

ŠIAULIŲ UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

INFORMATIKOS KATEDRA

Vaidas Gargasas

Informatikos specialybės II kurso magistro studijų studentas

SERVERIO RINKMENŲ PAIEŠKOS IR ANALIZĖS SISTEMA

SERVERS DATA SEARCH AND ANALYSIS SYSTEM

MAGISTRO DARBAS

Darbo vadovas:

Prof. habil. dr. L. Sakalauskas

Recenzentas:

Prof.habil. dr. G. Kulvietis

Šiauliai, 2010

Tvirtinu, jog darbe pateikta medžiaga nėra plagijuota ir paruošta naudojant literatūros sąraše pateiktus informacinius šaltinius bei savo tyrimų duomenis“

Darbo autoriaus _____

(vardas, pavardė, parašas)

Darbo tikslai ir uždaviniai

Tikslas - sukurti elektroninių dokumentų paieškos sistemą, kuri padėtų surasti rinkmenas vietinio tinklo serveryje, asmeniniuose kompiuteriuose bei vietiniame tinkle sujungtuose kompiuteriuose.

Uždaviniai

- 1) analitiškai apžvelgti elektroninio dokumento savybes, ištirti elektroninio dokumento paieškos metodus;
- 2) analitiškai apžvelgti elektroninių dokumentų paieškos ir analizės sistemas, išsiaiškinti esamų sistemų privalumus ir trūkumus;
- 3) ištirti pagrindinius paieškos ir analizės metodus ir pritaikyti sukurtą metodologiją rinkmenų paieškos sistemai realizuoti.

Darbo vadovo _____

(vardas, pavardė, parašas)

Anotacija

Šio darbo tikslas – išanalizuoti pasirinktas rinkmenų paieškos sistemas bei pasirinkta optimaliausia metodika sukurti elektroninių dokumentų paieškos sistemą, kuri leistų sumažinti laiko sanaudas ieškant rinkmenų vietinio tinklo serveryje ir vietiniame tinkle sujungtuose kompiuteriuose. Darbe aprašytos elektroninių dokumentų savybės, paieškos metodikos teoriniai ir praktiniai aspektai. Atlikta pasirinktų rinkmenų paieškos sistemų lyginamoji analizė.

Šio tyrimo problema: vidutinio dydžio įmonėje ar įstaigoje sunku greitai surasti ir peržiūrėti dokumentus, naudojamus vietiniame serveryje ar tinklo kompiuteriuose. Išanalizavus elektroninių dokumentų savybes bei jų paieškos technologijas buvo sukurta rinkmenų paieškos programa, pasinaudojant "Microsoft Visual C # 2005 Express Edition" programavimo platforma. Sukurta programa "Paieška", skirta surasti rinkmenas vietinio tinklo serveryje ir tinklo kompiuterių bendrai naudojamuose kataloguose. Tai ypač naudinga organizacijose, besinaudojančiose bendrais dokumentų resursais.

Darbą sudaro įvadas, elektroninių dokumentų analizė, rinkmenų paieškos sistemų analizė, paieškos technologijų apžvalga, realizuotos paieškos sistemos aprašymas, išvados, literatūros sąrašas ir 2 priedai. Darbo dydis yra 35 puslapiai be priedų, 3 lentelės, 4 paveikslėliai ir 17 literatūros šaltinių.

Annotation

This work main purpose to analyze the usage of main desktop search systems, choose one optimal system and to create electronic documents search engine, which would reduce the time searching for files on a local network server and LAN computers. There are described document search and managemant teoretical and practical aspects. Develope essential features, accomplished comparable analysis of main document search systems.

The following problems ware determined: there are no possibilities to find quickly and modify documents which are shared over Local Area Network in medium organization. After analysis of data search technologies it was decided to create a file search application using „Microsoft Visual C# 2005 Express Edition“ programming platform. Created application „Paieška“ is a tool designed for the purpose of searching for files and folders which are shared over a Local Area Network and servers. It is particularly useful in organizations where a lot of folders are shared on individual systems.

The work consists of introduction, the analysis of data managemant usage in the document management system, technologies of document seaching, solution for document search system, findings, references and 2 annexes. The size of work is 35 pages without annexes, 3 tables, 4 pictures and 17 references.

TURINYS

<u>Ivadas</u>	7
<u>Tyrimo tikslas ir uždaviniai</u>	8
<u>1. Elektroninio dokumento samprata</u>	9
<u>1.1. Elektroninio dokumento savybės</u>	9
<u>1.2. Laikmenos ir turinio santykis</u>	11
<u>1.3. Metaduomenys</u>	12
<u>1.4. Elektroninių dokumentų valdymas</u>	13
<u>2. Rinkmenų paieškos sistemų analizė</u>	14
<u>2.1. Failų sistemos</u>	14
<u>2.2. Windows rinkmenų paieškos sistemų apžvalga</u>	15
<u>2.3. Linux/MacOS rinkmenų paieškos sistemų apžvalga</u>	16
<u>3. Paieškos technologijos ir jų problemų sprendimų Apžvalga</u>	19
<u>3.1. LAN paieškos apibūdinimas</u>	21
<u>3.2. Reikalavimų serverio rinkmenų paieškos sistemai analizė</u>	22
<u>3.3. Vartotojų ir jų poreikių analizė</u>	23
<u>3.4. Nefunkciniai reikalavimai sistemai</u>	25
<u>4. SERVERIO RINKMENŲ PAIEŠKOS IR ANALIZĖS SISTEMOS MODELIS</u>	26
<u>4.1. Techninė užduotis</u>	26
<u>4.2. Programos „Paieška“ užklausų savybės</u>	28
<u>4.3. „Paieška“ programos kodo aprašymas</u>	30
<u>4.4. Programos „Paieška“ testavimas</u>	32
<u>IŠVADOS</u>	34
<u>LITERATŪROS SĄRAŠAS</u>	35
<u>1 Priedas</u>	37
<u>2 Priedas</u>	40

ĮVADAS

Įprastas popierinių dokumentų valdymas įmonėje apsiriboja dokumentų rengimu, kaupimu, archyvuavimu ir saugojimu. Dokumentai sugrupuojami pagal veiklos grupes, susegami į bylas ar segtuvus. Segtuvai archyvuoti ir saugomi. Prireikus, dokumentai, kopijuojami, siunčiami kitoms institucijoms paštu arba faksu. Per pastarąjį dešimtmetį ypač išaugo dokumentų apimtys. Pradėjo trūkti žmogiškųjų išteklių, apdorojančių vis didesnius dokumentų kiekius. Šitie faktoriai paskatino išvystyti efektyvesnius dokumentų valdymo metodus ir sistemas. Informacinėms technologijoms vis labiau skverbiantis į technikos, medicinos, verslo, gyventojų aptarnavimo ir kitas sritis, atsirado galimybė popierinius dokumentus pakeisti elektroniniais.

