istorija:	Užregistruotas 2005.04.21		

Lentelė 14 „9 reikalavimas“

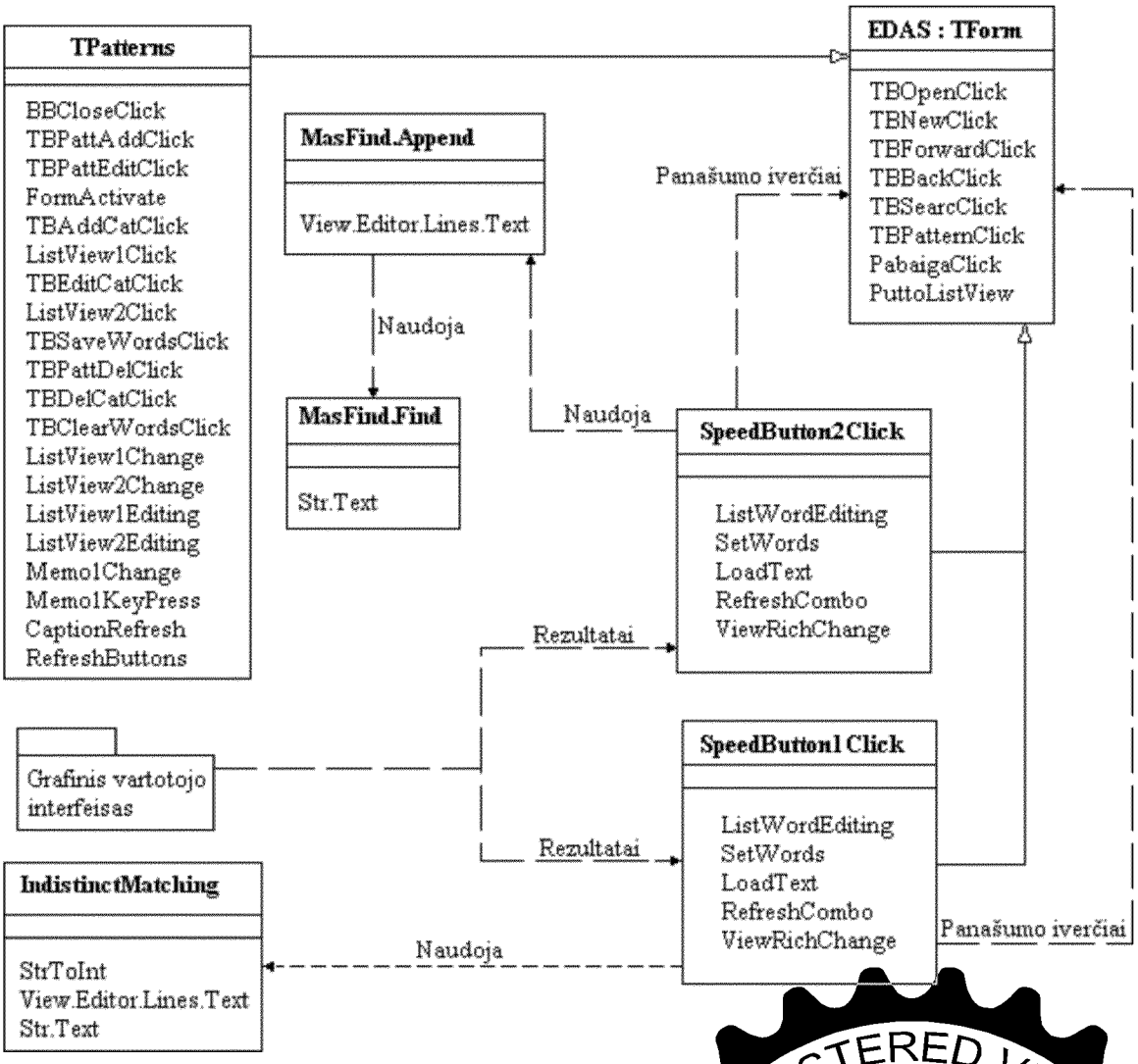
Reikalavimas#:	**9**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA2**
Aprašymas:	Sistema po analizės pabaigos turi išvesti vartotojui informacinį pranešimą, kuriame pateikiami analizės rezultatai.				
Pagrindimas:	Programos analizės rezultatai yra kaupiami jos vidinėje atmintyje. Ši atmintis vartotojui tiesiogiai neprieinama tam, kad vartotojas vaizdžiai juos pamatytų, programa turi suskurti specialų pranešimą, kuriame išveda analizės rezultatus patogioje vartotojui formoje.				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Pranešimas atvaizduojamas, rezultatai vaizdžiai matomi.				
Užsakovo tenkinimas:	5			*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra			*Konfliktai:*	Nėra
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

Lentelė 15 „10 reikalavimas“

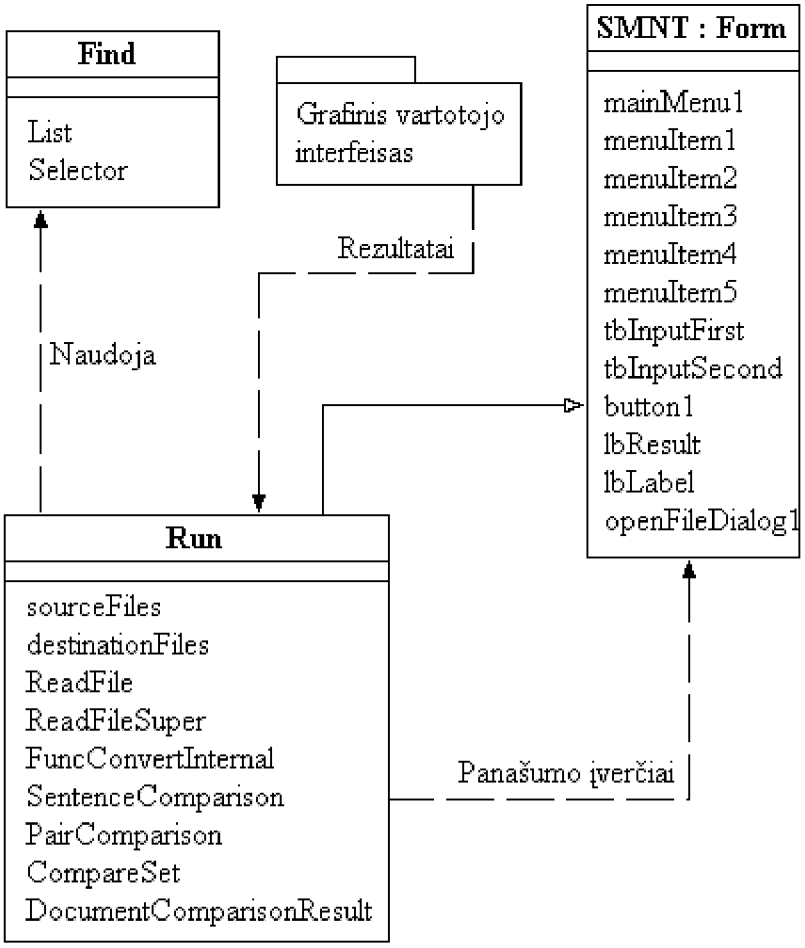
Reikalavimas#:	**10**	*Reikalavimo tipas:*	8	*Panaudojimo atvejis#:*	**PA2**
Aprašymas:	Sistema po paieškos ciklo pabaigos turi išvesti vartotojui pranešimą su paieškos rezultatais bei pažymėti raudona spalva dokumento tekste surastus žodžius.				
Pagrindimas:	Programos paieškos rezultatai yra kaupiami jos vidinėje atmintyje. Ši atmintis vartotojui tiesiogiai neprieinama, tam kad vartotojas vaizdžiai juos pamatytų, programa turi suskurti specialų pranešimą, kuriame išveda paieškos rezultatus patogioje vartotojui formoje. Surasti žodžiai dokumente pažymimi raudona spalva tam, kad juos būtų galima lengvai ir greitai surasti dokumento tekste.				
Šaltinis:	Užsakovas.				
Tikimo kriterijus:	Pranešimas atvaizduojamas, paieškos rezultatai vaizdžiai matomi. Surasti žodžiai dokumento tekste pažymėti raudona spalva.				
Užsakovo tenkinimas:	5			*Užsakovo netenkinimas:*	5
Priklausomybės:	Nėra			*Konfliktai:*	Nėra
Papildoma medžiaga:					
Istorija:	Užregistruotas 2005.04.21				

3.3 DUOMENŲ STRUKTŪRA

Tai dalykinės srities objektų specifikacija, kuri siejasi su kuriama sistema. Tai atitinka pradinį duomenų modelio variantą, kuris gali būti pateiktas ER (esybių – ryšių) diagramos arba klasių diagramos forma. Ši specifikacija kartais vadinama objektų arba srities modeliu. Žemiau pateikiamas pradinis duomenų modelis, kuris pavaizduojamas klasių diagrama.



3 pav. EDAS Duomenų struktūra

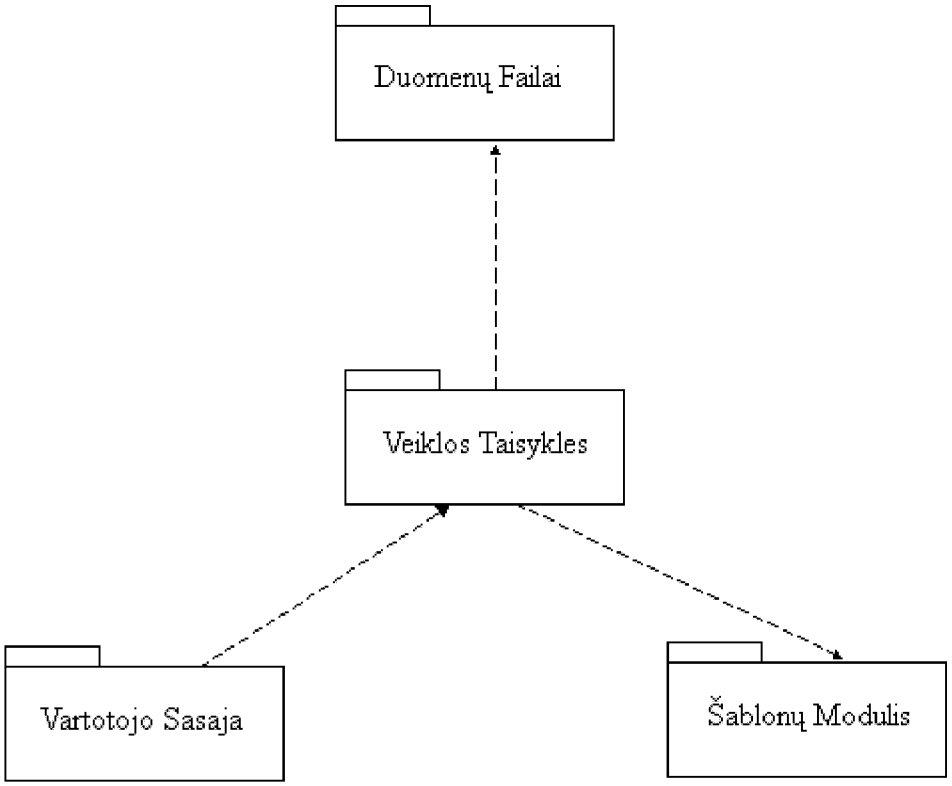


4 pav. SMNT Duomenų struktūra

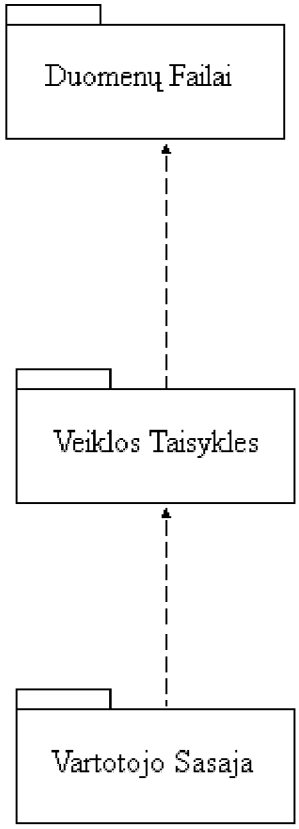
3.4. PROJEKTUOJAMOS SISTEMOS ARCHITEKTŪRA

3.4.1. SISTEMOS STATINIS VAIZDAS

Pateikiamas sistemos loginis vaizdas, kuris išskaidomas į paketus, kurie detalizuojami aprašant juos sudarančias klases.