Tyrimų sritis, objektas, problema

Dokumento samprata yra įvairi. Dokumentas gali būti atspausdintas popieriuje ar saugomas elektroninėje terpėje. Dokumento tipas taip pat gali būti labai įvairus: Microsoft Word rinkmenos, *.htm, *.pdf rinkmenos, video, audio rinkmenos ir t.t. Šiame darbe toliau dokumentu bus vadinama bet kokio tipo rinkmena. Vartotojas neturi būti priverstas kurti ir dirbti tik su vienu tam tikro tipo rinkmena. Visapusiška sistema turi leisti vartotojui naudotis bet kokio tipo rinkmenomis.

Pagrindinė tyrimo sritis – serverio rinkmenų paieškos ir analizės sistema.

Tyrimo objektas – serverio rinkmenų paieškos sistema.

Mokslinių darbų ir tyrimų atskirai apie paieškos sistemas arba dokumentų valdymo sistemas yra žymiai daugiau negu apie bendrą šių technologijų ir koncepcijų apjungimą bei naudojimą. Rinkmenų paieškos sistemos poreikis ir paskirtis dalinai skiriasi nuo atskiros organizacijos dokumentų valdymo sistemos poreikių ir paskirties.

Šio darbo užduotis – ištirti šiuos skirtumus, sukurti rinkmenų paieškos sistemos metodiką ir pritaikyti ją organizacijos reikmėms.

Projektuojant ir realizuojant rinkmenų paieškos sistemą, iškyla kelios esminės problemos. Organizacija, kuri operuoja dideliais dokumentų kiekiais, turi operatyviai ir efektyviai valdyti esamus ir turimus dokumentus. Taip pat, organizacijoje yra kaupiami daug įvairių, specifinių formatų dokumentų. Tai reiškia, kad prieigos prie dokumentų laikas turi būti kiek galima mažesnis. Ir paieškos sistema turi būti universali. Į visus šiuos faktorius reikia atsižvelgti projektuojant dokumentų paieškos technologiją.

Darbo aktualumas

Organizacijos tikslas – išsaugoti turimus ir naujai sukurtus dokumentus, visada žinoti, kur ir kokie dokumentai yra saugomi, operatyviai surasti saugomus dokumentus, o svarbiausia – panaudoti turimus duomenis greitesniam ir efektyvesniam sprendimų priėmimui. Tam

dokumentų valdymo sistemose naudojamas paieškos mechanizmas. Norint surasti reikiamą dokumentą ar informaciją dokumentuose, reikia žinoti, apie ką yra dokumentas ir kas jame saugoma. Tekstinio pobūdžio dokumentuose reikiamo dokumento galima ieškoti pagal dokumento turinį. Jei dokumentas ne tekstinio pobūdžio, atpažinti, kas yra saugoma dokumente, tampa labai sudėtinga. Žmogiškasis faktorius lemia tai, kad žmogus kaip sistemos vartotojas gali daryti klaidas. Jis gali įvesti klaidingą informaciją apie ieškomą dokumentą, užmiršti įvesti į sistemą parametrus ir panašiai.

Sistemos vartotojai – organizacijos darbuotojai. Tik nedidelė dirbančiųjų organizacijoje dalis turi galias žinias ir patirtį kompiuterijos srityje, todėl jiems reikalinga nesudėtinga ir jų galimybes atitinkanti rinkmenų paieškos sistema. Vartotojo sąsaja turi būti kuo paprastesnė sistemos vartotojams. Jei sistema sudėtinga ir apimanti daug uždavinių, iškyla problema, kaip suderinti sistemos efektyvumą ir paprastą vartotojo sąsają.

Tyrimo tikslas ir uždaviniai

Projekto tikslas – išanalizuoti organizacijos dokumentų struktūrą, problemas, kurios iškyla dokumentų valdyme, ir pateikti naują ir efektyvesnę būdą, padedantį surasti organizacijos dokumentus. Projekto pagrindinės probleminės sritys yra organizacijos dokumentų struktūra ir dokumentų paieška.

Esminis projektuojamos sistemos skirtumas nuo įprastų dokumentų paieškos sistemų yra tas, kad ši sistema orientuota į eilinį vartotoją ir nereikalaujanti įpatingų žinių bei įgūdžių informacinių technologijų srityje. Poreikis lengvai ir greitai rasti bet kurį dokumentą organizacijos kompiuteriuose yra labai didelis. Darbuotojai dalinasi informacija, kaupia, archyvuoja elektroninius dokumentus. Spartesnis informacijos gavimas, detalesnė bei tikslesnė dokumentų paieška padės organizacijai dirbti efektyviau, sutaupys žmogiškuosius išteklius, sumažins darbo sąnaudas.

Darbo tikslas – sukurti elektroninių dokumentų paieškos sistemą, kuri sumažintų laiko sąnaudas ieškant rinkmenų vietinio tinklo serveryje ir vietiniame tinkle sujungtuose kompiuteriuose.

Norint pasiekti šį tikslą, reikėjo išspręsti tokius uždavinius:

- analitiškai apžvelgti elektroninio dokumento savybes, ištirti elektroninio dokumento paieškos metodus;
- analitiškai apžvelgti elektroninių dokumentų paieškos ir analizės sistemas, išsiaiškinti esamų sistemų privalumus ir trūkumus;
- ištirti pagrindinius paieškos ir analizės metodus ir pritaikyti sukurtą metodologiją rinkmenų paieškos sistemai realizuoti.

1. ELEKTRONINIO DOKUMENTO SAMPRATA

1.1. Elektroninio dokumento savybės

Pastarojo meto informacinių technologijų raida sudarė sąlygas milžiniškam informacijos srautų šuoliui. Šiandien informacijos vaidmuo sparčiai didėja, informacija ir idėjos tampa savarankiška vertybe, strateginiu kiekvienos organizacijos ištekliu, veiklos pagrindu. Šalia tradicinių informacijos srautų vis didesnę reikšmę ir vietą užima elektroninė informacija. Dvidešimtojo amžiaus paskutiniuoju dešimtmečiu informacijos buvo sukurta daugiau nei per visą ankstesnę žmonijos veiklos istoriją.

Tačiau informacijos amžiaus paradoksas yra tas, kad didesnė dalis elektroninės informacijos yra sunkiau prieinama nei anksčiau sukurtoji arba neprieinama visiškai. Elektroninė informacija kuriama, valdoma, perduodama ir saugoma naudojant dinamiškas, greitai kintančias informacinių technologijų priemones. Dėl greito technologijų senėjimo informacija, sukurta jas naudojant, prarandama arba tampa sunkiai prieinama ir dėl to negalima naudoti. Informacijos elektroninėje aplinkoje atsiradimas ir sklaida neišvengiamai keičia požiūrius ir sampratas. Šiandienos daug diskusijų sukelia dokumento elektroninėje aplinkoje samprata, kuria grindžiama visų informacijos kūrėjų ir vartotojų veikla.

Akivaizdu, kad tradicinė dokumento, kaip materialaus objekto, samprata negali būti taikoma elektroniniams dokumentams. Elektroninėje aplinkoje nelieka medžiagiškumo, todėl tenka ieškoti kitų dokumento, kaip atskiro objekto, išskyrimo ir apibrėžimo kriterijų. 2001 m. Tarptautinė standartizacijos organizacija patvirtino tarptautinį dokumentų vadybos standartą ISO-15489, kuriame suformuluoti ir nustatyti pagrindiniai dokumentų sudarymo principai ir reikalavimai dokumentams, atvėrę galimybes organizacijų veikloje pradėti naudoti ir elektroninius dokumentus. Standarte **dokumentas** apibrėžiamas kaip „fiksiuota informacija ar objektas, kuris gali būti laikomas atskiru vienetu“, o **oficialūs dokumentai** apibrėžiami kaip „organizacijos ar asmens, vykdančių tesėtus įsipareigojimus ar veiklą, parengta, gauta ir kaip įrodymas išsaugota informacija“.[7]

Tai: dokumento autentiškumas, patikimumas, vientisumas ir tinkamumas naudoti. Toliau trumpai apžvelgsime kiekvienos šių savybių reikšmę.

Dokumento autentiškumą galima laikyti patvirtinimu, kad dokumentas yra tikras, įskaitant jo fizines savybes, struktūrą, turinį ir aplinką, kurioje dokumentas buvo sudarytas. Dokumento autentiškumas yra patvirtinamas įrodant, kad tai yra tas pats dokumentas, kuris buvo perduotas (pvz., saugoti) ir kad tai yra toks pat dokumentas, koks jis buvo iki perdavimo, t.y. jo patikimumas ir vientisumas nebuvo pažeisti nei perdavimo metu, nei po to. Tokiu būdu dokumento autentiškumas priklauso nuo dokumento sudarytojo, kuris turi užtikrinti dokumento

autentiškumą iki jo perdavimo, taip pat ir nuo dokumento perėmėjo (kitos įstaigos, kuriai perduotas dokumentas, ar archyvo, perėmusio dokumentą saugoti). Todėl dokumento autentiškumui užtikrinti turi būti naudojamos tokios priemonės, kuriais dokumentas apsaugomas nuo sąmoningo ar nesąmoningo pakeitimo, sugadinimo, suklastojimo, praradimo jo perdavimo ir saugojimo metu. [7]

Dokumento patikimumas turi garantuoti dokumento turinio tikrumą ir teisingumą dokumento sudarymo momentu, t.y. dokumentas patikimu laikytinas tada, kai jo forma, struktūra ir turinys yra pakankami įrodyti, jog jame fiksuota informacija dokumento sudarymo metu buvo tikra ir teisinga.

Dokumentas laikomas sudarytu tada, kai dokumente yra fiksuoti visi būtini jo elementai: data, asmenys, dalyvavę dokumento sudarymo procese (dokumento sudarytojas, rengėjas ir, jei reikia - adresatas), dokumento sudarymo pagrindas (t.y. veiksmai ar klausimai, sąlygoję dokumento sudarymą) ir nuorodos apie dokumento ryšius su kitais dokumentais toje dokumentų visumoje, kuriai priklauso šis dokumentas (pvz., dokumento registracijos numeris). [7]

Labai svarbią vietą čia užima **dokumento sudarymo procedūros** – tam tikros dokumentų sudarymo taisyklės. Šiomis taisyklėmis nustatomi veiksmai, būtini atlikti sudarant dokumentą (pvz., dokumento įforminimas, registravimas, priskyrimas konkrečiam vykdytojui), apibrėžiami konkrečių asmenų vaidmenys dokumento sudarymo procese, pavyzdžiui, kokie asmenys turi teisę sudaryti vienus ar kitus dokumentus, kas turi teisę pasirašyti dokumentus, kurie asmenys yra atsakingi už veiksmų, fiksuotų dokumente, vykdymą ir pan. Dokumento patikimumas, taip pat ir jo vientisumas, priklauso nuo šių taisyklių detalumo ir jų laikymosi kontrolės lygio, t.y. kuo daugiau nustatytos procedūros, būtinos atlikti sudarant dokumentą, tuo didesnis ir lengviau įrodomas dokumento patikimumo laipsnis.