5 pav. EDAS statinis vaizdas

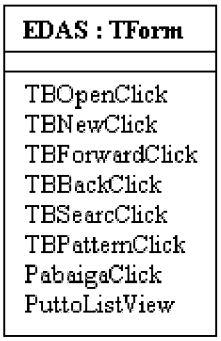


6 pav. SMNT statinis vaizdas

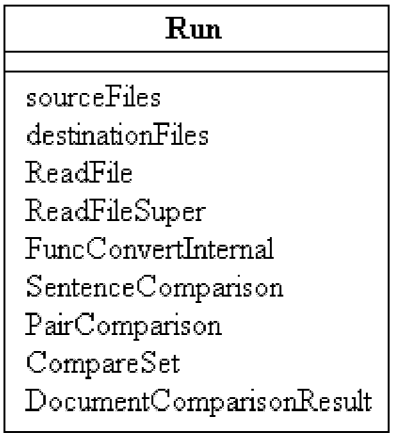
3.4.2. PAKETŲ DETALIZAVIMAS

Paketas **Veiklos Taisyklės**

Pakete pateikiamos klasės, realizuojančios visą sistemos funkcionalumą. Šis paketas naudoja paketą *Šablonų Modulis*, kai atliekama tekstinė paieška ir paketą *Duomenų Failai,* kai atliekama automatinė dokumentų palyginamoji analizė.



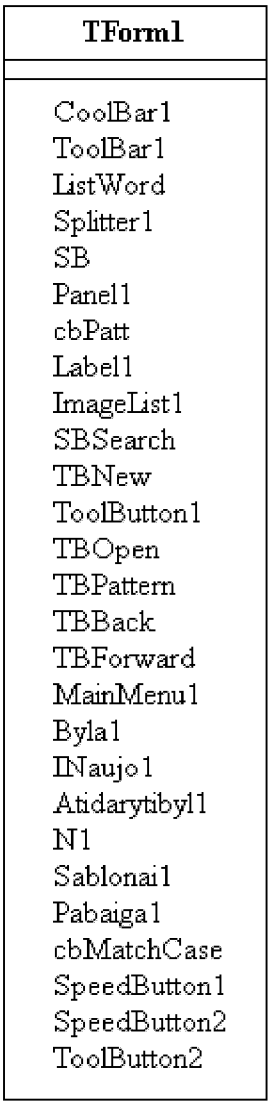
7 pav. EDAS Paketas *Veiklos Taisyklės*



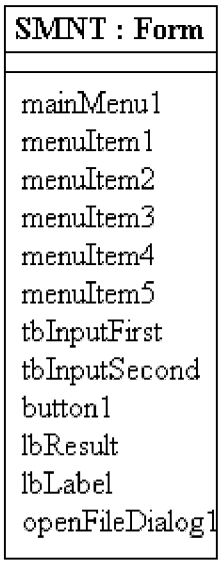
8 pav. SMNT Paketas *Veiklos Taisyklės*

Paketas **Vartotojo Sąsaja**

Paketas apima klases, skirtas vartotojui bendrauti su sistema.



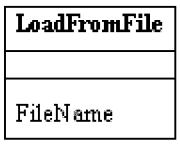
9 pav. EDAS Paketas Vartotojo Sąsaja



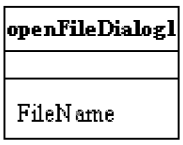
10 pav. SMNT Paketas Vartotojo Sąsaja

Paketas **Duomenų Failai**

Pakete pateikiamos klasės, skirtos duomenų failų pasirinkimui iš vartotojo darbo stoties (personalinio kompiuterio) duomenų saugyklos (HDD).



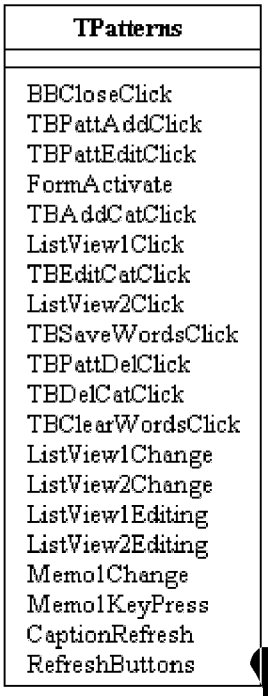
11 pav. EDAS Paketas *Duomenų Failai*



12 pav. SMNT Paketas *Duomenų Failai*

Paketas **Šablonų Modulis**

Pakete pateikiamos klasės, skirtos paieškos šablonų pasirinkimui bei šių šablonų kūrimui, redagavimui bei naikinimui.

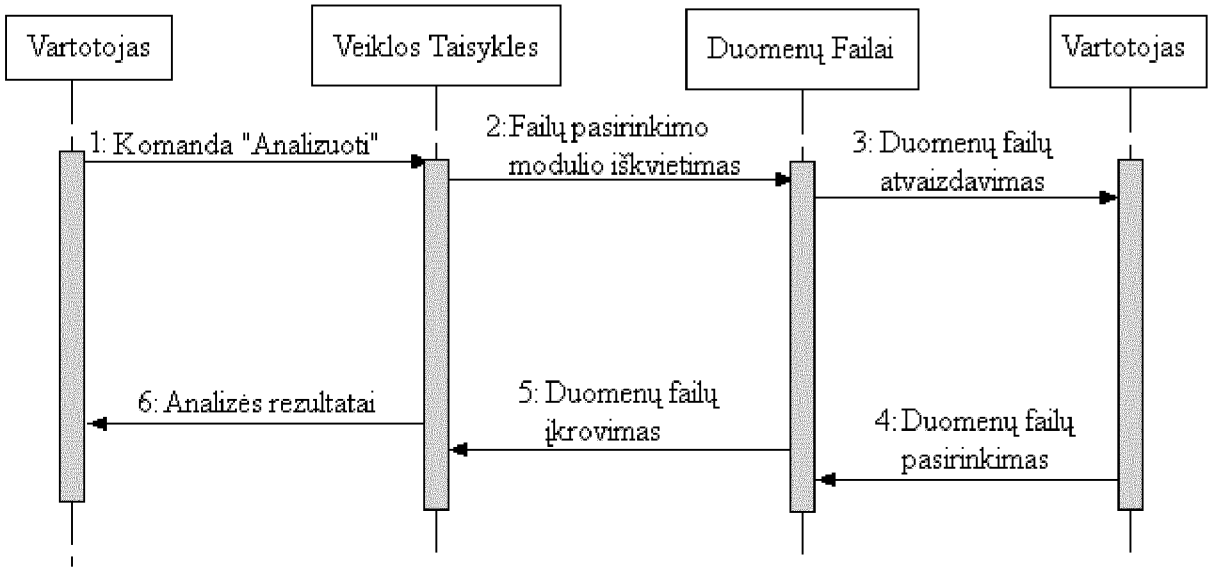


13 pav. EDAS Paketas Šablonų Modulis

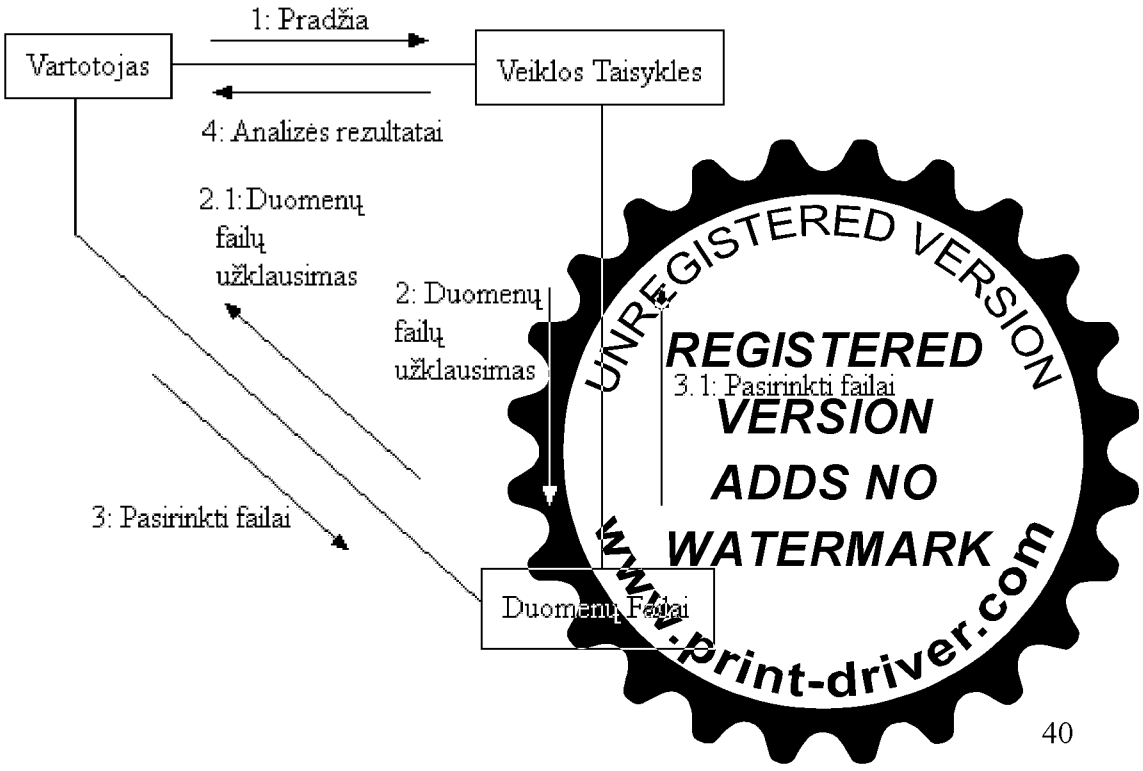
3.4.3. SISTEMOS DINAMINIS VAIZDAS

Šiame skyriuje pateikiamos sąveikos (*interaction*), būsenų (*state*) ir veiklos (*activity*) diagramos. Sąveikai atvaizduoti pasirinktos sekų (*sequence*) bei bendradarbiavimo (*collaboration*) diagramos.