Dokumento vientisumas reiškia, kad dokumentas yra užbaigtas ir nekeičiamas, t.y. sudarytas dokumentas turi būti apsaugotas nuo neteisėtų pakeitimų nustatant, kokie papildymai ar pastabos po dokumento parengimo gali būti atlikti ir kas įgaliotas juos atlikti.

Tinkamu naudoti laikomas toks dokumentas, kurį galima surasti, pateikti ir suprasti visu dokumento gyvavimo metu, ir kuris yra išsaugotas kartu su dokumentų tarpusavio ryšiais, paliudijančiais veiklos eigą. Šios dokumentų savybės yra būtinos visiems dokumentams, nepriklausomai nuo jų formos ir laikmenos, taigi ir elektroniniams dokumentams. Pagrindinė elektroninių dokumentų ypatybė yra ta, kad **šie dokumentai turi būti sudaryti, perduodami, naudojami ir saugojami naudojant informacinių technologijų priemones**, dėl ko atsiranda tik elektroniniams dokumentams būdingų savybių, atskiriančių juos nuo įprastų (rašytinių)

dokumentų. Išskirtini šie elektroninių dokumentų ypatumai, sąlygoti skaitmeninės aplinkos savybių, kuriais jie skiriasi nuo tradicinių (rašytinių) dokumentų:

- laikmenos ir turinio santykis;
- informacijos fiksavimo laikmenoje būdas;
- dokumento forma;
- metaduomenų santykis su dokumentu;
- dokumentų saugojimas.

1.2. Laikmenos ir turinio santykis

Dokumento laikmena paprastai yra suprantama kaip materialinė medžiaga, kurioje fiksuojama norima perduoti informacija, sudaranti dokumento turinį. Tradicinėje aplinkoje informacija yra neatsiejama nuo laikmenos (pvz., popieriuje sudarytų dokumentų turinys niekaip negali būti atskirtas nuo popieriaus, nebent sunaikinant patį dokumentą). Skaitmeninėje aplinkoje sukurtų dokumentų turinys, yra atskiriamas nuo laikmenos, kurioje dokumentas buvo sudarytas. Be to, dėl techninės ir programinės įrangos senėjimo ir trumpalaikiškumo, elektroninio dokumento laikmenos pakeitimas yra būtina ilgalaikio išsaugojimo sąlyga, nes elektroninio dokumento saugojimo terminas gali būti daug ilgesnis, nei priemonių, kuriomis dokumentas buvo sukurtas, galiojimo laikas.

Informacijos fiksavimo laikmenoje būdas. Tradicinių dokumentų turinys fiksuojamas laikmenoje (popieriuje, fotojuostoje ir pan.), naudojant simbolių (raidžių, skaičių ir pan.) reikšmes, kuriuos žmogus gali perskaityti ir suprasti iš karto, nenaudojant jokių papildomų priemonių. Tokiu būdu užtikrinama, kad dokumentą gali perskaityti visi, kurie atpažįsta simbolius. Elektroninių dokumentų turinys laikmenoje koduojamas (informacija įrašoma 0 ir 1 sekomis, t.y. skaitmeninama), todėl dokumentas negali būti perskaitomas ar atpažįstamas tiesiogiai, nenaudojant specialių priemonių. Kiekvieną kartą naudojant dokumentą, koduota informacija transformuojama iš savo pirminės – dvejetainės formos - į žmogui atpažįstamą ir suprantamą. Toks informacijos fiksavimo būdas sąlygoja dokumento priklausomumą nuo informacinių technologijų priemonių (techninės ir programinės įrangos), kuriomis jis yra sudaromas, naudojamas ir saugojamas visu dokumento gyvavimo metu, tuo tarpu kai tradicinėje aplinkoje sukurtas dokumentas nuo jo sudarymo momento tampa visiškai nepriklausomas nuo priemonių, kuriomis jis buvo sudarytas.

Dokumento forma. Tradicinių dokumentų forma (šriftas, spausdinimo būdas, kalba, spalvos, specialūs ženklai, anspaudai ir t.t.) yra tiesiogiai atpažįstama ir suprantama vartotojo.

Tuo tarpu visa elektroninio dokumento forma paprastai vartotojui yra nežinoma. Elektroninio dokumento sudarytojo sukurtas žmogaus atpažįstamas ir suprantamas elektroninio dokumento vaizdas, matomas kompiuterio ekrane, yra tik dalis elektroninio dokumento struktūros, kurią žmogus gali suvokti. Didesnė dalis elektroninio dokumento formos elementų vartotojo nėra matomi, jie priklauso nuo techninės ir programinės įrangos, kurią naudojant dokumentas buvo sukurtas, savybių. Kiekvieną kartą transformuojant elektroninį dokumentą (pvz., į vartotojui atpažįstamą ir suprantamą vaizdą), elektroninio dokumento forma kinta, o elektroninių dokumentų perkėlimo iš vienu laikmenų į kitas (pvz., kopijuojant juos į naujesnes laikmenas ar keičiant programinę įrangą) metu pakeičiama dalis formos elementų, nors dokumento turinys ir jo loginė forma, kurią atpažįstama ir suprantama vartotojas, gali likti visiškai nepakitę. [7]

1.3. Metaduomenys

Kontekstinė informacija atspindi aplinką, kurioje dokumentas buvo sudarytas, dokumento sudarymo metu atliktus veiksmus, procedūras, kurių privaloma laikytis sudarant dokumentą, ir asmenis, dalyvavusius dokumento sudarymo procese. Kontekstinės informacijos sudedamosios dalys yra dokumento sudarytojas, adresatas, dokumento antraštė (pavadinimas), dokumento sudarymo vieta, dokumento data, registracijos įrašai, įrašai apie dokumento autorių ar jį sudariusį asmenį, jų parašai, specialūs administracinės ir teisines procedūras fiksuojantys įrašai – dokumento tvirtinimo, derinimo, dokumento perdavimo, gavimo žymos ir t. t. Dokumente fiksuojamos kontekstinės informacijos kiekį, formą, išdėstymo struktūrą ir pan. lemia dokumento administracinės ir teisinės aplinkos nustatyti reikalavimai ir taisyklės, kuriomis remiantis dokumente fiksuota kontekstinė informacija atpažįstama ir atskiriama nuo dokumento turinio. Tradiciniuose dokumentuose kontekstinė informacija paprastai yra pateikiama pačiame dokumente. Tai – dokumento sudarytojas, dokumento sudarymo data, vieta, dokumento pavadinimas, dokumento ryšiai konkrečios organizacijos dokumentų sistemoje ir pan.