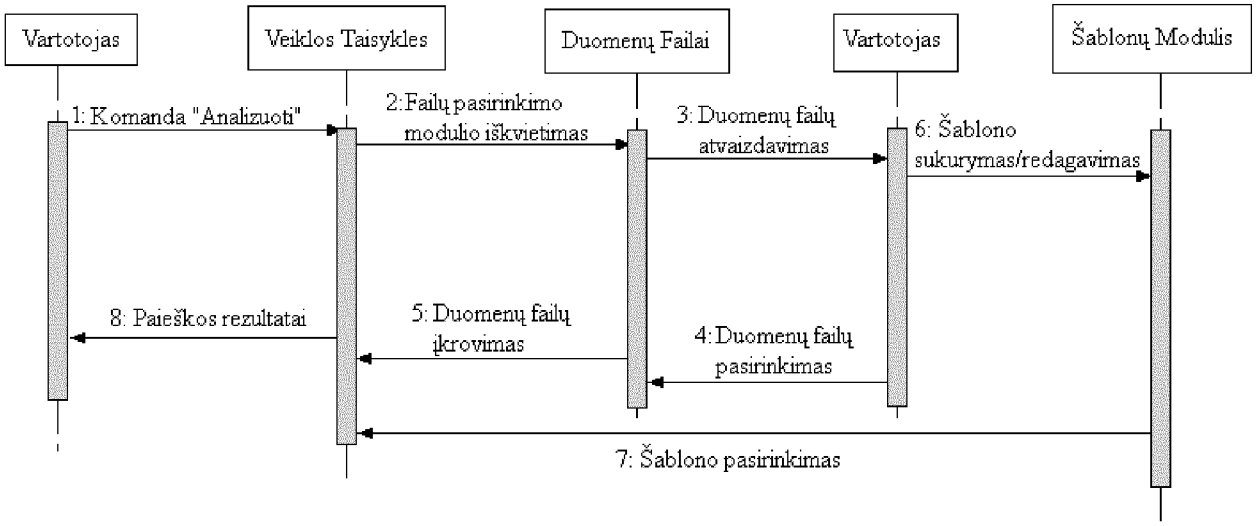
3.4.3.1. SĄVEIKOS DIAGRAMOS



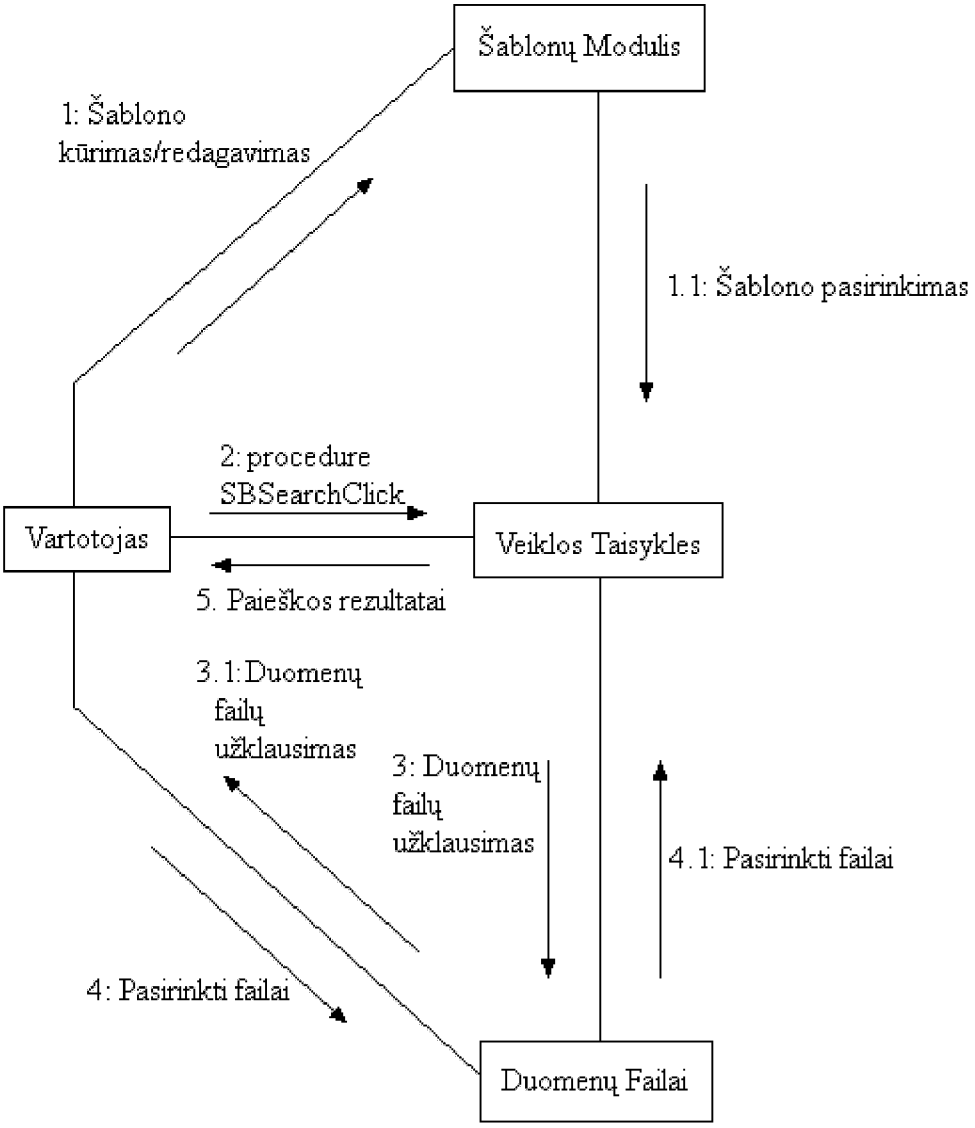
14 pav. Analizes proceso sekų diagrama



15 pav. Analizes proceso bendradarbiavimo diagrama

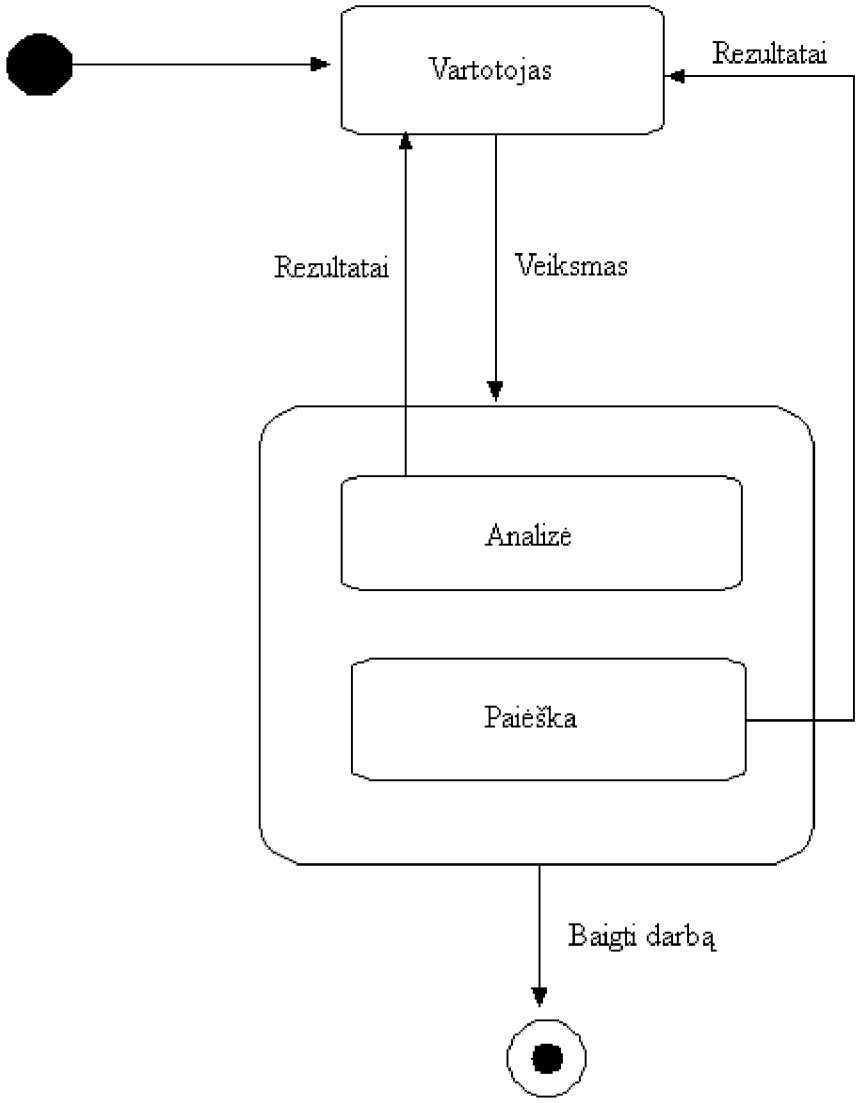


16 pav. Paieškos proceso sekų diagrama



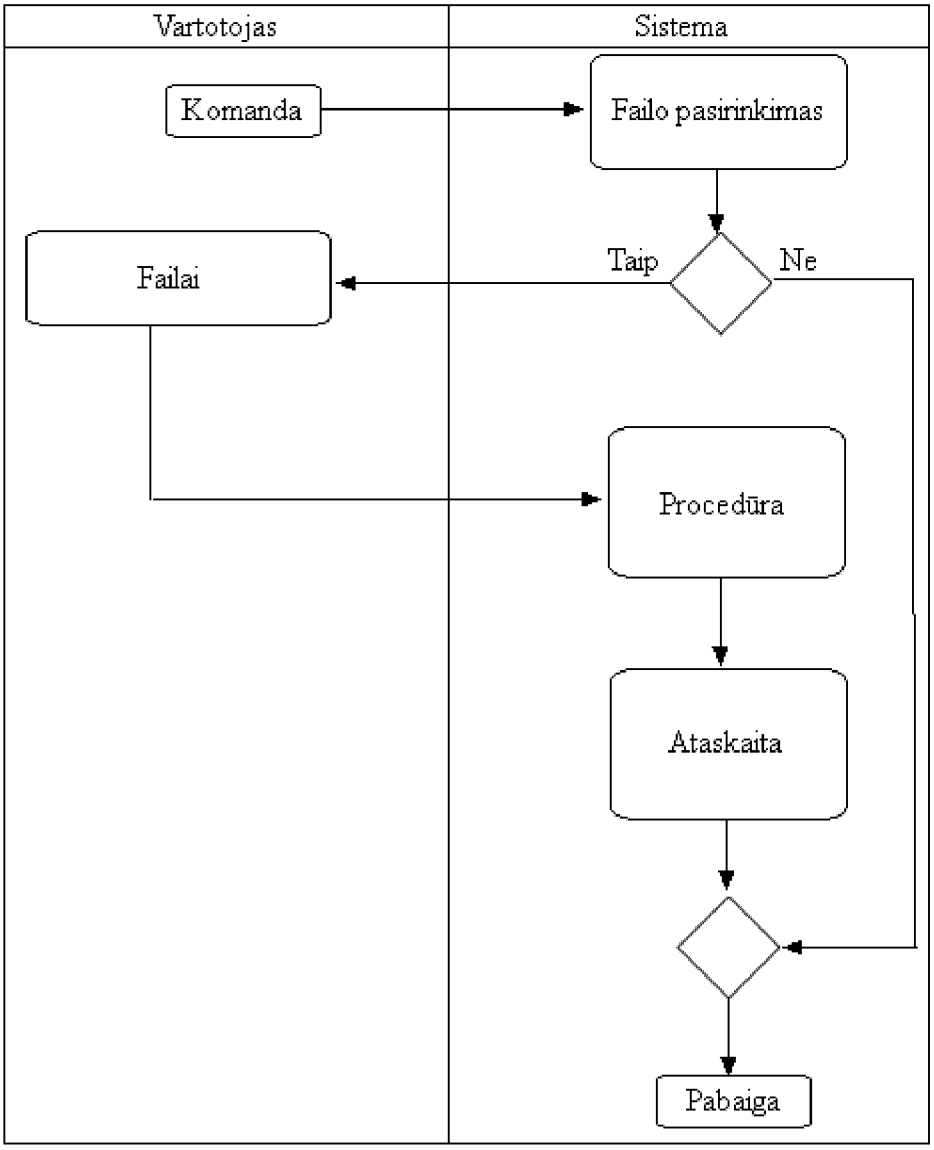
17 pav. Paieškos proceso bendradarbiavimo diagrama

3.4.3.2 BŪSENŲ DIAGRAMOS



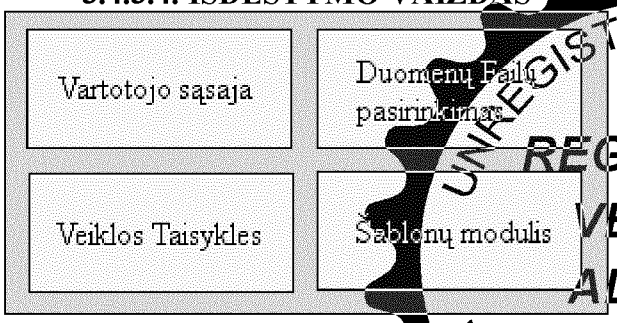
18 pav. Sistemos būsenų diagrama

3.4.3.3. VEIKLOS DIAGRAMOS



19 pav. Sistemos veiklos diagrama

3.4.3.4. IŠDĖSTYMO VAIZDAS



20 pav. Sistemos išdėstymo (*deployment*) vaizdas

3.5. NEFUNKCINIAI REIKALAVIMAI

Nusako sistemos savybes, kuriomis ji turi pasižymėti. Tai kokybinės funkciniuose reikalavimuose numatytų funkcijų vykdymo charakteristikos.

Reikalavimai sistemos išvaizdai

- lengvai skaitoma sąsaja;
- paprastas (nesudėtingas) panaudojimas;
- prieinamumas, kad vartotojas nesivaržytų naudodamas sistemą;
- programos vartotojo sąsaja turėtų būti artima standartinėms *Windows* aplinkos programų vartotojo sąsajoms (MS Word, Excel it pan.)
- neįkyri sąsaja (nereikalaujanti pastoviai ką nors kelis kartus patvirtinti);

Reikalavimai panaudojamumui

- paprastai panaudojamas bet kokio asmens be specialaus apmokymo (90% sėkmingas pasinaudojimas pirmo bandymo metu);
- nacionalinės kalbos panaudojimas;

Reikalavimai vykdymo charakteristikoms

- užduočių vykdymo greitis – ne ilgiau negu 5 min.;
- tikslumo koeficientas – paklaida tikslumui neturi būti didesnė negu 30%;
- patikimumas – pagrindiniai patikimumo reikalavimai sistemai – kuo mažesnė klaidos tikimybė.

Reikalavimai veikimo sąlygoms

Sistemos veikimui reikalingi – Personalinis kompiuteris su Windows operacine sistema (98/2000/XP).

Reikalavimai sistemos priežiūrai

(Prognozuojami pasikeitimai ir laiko sąnaudos jiems atlikti.)