Elektroninių dokumentų atveju šią funkciją atlieka metaduomenys, kurie apibrėžiami kaip duomenys apie duomenis, t. y. duomenys apie dokumento sudarymo aplinką ir dokumento struktūrą ir yra būtini dokumentui suprasti ir naudoti. Metaduomenys fiksuojami atskirai nuo dokumento turinio, dalis jų yra sukuriama automatiškai kompiuterio programinės įrangos, ir su dokumentu susiejami loginiais ryšiais.[2]

1.4. Elektroninių dokumentų valdymas

Tradicinių dokumentų saugojimas suprantamas kaip fizinių vienetų (popieriaus lapų, bylų, negatyvų, kino juostų ir kt.) saugojimas užtikrinant optimalias saugojimo sąlygas, kad dokumentai būtų išsaugoti reikiamą laiką ir išliktų prieinami ir galimi naudoti. Elektroninių dokumentų saugojimui taip pat keliami dideli reikalavimai, susiję su optimaliomis saugojimo sąlygomis. Tačiau vien optimalių saugojimo sąlygų, užtikrinančių tinkamą techninių priemonių saugomą, sudarymas negarantuoja elektroninių dokumentų išsaugojimo reikiamą laiką. Dėl greito techninės įrangos senėjimo, informacija gali tiesiog „dingti“, t. y. pasibaigus techninių priemonių galiojimo laikui (o jis yra gana trumpas – nuo keleto iki keliasdešimt metų) jose fiksuota informacija gali būti nebe prieinama ir nebegalima naudoti. Taip pat ir programinės įrangos senėjimas turi įtakos elektroninių dokumentų ilgalaikiam išsaugojimui. Todėl elektroniniai dokumentai periodiškai turi būti „atnaujinami“ – perkeliama (kopijuojami) į naujas technines saugojimo priemones ir prireikus perkeliama (konvertuojami) į naujus formatus, atitinkančius naujausios programinės įrangos reikalavimus.

Taigi elektroninis dokumentas yra sukuriama, perduodama, laikoma, naudojama ir saugojama skaitmeninėje aplinkoje visu jo gyvavimo metu. Elektroninis dokumentas susideda iš kompiuterio ekrane vartotojo matomo dokumento vaizdo, metaduomenų ir loginių ryšių, susiejančių ypatybių elektroniniai dokumentai yra priklausomi nuo technologinių priemonių – dėl techninės ir programinės įrangos senėjimo ir trumpalaikiškumo, dalis dokumentų gali būti prarasta. Todėl elektroninių dokumentų apyvartai organizacijoje užtikrinti yra būtinos papildomos priemonės, kurių pagrindinė – **elektroninių dokumentų valdymo sistema**. Toliau trumpai apžvelgsime, kokias funkcijas turi atlikti ši sistema, kad elektroniniai dokumentai atitiktų visus dokumentui keliamus reikalavimus.

Elektroninių dokumentų valdymo sistema – įstaigos dokumentų valdymo sistemos dalis, veikianti informacinių technologijų pagrindu, skirta rengti, tvarkyti, apskaityti ir saugoti elektroninius dokumentus. Elektroninių dokumentų valdymo reikalavimai turi atitikti bendrus dokumentų valdymo reikalavimus. [5]

2. RINKMENŲ PAIEŠKOS SISTEMŲ ANALIZĖ

2.1. Failų sistemos

Failų sistema (angl. filesystem) – duomenų struktūrų, algoritmų ir sąsajų visuma, leidžianti įrašyti, išsaugoti ir perskaityti laikmenoje ar kompiuterių tinkle esančią struktūruotą informaciją, nesigilinant į technines informacijos išsaugojimo detales. Daugumoje failų sistemų operuojama failo – tam tikro vientiso identifikuojamo informacijos rinkinio – sąvoka, daugelyje yra ir papildomų objektų, pvz., katalogas, failo atributai, itin išvystytose – yra ir sudėtingesnių objektų pvz., Files-11 – failo elementą apibrėžiantis įrašas, HFS – failo savybes ir naudojimą aprašanti resursų šaka ir pan. Failų sistemos atliekami veiksmai apibrėžiami rinkiniais funkcijų, leidžiančių šiuos objektus kurti, skaityti, keisti, šalinti ir t. t.

Tipiškos failų sistemos visada palaiko šias funkcijas ar panašiai besivadinančius jų ekvivalentus:

open(failo_vardas) – grąžina deskriptorių, identifikuojantį failą, su kuriuo dirbama.

seek(deskriptorius, pozicija) – perstumia failo rodyklę į atitinkamą poziciją.

read(deskriptorius, buferis, ilgis) – perskaito iš failo nurodytą kiekį informacijos į buferį.

write(deskriptorius, buferis, ilgis) – perrašo iš buferio nurodytą kiekį informacijos į failą.

close(deskriptorius) – uždaro failą.

Dauguma failų sistemų palaiko ir papildomas funkcijas, skirtas katalogų informacijos skaitymui, failų atributų, naudotojų teisių kontrolei ir pan. Kai kurios išvystytos failų sistemos leidžia atlikti ir žymiai sudėtingesnius veiksmus: dirbti su įvairiais įrenginiais (UNIX devfs), manipuluoti operacinės sistemos darbo parametrais (UNIX procfs) ar panaudoti daugumą operacinės sistemos funkcijų (Plan-9 9P).

FAT – grupė failų sistemų naudojamų DOS, Windows ir kitose OS;

HPFS – OS/2 naudota žurnalinė failų sistema;

NTFS – HPFS pagrindu sukurta failų sistema, naudojama Windows NT;

EXT2 – Linux naudojama failų sistema;

EXT3 – EXT2 pagrindu sukurta žurnalinė failų sistema, naudojama Linux;

ReiserFS – padidinto našumo žurnalinė failų sistema, naudojama Linux;

XFS – itin didelio našumo SGI kompanijos failų sistema, naudojama Irix ir Linux;

UFS – grupė itin patikimų failų sistemų, naudojamų įvairiose UNIX;

Files-11 – manipuliavimą įrašais palaikanti failų sistema, naudojama OpenVMS;

HFS – grupė objektinių failų sistemų, naudotų MacOS;

ISO 9660 – CD ir DVD diskuose naudojamos failų sistemos;

JFS – IBM firmos failų sistema, naudojama AIX, OS/2 ir Linux;

VxFS – Veritas firmos didelio našumo komercinė žurnalinė failų sistema, naudojama įvairių tipų UNIX;

ZFS – Sun firmos didelio našumo, išplėstas valdymo galimybes turinti failų sistema, naudojama Solaris;

9P – Plan-9 OS failų sistema ir protokolas;

NFS – tinklinė failų sistema, naudojama įvairiose OS;

SMB – tinklinė failų sistema, naudojama Windows tinkluose. [3]

Šiuolaikiniai kompiuteriai užtikrina didelių duomenų srautų gavimą ir jų saugojimą. Tačiau tebelieka didelė spraga tarp duomenų surinkimo bei saugojimo ir jų suvokimo. Dažna problema yra gauti svarbių žinių, suprasti duomenis, atskirti svarbią informaciją nuo menkavertės. Čia randa vietą žinių ir duomenų gavybos metodai (angl. *knowledge discovery, data mining*). Duomenų suvokimas yra sudėtingas uždavinys, ypač kai duomenys nurodo sudėtingą objektą, reiškini, kuris aprašytas daugeliu parametru, kurie gali būti skaitiniai, loginiai ir kt. Tokie duomenys vadinami daugiamačiais duomenimis. Duomenų suvokimas yra ilgo ir sudėtingo žinių gavybos proceso rezultatas. Jis apima daug etapų:

- suformuluojami analizės tikslai ir uždaviniai;
- iškeliamos pirminės hipotezės apie duomenų struktūras;
- formuojama duomenų imtis tyrimams;
- pasirenkami, kuriami nauji duomenų gavybos ir analizės metodai;
- analizuojami duomenų gavybos metodais gauti rezultatai;
- gautos žinios apibendrinamos, paneigiamos arba priimamos iškeltos hipotezės.

Šiuo metu pasaulyje yra sukurta nemažai įvairių statiško indeksavimo įrankių. Pradedant nuo griežtai hierarchiškai išrenkančių pagal rinkmenų pavadinimus iki semantinių paieškos sistemų. Dauguma vartotojų nėra linkę gilintis į reikšminių žodžių semantinę struktūrą, jų išdėstymą. Todėl semantinė rinkmenų paieškos sistema grindžiama tuo, kad sutelktų požymius, o ne vykdytų turinio analizę.

2.2. Windows rinkmenų paieškos sistemų apžvalga

„*Google desktop*“ paieškos sistema, kuri labai skiriasi nuo kitų įprastų darbalaukio paieškos sistemų. Su „*Google desktop*“ ieškoti rinkmenų yra lengva todėl, kad jau esame įpratę naudotis *Google* paieška internete. Programa randa tekstus, esančius elektroniniame pašte, muzikos, vaizdo rinkmenose. Indeksavimas vyksta ne realiu laiku, o po 30 sekundžių, kai kompiuteris yra ramiame stovyje, taip *Google* kūrėjai nori nesumažinti kompiuterio našumo. Šios programos išskirtinumas yra tas, kad ji indeksuojant rinkmeną, padaro jos atvaizdą. Taip

norima apsaugoti nuo ivykių, kai netyčia yra ištrinami tekstiniai failai. Išplestinėje paieškoje galima ieškoti pagal datą, žodžius, kurie įeina ir kurie neįeina į rinkmenas. [14]

Paieškos sistema „*Everything*“. Tai viena paprasčiausių paieškos programų. Jos minusas tai, kad neatlieka išplestinės paieškos, tačiau tai yra viena iš sparčiausiai veikiančių paieškos sistemų. Programa skirta žmonėms, kurie retai naudojami paieškos sistemomis bei taupo kompiuterio resursus. *Windows XP* terpėje sistemą „*Everything*“ indeksaciją atlieka per 1 sekundę, o 1 milijono rinkmenų indeksavimas užtrunka apie 1 minutę. Indeksavimas vyksta greitai todėl, kad indeksuojami yra tik failų pavadinimai. [15]

Dar viena labai galinga paieškos sistema tai „*DtSearch*“. Ši programa gali greitai atlikti paiešką tarp 1000 GB užimančių failų. Paieškos lauke rašant užklausą, yra pateikiamas sąrašas galimų žodžių, frazių, kas kartais labai palengvina informacijos paiešką. Norint, kad kompiuteris nenaudotų papildomų resursų, sistemoje galima atjungti indeksavimą tam tikrų duomenų, pvz.: neindeksuoti skaičių, papildomos rinkmenos informacijos, teksto esančio rinkmenoje. Ieškant teksto dokumentų, pateikiama yra ne tik tekstinė rinkmena, bet ir jos išvaizda ją atidarius. Taip pat atidarius šią programą, galima matyti, kada kokias užklauses buvo parašytos ieškant prieš tai. Tiesa, ši programa yra mokama.

Paieškos sistema „*Turbo Searcher*“ gali rasti duomenis, rinkmenas, nurodydama jų dydį. Su šia programa taip pat galima surasti atvaizdus, muzikos, video bei suarchyvuotas rinkmenas. Visa tai atliekama vienu spragtelėjimu pasirinkus kategoriją mp3, jpg, doc, pdf ir t.t. Norint surasti tam tikrą failą, galima panaudoti loginius operatorius (pvz.: XOR, AND, OR, NOT). „*Turbo Searcher*“ privalumas yra greitis, bet kokio tipo failą ji gali surasti tik per 1 sekundę. Kitas šios paieškos sistemos privalumas yra ne tik paieška sistemoje, bet ir tinkle.