Numatoma, kad sistemos patobulinimas/pakeitimas ateityje bus laikomas, kai iškils būtinybė padidinti sistemos patikimumą. Laiko sąnaudų kol kas nustatyti neįmanoma.

Reikalavimai saugumui

Sistema neskirta darbui su slaptais dokumentais, bet tam, kad sistema negalėtų naudotis tam tikros amžiaus grupės vartotojai, būtina apsaugoti ją slaptažodžiu.

Kultūriniai-politiniai reikalavimai

- sistemoje negalima naudoti ką nors įžeidžiančių terminų ar iliustracijų;

Teisiniai reikalavimai

Sistema turi pilnai atitikti šios dokumentacijos „*Funkciniai reikalavimai*“ punkto reikalavimus.

4. VARTOTOJO DOKUMENTACIJA

4.1. EDAS SISTEMOS FUNKCINIS APRAŠYMAS

„Elektroninių dokumentų analizės sistema“ (EDAS) skirta studentų rašto darbų tyrimui.

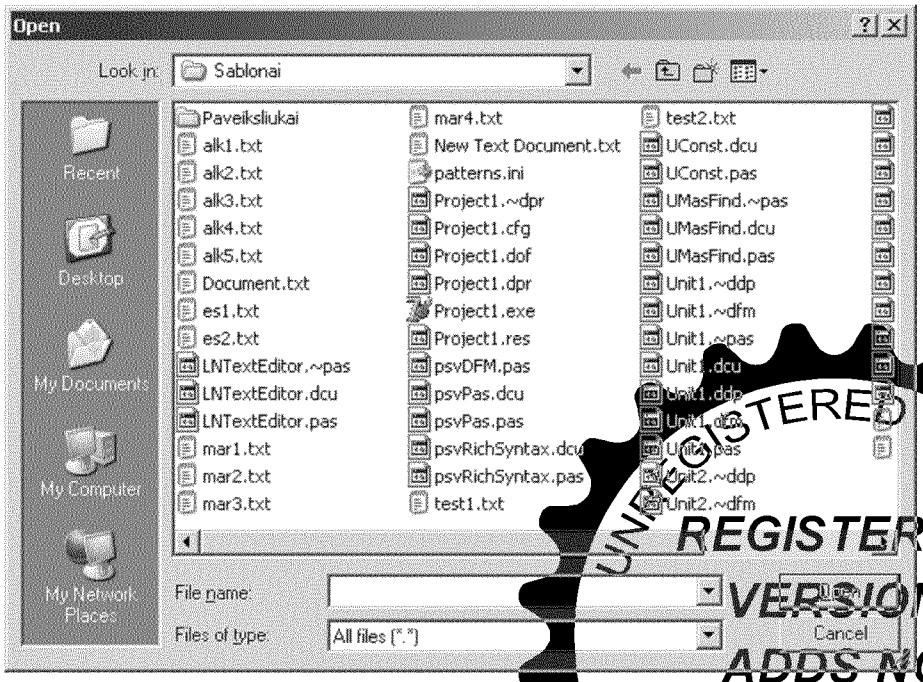
Sistema turi galimybę tarp kelių studentų rašto darbų nustatyti nuplagijuotus (tuos, kurių tarpusavio panašumas labai didelis). EDAS analizuoja pasirinktus txt failus.

4.1.1. EDAS SISTEMOS VADOVAS

Sistema galima naudotis dvejopai.

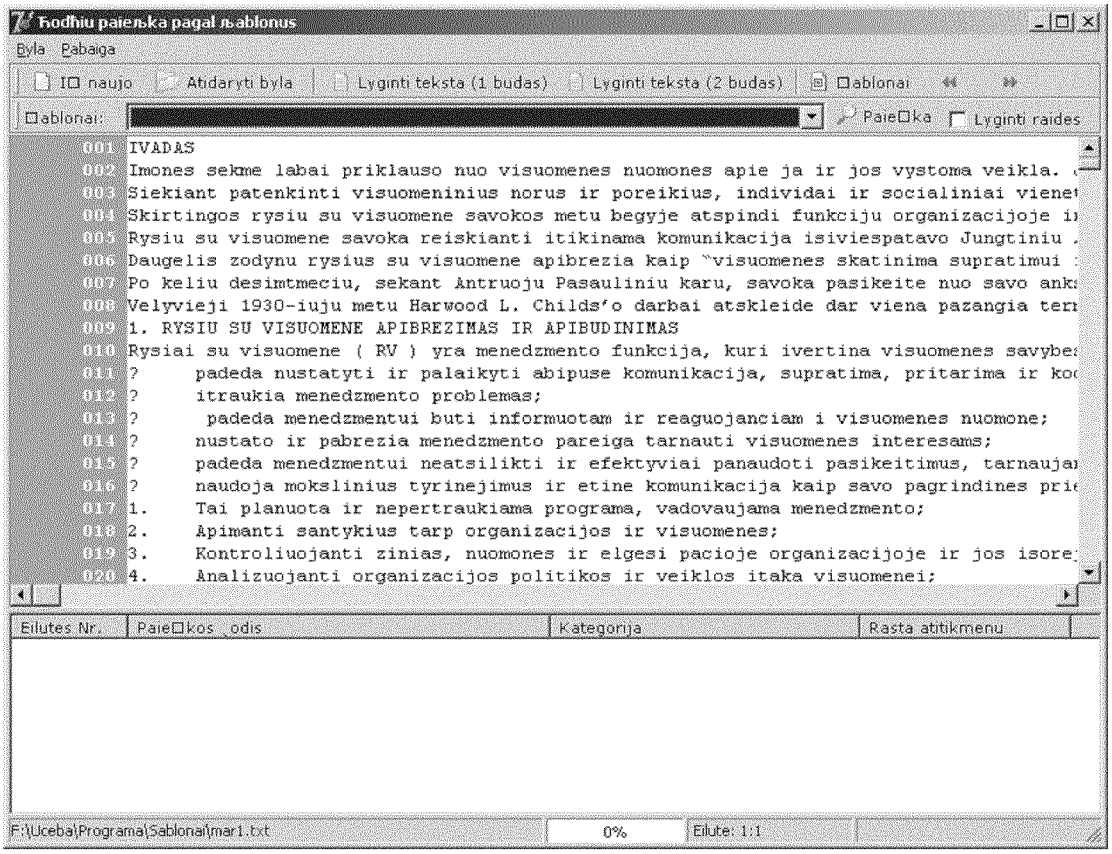
Žodžių paieška dokumentų tekstuose pagal šablonus

Šiam tikslui pradžioje pasirenkamas elektroninis dokumentas - tam pasirenkama funkcija „*Atidaryti bylą*“. Atsidarys langas, kuriame galima pasirinkti failus tyrimui (1 pav.).



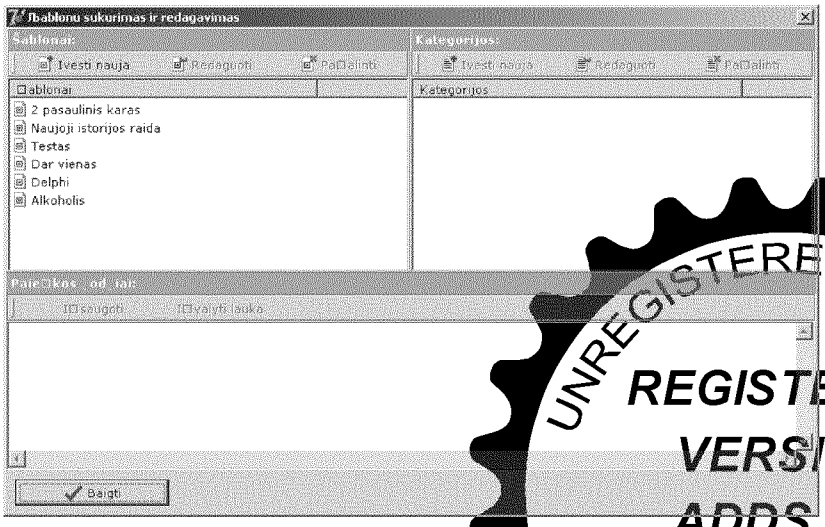
21 pav. Failų pasirinkimas

Programos langas atrodys taip:



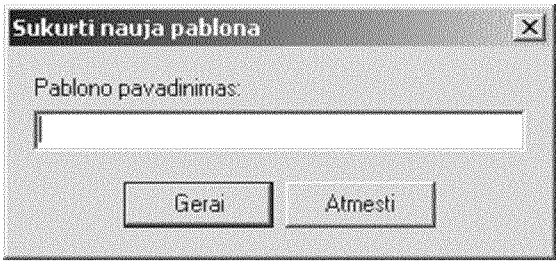
22 pav. EDAS sistemos vaizdas

Dabar, norint atlikti paiešką dokumento tekste, reikia sukurti paieškos šabloną. Tam reikia pasirinkti bylą „*Šablonai*“. Atsidarys šablonų sukūrimo langas 23 pav.



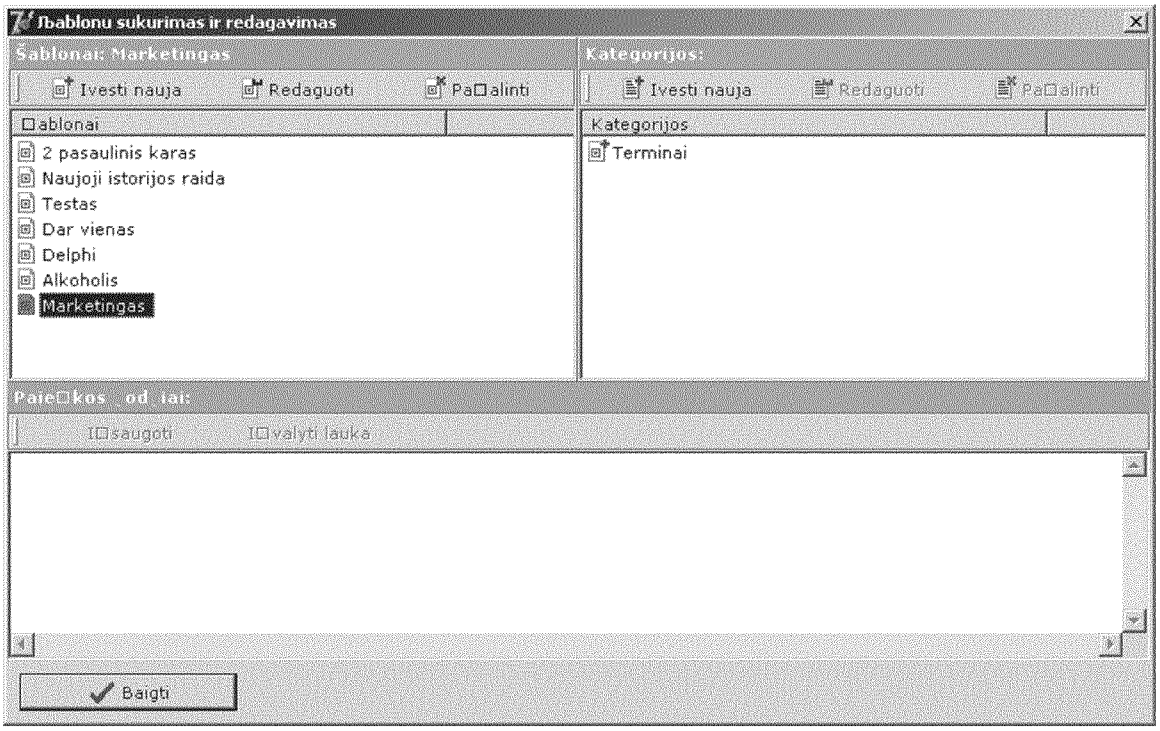
23 pav. Šablonų sukūrimo langas

Šiame lange galime pasirinkti jau sukurtus šablonus. Pažymėti šabloną ir paspausti mygtuką „*Baigti*“. Taip pat galime sukurti naują šabloną. Spaudžiamas mygtukas „*Įvesti naują*“.