„*Windows Search*“, seniau buvo žinomas kaip „*Windows Desktop Search*“. Ši sistema sukurta Microsoft Windows operacinėms sistemoms. „*Windows Search*“ turi daugybę išskirtinių bruožų, pvz.: *Ifilters* palaikymas, kurio dėka galima rasti bet kokio tipo failus. Taipogi dėka *Ifilters* „*Windows Search*“ suranda formato failą. Programos privalumas yra tai, kad ši paieškos sistema yra integruota įrankių juostoje, kas padeda vartotojui patogiau dirbti su šia paieškos sistema, tereikia paieškos laukelyje įvesti raktinį žodį ir paieškos rezultatas bus pateiktas.

2.3. Linux/MacOS rinkmenų paieškos sistemų apžvalga

Linux operacinei sistemai skirta paieškos programa „*Beagle Search*“. Ši paieškos sistema kaip ir kitos, skirtos ivairiems dokumentams rasti. Ją galima išskirti tuo, kad gali rasti web puslapius, kurie yra išsaugoti naršyklių istorijoje. Rinkmenų permainos yra indeksuojamos

realiame laike, failai indeksuojami iškart po jų modifikavimo, sukūrimo, elektroniniai laiškai iškart po gavimo, o web puslapiai indeksuojami, įvedus puslapio adresą. Kaip ir su „Apple“ kompanijos paieškos sistema, taip ir su „Beagle Search“ galima ieškoti žodžius, frazes dokumentų viduje, kas kartais labai praverčia. Unikalus „Beagle Search“ bruožas yra tai, kad ši paieškos sistema indeksuoja netgi kompaktinius diskus, atmintukus, prijungtus prie kompiuterio, kas palengvina paiešką išplėstinėje informacijos talpykloje. [16]

„Strigi“ tai lanksti, lengvai naudojama paieškos sistema. Ieškant su „Strigi“, rezultatai yra pateikiami labai greitai. Ši paieškos sistema išsiskiria iš kitų tuo, kad čia indeksavimas atliekamas realiu laiku, nenaudojant daugybės kompiuterio resursų. Kaip ir dauguma paieškos sistemų, paiešką galima atlikti pagal datą, dydį, adresą, „Strigi“ indeksuoja *.pdf, *.mp3 bylas, tekstinius failus ir t.t. Nors ši programa atlieka paiešką nesigilintant kas yra rinkmenos viduje, tačiau atlieka tai labai greitai. Nors ši sistema yra intergruota neseniai išleistoje ubuntu 9.04 distribucijoje, tačiau ją galima atsisisiusti ir Macintosh operacinėse sistemose. [17]

Su paieškos sistema „Tracker search“ galima savo kompiuteryje rasti dokumentų, paveikslėlių, muzikos rinkmenų. Ši paieškos sistema yra inergruota Linux Ubuntu operacinėje sistemoje. „Tracker search“ privalumas yra tas, kad paieškos rezultatą galima išsaugoti, kuris yra pateikiamas teksto pavidalu. Žvelgiant iš kitos pusės, galima išsaugoti tik pasirinktus rezultatus, kuriuos parodo paieškos sistema.

Vykdamas paiešką su „Tracker search“ rezultatuose yra pateikiami dokumentai, kurie neatitinka raktinio žodžio, jie yra pateikiami nuo pirmutinio rezultato, pvz.: kai įvedam žodį „linux“, rezultate yra pateikiamas žodis „linux ubuntu“, sekanti pateikiama rinkmena su pavadinimu „ubuntu 8.10“. Iš vienos pusės programa ieško platesniu mastu, tačiau iš kitos pusės, jei vartotojas ieško tam tikro failo, jam yra pateikiama daugybė nereikalingų atsakymų.

Jau nuo 2005 metų Macintosh 10.4 operacinėse sistemose failus galima buvo ieškoti su „Spotlight“ paieškos sistema. Su ja galima ne tik greitai, lengvai rasti dokumentų pagal jų rūšį, bet ir ieškoti rinkmenų pagal įdėjimo, modifikacijos datą, laiką. „Spotlight“ nuo kitų paieškos sistemų galima būtų išskirti tuo, kad ji turi galimybę ieškoti žodžius, frazes dokumentų viduje. Taip pat su šią paieškos sistema galima naudoti operatorius („AND“, „OR“, „NOT“, „|“, „-“), kas padeda išplėsti arba susiaurinti paieškos parametrus. Vizualiai šią sistemą galima rasti paspaudus meniu juostoje, dešinėje pusėje esančią ikoną, arba galima rasti per patį Macintosh meniu.

Programos vykdymas skiriasi tik tuo, kad vienur galima atlikti greitą paiešką, įvedus tik tekstą, o kitur išplėstinę. Ši paieškos sistema naudojama išskirtinai tik Apple kompiuteriuose ir Apple Iphone mobiliuosiuose telefonuose.

Nėra paieškos sistemos, kuri visiškai atitiktų visus vartotojų poreikius. Vieniems reikia paieškos sistemos, kuri praktiškai galėtų surasti viską. Antriems reikalinga tokia paieškos sistema, kuri indeksuotų tik tam tikras rinkmenas, kad išsaugoti kompiuterio resursus.

1 lentelė. Rinkmenų paieškos įrankių palyginimas

Įrankis	<i>Google desktop</i>	<i>Windows Search</i>	<i>Everything</i>	<i>DtSearch</i>	<i>Turbo Searcher</i>	Paiška	<i>Strigi</i>	<i>Spotlight</i>	<i>Beagle Search</i>
Vertinimo Kriterijus									
Greitas indeksavimas	+	+	+	+	+	+	+	+	+
Išplėstinė paieška	+	+	-	+	+	+	-	+	+
Paieška tinkle	+	+	-	-	+	+	-	-	-
Tekstinės informacijos paieška pagal turinį	-	+	-	+	+	-	-	+	+
Paprastas administravimas pradedančiajam naudotojui	+	+	+	+	+	+	+	-	+
Mokama	-	+	-	+	-	-	-	-	-
Nemokama	+	-	+	-	+	+	+	+	+

Vartotojams, kurie kompiuteriuose saugo ypač daug tekstinių failų, tačiau nežino jų pavadinimų, patogiausia naudoti „*Beagle Search*“ (Linux sistemoje) ir „*Spotlight*“ (Apple kompiuteriuose), šios programos be jokių kliūčių apdoroja vidinę tekstinę informaciją. „*Windows Search*“ funkcijos *Ifilters* dėka, Windows XP/Vista/Seven vartotojai taip pat gali ieškoti rinkmenas pagal tekstus viduje.