24 pav. Šablono pavadinimas

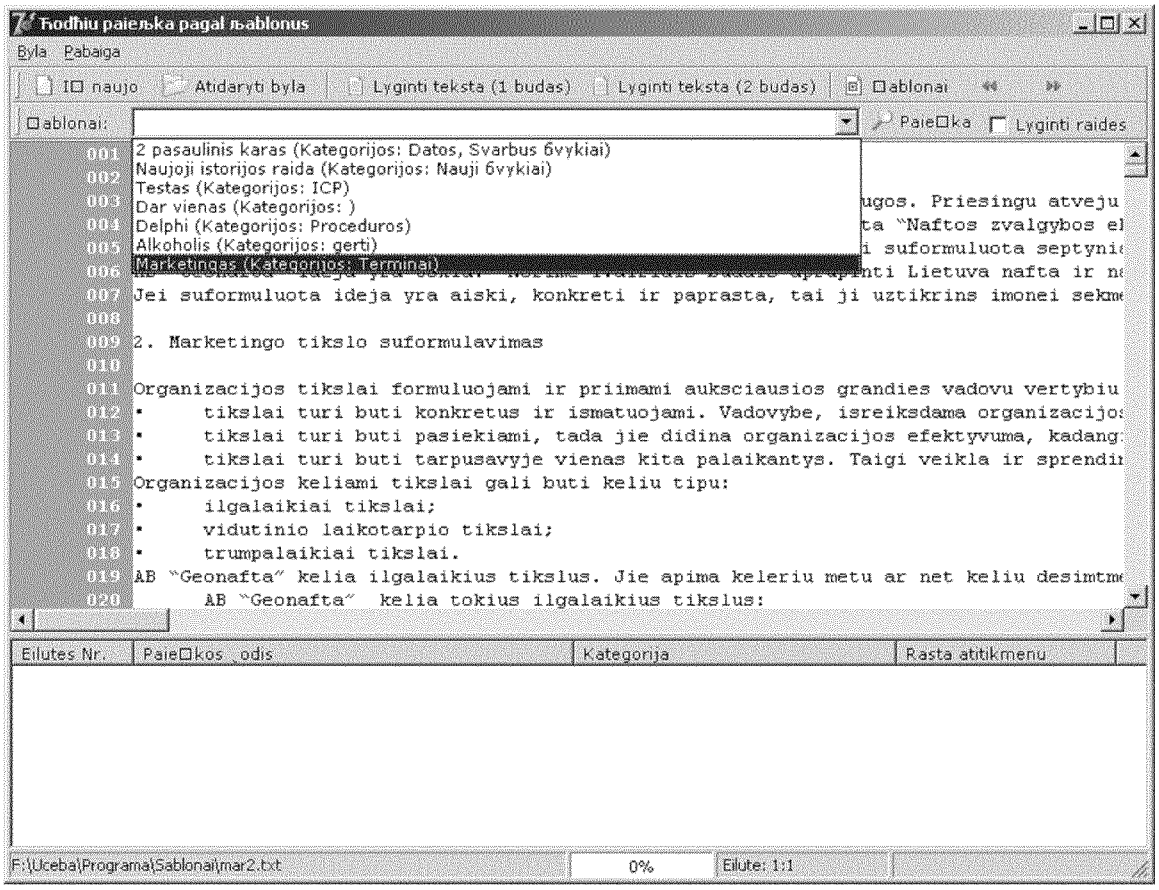
Įvedamas šablono pavadinimas ir pasirenkama „*Gerai*“ arba „*Atmesti*“.



25 pav. Šablonų pagrindinis vaizdas

Kaip matoma 25 pav., kairėje lango pusėje yra šablonų sąrašas, dešinėje šablono kategorijos. Pažymėjus kategoriją dešinėje, apatiniame lange galima surašyti paieškos žodžius. Po visų veiksmų pabaigos paspaudžiamas mygtukas „*Baigti*“, šablonas bus išsaugotas.

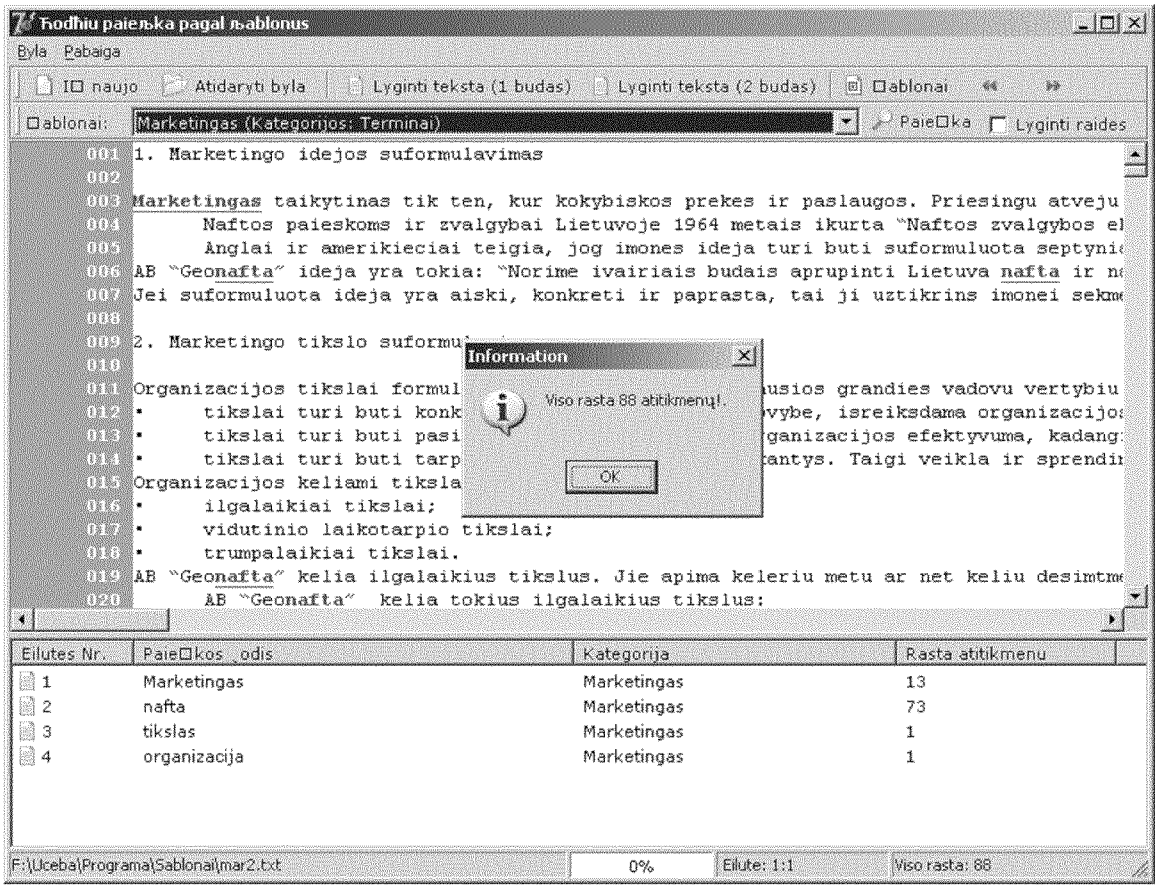
Po to kai šablonas sukurtas, jis gali būti pasirinktas per meniu virš teksto (26 pav.).



26 pav. Šablono pasirinkimas

Pasirinkus norimą šabloną spaudžiamas mygtukas „*Paieška*“

Paieškos rezultatai yra pavaizduojami apatiniame lange 27 pav.

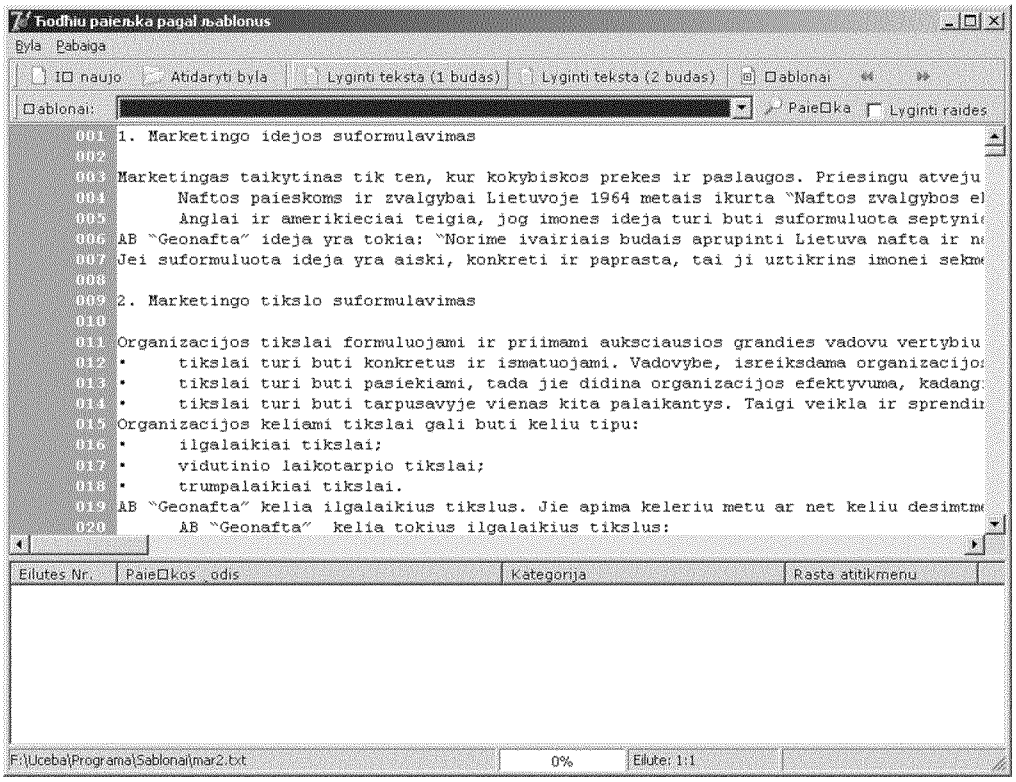


27 pav. Paieškos rezultatai

Elektroninių dokumentų tarpusavio palyginimas 2 metodais.

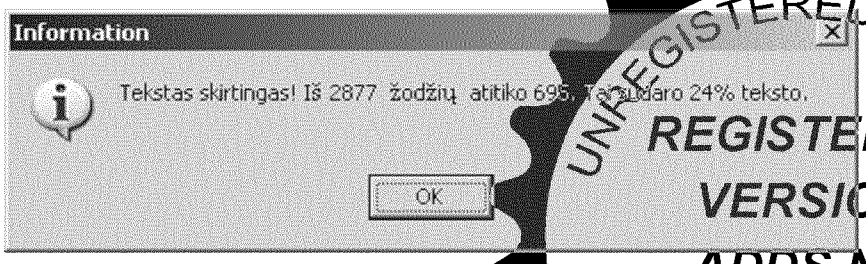
Sistema leidžia automatiškai patikrinti dokumentus 2 būdais.

Pradžioje būtina pasirinkti vieną dokumentą. Spaudžiamas mygtukas „*Atidaryti bylą*“ ir atsivėrusiame lange pasirenkamas norimas dokumentas. Kai dokumentas įkeltas į sistemą, spaudžiamas vienas iš 2 mygtukų „*Lyginti tekstą (1 būdas)“* ir „*Lyginti tekstą (2 būdas*)“ 28 pav., priklausomai nuo to koks analizės metodas bus naudojamas.



28 pav. Dokumentų tarpusavio palyginimas

Pabaigus analizę, sistema automatiškai išves į ekraną pranešimą su analizės rezultatais



29 pav. Analizės rezultatai

4.2. SMNT SISTEMOS FUNKCINIS APRAŠYMAS

„*Semantinės dokumentų analizės sistema*“ (SMNT) skirta studentų rašto darbų analizei.

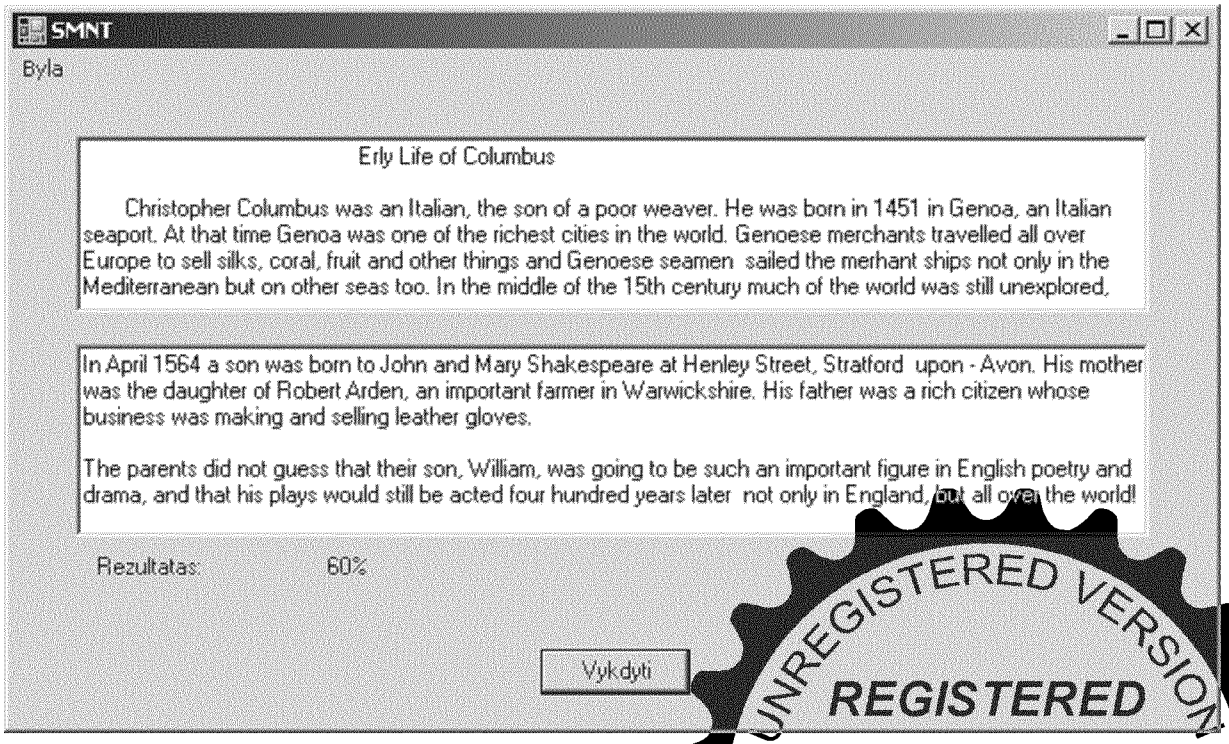
Sistema turi galimybę tarp kelių studentų rašto darbų nustatyti nuplagijuotus (tuos, kurių tarpusavio panašumas labai didelis). SMNT analizuoja pasirinktus txt failus.

4.2.1. SMNT SISTEMOS VADOVAS

Sistema leidžia automatiškai patikrinti dokumentus.

Pradžioje būtina pasirinkti dokumentus. Spaudžiamas mygtukas „*Byla*“ ir atsivėrusiame lange pasirenkama byla „*Atidaryti bylą1*“ ir „*Atidaryti bylą2*“. Kai dokumentai įkelti į sistemą spaudžiamas mygtukas „*Vykdyti*“.

Pabaigus analizę, sistema automatiškai išves į ekraną analizės rezultatą.



30 pav. SMNT sistemos vaizdas

4.3. SISTEMOS INSTALIAVIMO DOKUMENTAS

Reikalavimai vartotojų programinei įrangai:

Windows 9x/2000 operacinė sistema;

- CPU Intel Pentium III, 600 MHz;
- RAM 64MB;
- HDD 12 GB;
- Vaizdo plokštės (32 MB);

Windows XP operacinė sistema;

- CPU Intel Pentium 4A, 2800 MHz (5.25 x 533);
- RAM 524 MB, 400MHz DDR;
- HDD, ST340014A (40 GB, 7200 RPM, Ultra-ATA/100);
- Vaizdo plokštės – Intel(R) 82865G Graphics Controller (64 MB);
- 3.2 FUNKCINIAI REIKALAVIMAI

5. EKSPERIMENTINIS TYRIMAS

5.1. TESTAVIMAS

Testavimo esmė – išnagrinėti sukurtos sistemos veikimą bei pasirinktų metodų įvertinimo tikslumą. Taikoma sekanti testavimo eiga:

- Parinkti tam tikrą kiekį rašto darbų
- Rašto darbai turi būti parašyti skirtinga tematika
- Darbai bus lyginami tarpusavyje
- Rašto darbų originalai neturi būti vienodi (ta prasme neturi būti plagiatų iš anksto)
- Pirmajame testavimo etape siekiama nustatyti, kaip sistema analizuoja turinio prasme skirtingus darbus
- Antrajame etape darbai bus sąmoningai modifikuojami taip, kad aiškiai taptų plagiatu.

Testams parinkti 8 referatai:

1. American people
2. Lithuanian basketball
3. Bill Gates
4. Christmas
5. Life of Columbus
6. Four seasons of the year
7. Light
8. Lithuania

Šie referatai parinkti naudojantis nemokama, laisvai prieinama Internete referatų duomenų baze - http://mokslo.centras.lt .

5.2. TESTAVIMO EIGA

1 Etapas – Testuojama, siekiant nustatyti, kaip sistema analizuoja nepanašius dokumentus (ne plagiatas). Kadangi analizuojami skirtingi dokumentai, palyginimo rezultatas turėtu būti 0%, bet tekstuose dažnai gali pasitaikyti panašių žodžių (artikeliai, išsireiškimai ir pan.), atsižvelgiant į tai, lyginimo paklaida nustatoma 0-35% ribose.

1 – Tapatingo palyginimo metodas

2 – Netikslaus palyginimo metodas

3 – Prasminio palyginimo (Semantinės analizės) metodas

Lentelė 16 „Ne plagijavimo atvėjo analizės rezultatai“

	Four seasons of the year	American people	Lithuanian basketball	Bill Gates	Christmas	Lithuania	Life of Columbus	Light
Four seasons Of The year		1 – 17 % 2 – 66 % 3 – 20 %	1 – 17 % 2 – 66 % 3 – 20 %	1 – 17 % 2 – 62 % 3 – 21 %	1 – 23 % 2 – 66 % 3 – 20 %	1 – 18 % 2 – 61 % 3 – 19 %	1 – 24 % 2 – 66 % 3 – 30 %	1 – 31 % 2 – 66 % 3 – 30 %
American people	1 – 17 % 2 – 66 % 3 – 20 %		1 – 18 % 2 – 60 % 3 – 17 %	1 – 30 % 2 – 59 % 3 – 31 %	1 – 18 % 2 – 80 % 3 – 21 %	1 – 11 % 2 – 75 % 3 – 22 %	1 – 23 % 2 – 66 % 3 – 24 %	1 – 24 % 2 – 63 % 3 – 30 %
Lithuanian basketball	1 – 17 % 2 – 66 % 3 – 20 %	1 – 18 % 2 – 60 % 3 – 17 %		1 – 20 % 2 – 71 % 3 – 25 %	1 – 27 % 2 – 73 % 3 – 30 %	1 – 18 % 2 – 60 % 3 – 22 %	1 – 26 % 2 – 77 % 3 – 30 %	1 – 23 % 2 – 69 % 3 – 26 %
Bill Gates	1 – 17 % 2 – 62 % 3 – 21 %	1 – 30 % 2 – 59 % 3 – 31 %	1 – 20 % 2 – 71 % 3 – 25 %		1 – 23 % 2 – 85 % 3 – 18 %	1 – 17 % 2 – 70 % 3 – 30 %	1 – 27 % 2 – 78 % 3 – 30 %	1 – 22 % 2 – 60 % 3 – 37 %
Christmas	1 – 23 % 2 – 66 % 3 – 20 %	1 – 18 % 2 – 80 % 3 – 21 %	1 – 27 % 2 – 73 % 3 – 30 %	1 – 23 % 2 – 85 % 3 – 18 %		1 – 24 % 2 – 70 % 3 – 34 %	1 – 28 % 2 – 80 % 3 – 20 %	1 – 27 % 2 – 78 % 3 – 30 %
Lithuania	1 – 18 % 2 – 61 % 3 – 19 %	1 – 11 % 2 – 75 % 3 – 22 %	1 – 18 % 2 – 60 % 3 – 22 %	1 – 17 % 2 – 70 % 3 – 30 %	1 – 24 % 2 – 70 % 3 – 34 %		1 – 36 % 2 – 64 % 3 – 37 %	1 – 34 % 2 – 66 % 3 – 31 %

Life of Columbus	1 – 24 % 2 – 66 % 3 – 30 %	1 – 23 % 2 – 66 % 3 – 24 %	1 – 26 % 2 – 77 % 3 – 30 %	1 – 27 % 2 – 73 % 3 – 30 %	1 – 28 % 2 – 80 % 3 – 20 %	1 – 36 % 2 – 64 % 3 – 37 %		1 – 24 % 2 – 65 % 3 – 36 %
Light	1 – 31 % 2 – 66 % 3 – 30 %	1 – 24 % 2 – 63 % 3 – 30 %	1 – 23 % 2 – 69 % 3 – 26 %	1 – 22 % 2 – 60 % 3 – 37 %	1 – 27 % 2 – 78 % 3 – 30 %	1 – 34 % 2 – 71 % 3 – 31%	1 – 24 % 2 – 65 % 3 – 36 %	

2 Etapas – Testuojama, siekiant nustatyti, kaip sistema analizuoja panašius dokumentus (plagiatas). Kiekvienas referatas perrašomas taip, kad plagijuotų originalą. Nagrinėsime tris plagijavimo būdus: **Copy / Paste** plagiatas**, Žodžių sukeitimo** plagiatas, **Stiliaus** plagiatas

1 – Tapatingo palyginimo metodas

2 – Netikslaus palyginimo metodas

3 – Prasminio palyginimo (Semantinės analizės) metodas

Lentelė 17 „Plagijavimo atvėjo analizės rezultatai“

	Copy / Paste ~ 100%	**Žodžių sukeitimo ~ 70%**	**Stiliaus ~ 70%**
Four seasons of the year	1 – 100 % 2 – 100 % 3 – 100 %	1 – 67 % 2 – 91 % 3 – 70 %	1 – 59 % 2 – 88 % 3 – 72%
The American people	1 – 100 % 2 – 100 % 3 – 100 %	1 – 78 % 2 – 93 % 3 – 80 %	1 – 62 % 2 – 85 % 3 – 70%
Lithuanian basketball	1 – 100 % 2 – 100 % 3 – 100 %	1 – 67 % 2 – 90 % 3 – 68 %	1 – 57 % 2 – 85 % 3 – 60%
Bill Gates	1 – 100 % 2 – 100 % 3 – 100 %	1 – 62 % 2 – 89 % 3 – 75 %	1 – 46 % 2 – 83 % 3 – 73%
Christmas	1 – 100 % 2 – 100 % 3 – 100 %	1 – 59 % 2 – 88 % 3 – 72 %	1 – 58 % 2 – 85 % 3 – 71%

Lithuania	1 – 100 % 2 – 100 % 3 – 100 %	1 – 69 % 2 – 92 % 3 – 72 %	1 – 59 % 2 – 85 % 3 – 70%
Life of Columbus	1 – 100 % 2 – 100 % 3 – 100 %	1 – 76 % 2 – 94 % 3 – 84 %	1 – 56 % 2 – 85 % 3 – 76%
Light	1 – 100 % 2 – 100 % 3 – 100 %	1 – 18 % 2 – 71 % 3 – 84 %	1 – 37 % 2 – 81 % 3 – 50%

5.3. TESTAVIMO REZULTATŲ ANALIZĖ

Pradžioje bus apžvelgiama, kaip šie trys metodai įvertina nepanašius dokumentus, kitaip sakant, kaip įvertinamas ne plagiatas. Iš lentelės rezultatų matoma, kad *Netikslaus palyginimo* metodas rodo gan aukštus sutapimo koeficientus. *Tapatingo palyginimo* metodo ir *Prasminio palyginimo (Semantinės analizės)* metodo rezultatai yra mažesni. Iš to galima daryti išvadą, kad *Netikslaus palyginimo* metodas ne plagijavimo atvejį nustato klaidingai. O *Tapatingo palyginimo* metodas ir *Prasminio palyginimo (Semantinės analizės*) metodo rezultatai yra "arti tiesos". *Prasminio palyginimo (Semantinės analizės*) metodo rezultatai yra šiek tiek didesni negu *Tapatingo* metodo, bet tai galima paaiškinti tuo, kad *Tapatingo palyginimo* metodas aptinka tik absoliučiai panašius žodžius, taigi kai kuriais atvėjais panašumo jis nenustatė.

Dabar bus apžvelgimas kitas variantas – plagiato nustatymas. Kaip matoma iš lentelės rezultatų, visi trys metodai taisyklingai nustatė pirmą plagijavimo atvejį – *Copy / Paste* plagiatą.

Įvertinant kitus du atvejus, gauti rezultatai yra skirtingi. *Netikslaus palyginimo* metodo rezultatai ir *Žodžių sukeitimo* ir *Stiliaus sukeitimo* ganėtinai aukšti – galima būtų daryti išvadą apie teisingą įvertinimą, bet stebina tas faktas, kad absoliučiai visų analizuotų dokumentų sutapimo koeficientai yra virš 80%. O jei dar atsižvelgti į ankstesnius ne plagiato palyginimo rezultatus tai ir šį kartą galima padaryti išvadą apie šio metodo patikimą netikslų įvertinimą arba, kitaip sakant, ir plagijavimo atvejus šis metodas nustato klaidingai.

Tapatingo palyginimo metodas – *Žodžių sukeitimo* plagijavimo atvejį, kaip matoma iš rezultatų, nustato gana teisingai, o *Stiliaus* rezultatai jau labai žemi, galima daryti išvadą, kad šį atvejį šis metodas nustatė klaidingai.

Galiausiai išanalizavus *Prasminio palyginimo (Semantinės analizės)* metodu gautus rezultatus – matoma, kad ir *Žodžių sukeitimo* ir *Stiliaus* atvejai šiuo metodu yra nustatyti teisingai.

6. IŠVADOS

1. Šiame darbe buvo nagrinėjama aktuali mokslinėje sferoje problema – plagiatas studentų darbuose.
2. Atlikus egzistuojančių sistemų analizę, buvo nustatyti rašto darbų analizės programinei įrangai keliami reikalavimai bei vartotojų poreikiai.
3. Buvo nuspręsta panaudoti programoje, bei tuo *pačiu palyginti tarpusavyje 3 analizės metodus - Netikslaus palyginimo metodą, Tapatingo palyginimo* metodą ir *Prasminio palyginimo (Semantinės analizės)* metodą.
4. Buvo sukurta programinė įranga, kurioje buvo realizuotas aukščiau išvardintų metodų funkcionalumas ir atlikti bandymai, siekiant nustatyti pasirinktų metodų panaudojimo racionalumą.
5. Atlikus testavimus paaiškėjo, kad *Netikslaus palyginimo* metodas – visiškai netinka dokumentų palyginimo uždaviniams, nes šio metodo lyginimo paklaida labai didelė.
   Pvz: *Netikslaus palyginimo* metodo plagijavimo atvėjų analizės rezultatai buvo didesni negu *Prasminio palyginimo (Semantinės analizės)*, bet ir ne plagijavimo atvėjų nustatymas šiuo metodų parodė aukštus rezultatus (kitaip tariant šiuos atvėjus nustatė kaip plagiatą). O tai visiškai netenkina užduoties sąlygų.
   Palyginimas *Tapatingo palyginimo* metodu parodė, kad šis metodas žymiai kokybiškiau atliko analizę. Tačiau jis irgi negali būti priimtas, kaip tinkamas variantas, nes kai kurių plagijavimo atvejų jis nesugeba nustatyti.
   Tuo tarpu *Prasminio palyginimo (Semantinės analizės)* metodo panaudojimas pasiteisino, nes visus plagijavimo atvejus jis nustatė sėkmingai.

7. LITERATŪRA

[1] Paul Clough. Plagiarism in natural and programming languages: an overview of current tools and technologies. 2000

[2] А. Б. Бушев ПЛАГИАТ И КОПИРАЙТ В ЭЛЕКТРОННУЮ ЭПОХУ. 2000

[3] Graham A. Stephen String analysis. October 1992

[4] P. Hall, G. Dowling. Approximate String Matching.

[5] T. R. Girill, Clement H. Fuzzy Matching as a Retrieval-Enabling Technique for Digital Libraries.

[6] В. Максимов Алгоритмы поиска, или «Как искать неизвестно что.». Монитор № 6, 95.

[7] Автоматическая Обработка Текста
Prieiga per internetą: http://www.aot.ru/product.html#top

[8] Text Analysis, Text Mining, and Information Retrieval Software
Prieiga per internetą: http://www.kdnuggets.com/software/text.html

[9] L. Prechelt, G. Malpohl, M. Philippsen. JPlag: Finding plagiarisms among a set of programs. Technical report 2000-1, Fakultat fur Informatik, Universitat Karlsruhe, Germany, 2000.

[10] G. Whale. Software metrics and plagiarism detection. Journal of Systems and Software, 13, 1990.

[11] M. J. Wise. YAP3: Improved detection of similarities in computer program and other texts. SIGCSE Bulletin, vol. 28, 1996.

8. SUMMARY

Detection of plagiarism. Application of semantic message analysis method.

With the Internet being widely spread, databases are becoming more and more accessible. Unfortunately, this process brings some problems alongside obvious advantages. Easily accessible information causes raise of plagiarism, a phenomena of people finding some creative works, slightly (if at all) changing them and presenting as their own pieces.

This work presents an overview of various methods that can be used in computer software in order to automatically perform document analysis and comparison for detecting similarity instances. In the work, such software development stages are given as well as software testing and considerations on semantic message analysis method integration in the system.

9. TERMINŲ IR SANTRUMPŲ ŽODYNAS

Lentelė 18 „Terminai ir santrumpos“

Pavadinimas	Paaiškinimas
HDD	Hard Disk Drive – Personalinio kompiuterio duomenų laikmena.
EDAS	Elektroninių Dokumentų Analizės Sistema
SMNT	Semantinės dokumentų analizės sistema

10. PRIEDAI

Analizės dokumentų ištraukos:

Four seasons of the year
The twelve months of the year are divided into four seasons.
March, April and May are spring months. March is the month when the ground starts to thaw and the snow melts again. At the end of this month you can see many violets and anemones in the forests with some butterflies and bees on them spring is my mother's favorite season. She says: "It's exiting when you wake up and that the nature is different every morning. At first you see only the grey ground, then you see the buds on the trees, some days later everything blossoms and finally you can see that all nature is green." The weather is changeable. The cold days of winter turn to the chilly days of spring. Usually there are some thunderstorms in April. But during these months farmers begin to work in their farms; they work from the early spring until the late autumn. June, July and April are summer months. The weather is changeable: some days are hot, some - quite cool, some days are dry, during some days it's pouring with rain. Summer is a time of holiday. Then people go near the water, to the country. If you ask in the street, what the summer is, most of people would answer: “summer is a period of entertainment!” And it’s the truth. If you’d go to Palanga, Vilnius, Kaunas or some other big town and you can spend your free time very well there.

Žodžių sukeitimo plagiatas
The year have twelve months, they divided into four seasons.
Spring months are March, April, May. In March the ground starts to thaw also the snow melts in other words is a months of the “beginning”. This month ends with a many violets, butterflies and bees in the forests. Spring is a favorite season for my mother. She always says: "How exiting when you wake up every morning and the nature is always different. The grey ground at first, then you see the buds on the trees, and later everything is blossoms and at the end you can see the “Green nature”." The weather is very changeable. Winter cold day’s turns in to the chilly days of spring. April is usually a months of thunderstorms. In this month farmers begin to work in their farms. Work starts in the early spring and continuous until the late autumn. Summer months are June, July, August. The weather is also very changeable. Some days are hot, others are quite cool, another day is dry, in others the rain is pouring. Summer is a time of holidays. A person goes to the sea or to the rivers or lakes in the country. Most of people in the street if you ask them, what the summer is, would answer: “summer is an entertainment!” You can spend your free time very well in Palanga, Vilnius, Kaunas or some other big town.

Stiliaus plagiatas
There are twelve months in a year, but they are divided into 4 seasons. March, April May are spring months. The grow of nature starts usually in March when the snow melts. When this month comes to an end it can see violets and anemones in the forests, butterflies, bees on them. Spring is favourite season. My mother use to said: "How wonderful when you wake up and and your can see how nature differs every day. First you see the grey ground, then buds, some more days and everything is blossoms and after all all turns into green." The weather changes. The cold days turns by the chilly days. The April comes with a thunders. These months are the time of farmers. They start works in their farms; those works continuous from the early spring until the late autumn. June, July, April - we call them months of summer. This month are not stable: some days are hot, some cool, other are dry, and some days is rainy. Summer the time for holidays. Then people use spend days near the water, went out of the town. People to the question “what

the summer is?" usually answers: "It's a period of entertainment!" So they probably right. Go to Palanga or Vilnius or Kaunas or to any other big town. You can spend your time pithy there.

The American people

Americans can be very generous. Once a girlfriend and I hitchhiked from Los Angeles to San Francisco. What I liked was that the people who picked up were very worried about us. They always wanted to help us. One couple asked us if we wanted money so that we could take a bus instead of hitchhiking. And they wanted nothing back. Later we hitched a ride to Salinas with a truck driver. He was also worried about our safety and tried over his CB radio to fix us up with a ride with another trucker from Salinas to San Francisco. When he could not, he told us, "I do not want to leave you on the street, so I will take you up myself to make sure you get there safely." And then he drove us to San Francisco and dropped us off on Market Street where we were going to stay. And he didn't want anything back. He would not let pay him. That trip was a highlight of my stay in America.

Žodžių sukeitimo plagiatas

Americans are generous. Once my girlfriend with me hitchhiked between Los Angeles and San Francisco. The people who picked up us very worried about us, I liked that very much. They would like to help us. If we wanted money so that we could take a bus instead of hitchhiking, one couple asked us. They wanted nothing for. Later we had a ride to Salinas with a truck. He also very worried about us, did we are safety and over his CB radio tried with a ride with another trucker to fix us up from Salinas to San Francisco. When he couldn't, he said us, "I will not leave you on the street, I will make sure you get there safely , I take you up myself." And he drove us to San Francisco and dropped us off where we were going to stay. And he didn't want back anything. He would not let pay him, that trip of my stay in America was a highlight.

Stiliaus plagiatas

American people sometimes can be generous. Once I travel from Los Angeles to San Francisco with a familiar woman. What is fascinating me mostly that the people who had tracked us very worried about us. They wanted to help us all the time. They said that who wants money, could take a bus. And they wanted no pay-off. Later we had travel to Salinas with a truck driver. He was also very welcome pearson, he even spoke over his CB radio to another drivers, to find somebody to ride with from Salinas to San Francisco. But hi wont't find anybody, so he told us, "I do not want to leave you on the street, I will drive your there myself." And then he drove us to San Francisco and dropped us off we were going to stay. And he did not take any money. He would not let pay him. That trip leave me only pleasant remebrance of my stay in America.

Lithuanian basketball

In 1891 American teacher James Naismith invented basketball. Year after year it became more and more popular, people played basketball in more and more countries. In 1922 Steponas Darius brought it to Lithuania. It became popular so fast, as TV after World War II.
Level of Lithuanian basketball was so high, that in 1935 Lithuanian team played in the European championship. In 1935 championship took place in Riga (Latvia). Lithuanians were debutantes, and nobody could hope, that debutantes could be among leaders, but Lithuanians were!
Lithuanians weren`t champions!
In 1939 the occurred in Kaunas. A new Sport Hall was built for this occasion, and sportsmen still play basketball in it now. Lithuanian team became the winner of this competition again.
Women have never remained in the background of men. All time they tried to be in the same level like men and they always do it. European championship in 1938 was very successful and Lithuanian sportswomen won silver medals.

Žodžių sukeitimo plagiatas

Basketball was invented by American teacher James Naismith in 1891. People had started to play basketball in more and more countries, and years after years basket became more and more popular. In 1922 bascetball were brought to Lithuania by Steponas Darius, it became popular so fast there after World War II as TV.
Lithuanian's basketball level was very high. In 1935 in the European championship Lithuanian team already played. This 1935's championship took place in Latvia, in Riga town. Because Lithuanians were debutante's nobody hoped, that debutants could be among leaders, but Lithuanian's was! Lithuanians wasn`t became champion!
Next championship occurred in Kaunas in 1939. For this occasion a new Sport's Hall was built. Sportsmen still play basket in it. Lithuanian team won this competition.
Women never remained in the background of men's, like men they had tried to be in the same level and they did it. Lithuanian sportswomen won silver medals in 1938 in European championship. It was a success.

Stiliaus plagiatas

Basketball was invented in 1891 by James Naismith. Through the times it became very popular, among people in different countries. In 1922 it reached Lithuania, the culprit of it was Steponas Darius. Basketball became popular there very soon, especially as TV after WWII.
The lithuanians reached high level so soon that in 1935 Lithuanian team played in the European championship. In 1935 on championship in Riga (Latvia), lithuanians were at first time. And thought the won't win that challenge they were among leaders, that won't bad for debutantes!
In 1939 the championship were played in Kaunas. New modern sport hall was built specially for this event, btw Sports Hall is still the main building for playing different basketball competitions. Lithuanian team became the winner of that championship.
Women also wanted to show that basketball not only mans prerogative. All time they tried to reach the same level. And they won silver medals in European championship in 1938.

Bill Gates

William H. (Bill) Gates - Chairman and chief executive officer of Microsoft. Born October 28, 1955 shortly after 9 PM. He is reported to be the richest private individual in the World topping the Forbes list of richest people for both 1996 and 1997
Soon after Gates unveiled his Windows 3.0 program in 1990, the applications software industry was crying uncle. Over 60 million copies of the Windows progam were sold, which established Microsoft's operating system as the PC software standard and left companies like Lotus and WordPerfect scrambling because they had been creating applications for IBM's system, the OS/2. Six years after the Windows launch, Microsoft dominates the word processing and spreadsheet mark.

Netscape, Oracle and Sun have publicly made thwarting Gates's "plan for world domination" a holy crusade. They accuse him of trying to leverage Microsoft's near-monopoly in desktop operating systems unfairly with the goal of dominating everything from word processing and spreadsheet applications to web browsers and content.

"Where will it stop? They'll go on to bundle in content, their Microsoft Network, financial transactions, travel services, everything. They have a game plan to monopolize every market they touch," says Gary Reback, the Silicon Valley antitrust lawyer representing Netscape.

Žodžių sukeitimo plagiatas

William (Bill) Gates are the chief executive officer and the chairman of Microsoft company. He was born in October 28, 1955. The Forbes list of most richest people for 1996 and 1997 named him The richest private individual in the

World. After Bill's Windows 3.0 program was unveiled in 1990, the industry of applications software was crying uncle. Were sold over 60 million copies of this program. Microsoft operating system's was established as the Personal Computers software standard and companies like Lotus and WordPerfect were scrambled because they had been creating applications for IBM systems. Microsoft dominated as word processing and spreadsheet mark, for 6 years after the Windows launch.
Netscape, Oracle and Sun companies have public a holy crusade against Gates " world domination plan". They accuse Microsoft company in trying to leverage they monopoly in desktop computer systems and dominate in everything from word processing and spreadsheet applications to web browsers and content.
"It must be stopped? They'll go on to bundle in content. Microsoft's Network's, financial transactions, services, everything. They plan is to monopolize every market they work on" said Gary Reback, the antitrust lawyer of Silicon Valley, the represent of Netscape.

Stiliaus plagiatas
William H. (Bill) Gates - chairman and leader of Microsoft company. Date of his birth is October 28, 1955. At this time he is richest man in the World (by Forbes list of richest people). His career start to grow after Bill propose his Windows 3.0 program in 1990, for the applications software industry it was like an hurrycane. There were sold over 60 million copies of this program. Also this Microsoft's operating system became the Personal Computer software standard and became a serious competitor to Lotus and WordPerfect because they proposed applications for IBM's system. Microsoft domination continuous six years after the Windows launch.
But soon after that Netscape, Oracle and Sun companies declare a holy crusade to Gates's "plan for world domination". They accuse him of intention to monopoly everything from word processing applications to web browsers. By the words of Gary Reback the representor of Netscape - Microsoft newer stops, it has a plan to monopolize every market they touch.

Christmas
Maybe one of most excited holidays. Everyone likes Christmas and Christmas Eve - some because they shouldn't work, some - because they get a lot of beautiful presents.
In Lithuania everyone celebrates Christmas together with their families. But it's not like that in other countries. For example, in Belgium Christmas are celebrated mostly by young people - they go to the coffees, to the towns, discos. It's like New Year in Lithuania.
Actually, a Christian feast commemorates the birth of Jesus. This feast is celebrated on the 25th of December. The name 'Christmas' came from middle English Cristemas, from Old English Cr?stes m?sse, Christ's festival.
In the Christian calendar, it ranks after EASTER, PENTECOST, and EPIPHANY in liturgical importance and was not widespread until the 4th cent. The customs of the caroling, mistletoe, and gifts at Christmas are English. Elsewhere, gifts are given at other times, as at Epiphany in Spain. Christmas cards appeared c.1846. The concept of a jolly Santa Claus was first made popular in 19th-cent. New York City. The Christmas tree was a medieval German tradition. Midnight Mass is a familiar religious observance among Roman Catholics and some Protestants.

Žodžių sukeitimo plagiatas
This holiday probably one of most excited. Christmas and Christmas Eve likes all. Because of shouldn't work, a lot of beautiful presents and etc.

Christmas celebration in Lithuania usually underway together with families. But in other countries Christmas celebrates different. In Belgium Christmas celebration are more for young people. They celebrate it in coffees, towns, discos, like New Year celebration in Lithuania. This Christian feast commemorates the Jesus birth. Christmas celebrates on the December 25. The "Christmas" name came from a Middle English Christmas. To Old English this meant "Cristes messe" - Christ festival.
Calendar of Christians ranks it on 4 places; in liturgical importance it goes after EASTER, PENTECOST, and EPIPHANY. It even was not widespread celebrated until the 4 century. Carols, mist lets, and gifts at Christmas are English tradition. But also in Spain Epiphany gifts were given at other times. Cards appeared in 1846. Jolly Santa Claus was first became popular in 19 century in New York. The Christmas tree came from a medieval German's tradition. Midnight Mass is a religious observance among Roman Catholics and Protestants.

Stiliaus plagiatas
Christmas probably one of the most desired holiday. Not a person o earth which would't like Christmas and Christmas Eve - because of not working, and gave lot of presents. In Lithuania it is usual to celebrates Christmas with families. But it differs from other countries. In other countries christmas celebration is more for young people - they go to the public places. In Lithuania this way celebrates New Year.
The meaning of Christian celebration is the birth of Jesus. Christmas celebrates on the 25th of December. The name 'Christmas' came from middle England - it means Christ's festival. Christmas celebration for Christian's, stands on 4 place after EASTER, PENTECOST, and EPIPHANY, and was not celebrated until the 4th centuty. Tradition of gifts at Christmas are came from English. But such tradition were at other times, as at Epiphany in Spain. Cards appeared in 1846. The jolly Santa Claus was firstly origin in 19th centuty in New York. The Christmas tree came into this celebration from a medieval German. Midnight Mass is a religious ritual among Catholics and Protestants.

Lithuania
Lithuania lies on the road between East Europe and West Europe, as the straightest road from Germany to Russia crosses Lithuania.
The Lithuanian-speaking territory embraced 110.000
sq. km. However, this area continued to grow narrower due to the process of Slavonization, especially in the 19th century. The calamities of the 20th century dipersed the Lithuanians all around the world. At present 80% of all Lithuanians live in the Republic of Lithuania, 20% - in other countries.
Lithuania was the last country in Europe to adopt Christianity (1387). The main confession today is Catholicism.
Lithuania is an agricultural country. Lithuania has a comparatively well-developed agriculture and food-processing industry, production of radioelectronic and electrotechnical goods and a light industry.
There are 500 km of hard surface roads in every one thousand square kilometers in Lithuania.

Žodžių sukeitimo plagiatas
On the road between East Europe and West Europe lies Lithuania. This road crosses Lithuania, as the straightest road from Germany to Russia.
The old Lithuanian-speaking territory embraced 110.000 square kilometers. This area grows narrow due to the Slavonization process. Especially in the 19 century. The Lithuanians dispersed all around the world due to the calamities of the 20 century. Now - 80% of all Lithuanians live in the Republic of Lithuania, 20% - in other countries.

Christianity adopted in Lithuania very late in 1387. Catholicism is the main confession today in Lithuania.
This country is more agricultural. There are developed agriculture and food-processing industry. Also there highly developed production of radio electronic and electro technical goods and a light industry.
In Lithuania in every one thousand square kilometers more than 500 km of hard surface roads.

Stiliaus plagiatas
Lithuania is placed on the road from Eastern Europe to Western Europe, this is the straightest road between Germany and Russia. The territory where are Lithuanians embraced 110.000 sq. km. But is still grow because of the process of slavonization, especially in 19 century. The events of the 20 century throws out the Lithuanians all over the world. Now 80% of all Lithuanians are there in a Republic, 20% - at other countries. Lithuanias are the last which adopted Christianity. That was in 1387. The main religion today in Lithuania is Catholicism. Lithuania is a country of farmers. Agriculture and food-processing industry well gathered in Lithuania, as well as light industry (we produce radioelectronic and electrotechnical goods). More than 500 km of Lithuanians territory are covered with a roads.

Life of Columbus
Christopher Columbus was an Italian, the son of a poor weaver. He was born in 1451 in Genoa, an Italian seaport. At that time Genoa was one of the richest cities in the world. Genoese merchants travelled all over Europe to sell silks, coral, fruit and other things and Genoese seamen sailed the merhant ships not only in the Mediterranean but on other seas too. In the middle of the 15th century much of the world was still unexplored, and most European countries were eager to find and lay claim to new territory and thus become rich. Consequently there was much fighting on the seas.

The Mediterranean galleyswere constantly passing in and out of the port of Genoa to load or unload cargoes. Their hardly crews had often been engaged in dangerous adventures and their fine and graceful ships were in a battered condition, and the seamen had plenty of exciting stories to tell little Christopher Columbus.

The boy helped his father to weave wool, but he did not like this work. He was intersted in the big ships which came from or left for strange and distant lands, and he liked to sit out-of-doors and whath them for hours.

Žodžių sukeitimo plagiatas
Christopher Columbus was from Italy, son of a poorest weaver. He was born in 1451 in Genoa. Genoa was the richest citie in the world, at that time. Genoese merchants travelled all over Europe to sell silks, coral, fruit and other things and Genoese seamen sailed not only in the Mediterranean but they sailed merhant ships on other seas too. Much of the world was still unexplored In the 15th century, and most European countries were eager to find new territories and lay claim there and thus become rich. There was much fighting consequently on the seas.

The Mediterranean galleys were passing constantly in and out of the port of Genoa to load or unload goods. Their crews had often been engaged in dangerous adventures and their fine and graceful ships were in a battered condition, and the seamen had plenty of exciting stories to tell little Christopher Columbus.

The boy weaved wool he helped his father, but he did not like this work. he liked to sit out-of-doors and whaith for the big ships which came from or left for strange and distant lands.

Stiliaus plagiatas

Christopher Columbus was an Italian. He was the son of a poor weaver. Italian seaport (Genoa) thats the place where he was born in 1451. Genoa were the mostly reachest town in that time. Genoe sellers travelled all over Europe to sell goods. In the middle of the 15th century the world was known wery small, and people of Europe would like to find and go to live to new territory and there become rich. So there was much fighting on the seas.

The Mediterranean ships were passing in and out of the Genoa's port to put goods. Those crews had often been in dangerous and unreal adventures, and the seamen had so much of fine stories to tell Christopher Columbus.

Cristopher helped his father to weave wool, but he doesn't like to work. The Ships from all around the world has interested him more, and he beter sit out-of-doors and whaith them for hours.

Light

Especially since the launch of HST and the unprecedented clarity of the images satellites have given us, You've all seen on the news or in books, beautiful color pictures of various sights in the cosmos. But is this the way you would see these objects if you went there? Well, to tackle that question, first we have to talk about the nature of light and color. Light is made of waves of electromagnetic radiation. We perceive different wavelengths as different colors. All solid bodies emit light: stars, rocks and people included. The temperature of the star, rock or person determines which wavelength of light will be most strongly radiated. In the constellation Orion, the upper left star is Betelgeuse (Armpit of the giant), 520 l-y distant. Betelgeuse is a supergiant star, 14,000 times brighter than our sun. and so big, if you were to put Betelgeuse in place of our sun, its surface would reach all the way out to Jupiter. Betelgeuse's color is bright red. On the other hand, another supergiant star, Rigel, with a luminosity 57,000 times that of the sun, appears whitish-blue. The reason that Betelgeuse is red and Rigel is blue is that their surface temperatures are different.Hot stars at 30,000 degrees emit a lot more blue light than red light, and so hot stars look blue or bluish-white. Cool stars at 3,000 degrees give off more red light than blue, and so these stars look red.

Žodžių sukeitimo plagiatas

Stars, rocks and people all emit light, and which wavelength of light will be most strongly radiated depends on the temperature of the star, rock or person. For example, the star Betelgeuse in the constellation Orion, Armpit of the Giant, is a supergiant star, 14,000 times brighter than our own sun.

Stiliaus plagiatas

The beautiful pictures that the space telescope has given us show spectacular color. But is the color real? First, we have to consider what light and color are. Different wavelengths of light correspond to different colors, and light is called electromagnetic radiation. The temperature of an object determines the color of light emitted, and all things, including people, emit light. In the constellation Orion, the star Betelgeuse is a huge, giant star, as big as the orbit of Jupiter. Betelgeuse is red. Another star in Orion, Rigel, is blue. The reason that they are different colors is that they each have a different surface temperature. Cold stars are at about 3,000 degrees and emit more red than blue light and very hot stars emit blue light since they have temperatures of about 30,000 degrees.