Vartotojams, kurie rengia paieškos ataskaitas, tinkamiausia „*Tracker search*“ paieškos sistema, nes joje paieškos rezultatą galima išsaugoti dokumentuose tekstinio failo pavidalu.

Žmonės, kurie labai dažnai naudojami paieškos sistemomis, kurie tikisi greitos paieškos, kurie žino ko nereikia indeksuoti, kad sutaupyti kompiuterio resursus, gali naudotis „*DtSearch*“ sistema, kuri tarp 1 TB rinkmenų, gali greitai ir tiksliai rasti reikalingą failą.

3. PAIEŠKOS TECHNOLOGIJOS IR JŲ PROBLEMŲ SPRENDIMŲ APŽVALGA

Šiuolaikinių paieškos sistemų branduolį sudaro failų indeksavimo mechanizmas, kuris apdoroja duomenis apie failus ir sukurią efektyvų paieškos nuorodų įrankį. Daugelis šių sistemų naudoja, aukšto lygio, įvairiapusias informacijos paieškos biblioteksa, kurios generuoja bet kokio skaitmeninio formato metaduomenis, paversdamos į tekstinius formatus nepriklausomai nuo duomenų šaltinio vietos, dydžio, kalbos ir t.t. Indeksavimą galima suskirstyti į tris pagrindines dalis: duomenų pavertimas į tekstinę informaciją, gautos tekstinės informacijos analizavimas ir išsaugijimas.

Šiuolaikinės paieškos sistemos gali indeksuoti bet kokio formato failą. Gauti metaduomenys kaip virtualus dokumentas susideda iš atitinkamų laukų aprašančių dokumentą. Kiekvienas laukas atitinka dalį duomenų, kurie yra peržiūrėti ir susisteminti indeksavimo metu.

Analizuodami šiuolaikines paieškos sistemas, pastebime, kad vykdydama užklausa, sistema tarp rezultatų pateikia ir daug visai su užklausa nesusijusių elementų. Taip atsitinka tada kai pasirenkamas per daug platus paieškos laukas. Siekiant to išvengti, atsižvelgiant į dokumentų rinkinį, būtina formalizuoti paieškos kriterijus taip, kad tinkmų paieškos rezultatų tarpe kuo atsirastų mažiau nereikalingo balasto.

Tegu S bus paieškos sritis, paieškos kriterijai – I , o užklausa q . Tuomet paieškos rezultatai – R . Tinkami paieškos rezultatai – R_a .

Atranka: paieškos srities ir tinkamų rezultatų santykis $\frac{|R_a|}{|S|}$

Tikslinimas: visų paieškos rezultatų ir tinkamų paieškos rezultatų santykis $\frac{|R_a|}{|R|}$

Atranka ir *Tikslinimas* yra būtini procesai, siekiant sutrumpinti paieškos laiką ir užkirsti kelią nereikalingų failų patekimui į sąrašą.

Tai, ar dokumentas yra tinkamas ar ne, turi būti nustatoma vykdant užklausa. Tačiau kai kuriais atvejais atliekant šią analizę neišvengiama klaidų. Kuomet vartotojas nepakankamai tiksliai apibrėžia paieškos sritį. Tada užklausa sistemai nėra aiški ir paieškos rezultatuose atsiranda daug nereikalingų elementų.

Yra žinomi 2 pagrindiniai dokumentų paieškos būdai. [6] Pirmasis būdas – papildomai saugoti reikalingą informaciją apie dokumentą. Tai yra metaduomenų paieškos modelis. Antrasis būdas – pilna tekstinė paieška. Einama per visą dokumentą ir tikrinamas kiekvienas žodis. Šis metodas turi nemažai trūkumų. Visų pirma, tokia paieška yra žymiai lėtesnė už paiešką naudojant metaduomenis. Antra, dokumentas gali būti ne tik tekstinis, arba jo formatas

nesuprantamas paieškos sistemai. Šio projekto tikslas – dokumentų valdymo sistema, kuri nėra priklausoma nuo dokumentų, taigi pilna tekstinė paieška nėra tinkamiausias paieškos būdas. Antra vertus, galima naudoti šį būdą kaip papildomą, kuris praplėstų paieškos sistemos galimybes.

Naudojant metaduomenis, svarbiausia žinoti, ką kiekvienas metaduomenų laukas reiškia. Pavyzdžiui, turime metaduomenis, kurių vienas laukas reiškia autorių. Lauko reikšmė – „Me Lung“. Žinome, kad tai yra vardas ir pavardė, bet kuris žodis reiškia vardą, kuris pavardę, nėra aišku. Šis pavyzdys iliustruoja tai, kad svarbu ne tik turėti metaduomenis, bet kartu jie turi būti tinkamai struktūrizuoti, kad neiškiltų dviprasmybių paieškos sistemai, o kartu ir vartotojui. Gera paieškos sistema galėtų žinoti, kad „Me“ reiškia vardą, o „Lung“ pavardę.

Kai kurie dokumento tipai, pavyzdžiui, Microsoft Office dokumentai, *.pdf rinkmenos, saugo metaduomenis apie dokumentą kartu su dokumento turiniu, t.y. įterpia save į dokumento failą. Šias dokumento savybes galima panaudoti indeksuojant dokumentus, apie kuriuos neturima metaduomenų.

Kaip teigia organizacija Infocap [10] neišvengiamai egzistuoja problemos dokumento turinio aprašymui naudojant realius žodžius, klasifikacijos schemas arba raktinius žodžius (keywords). Dažniausiai dokumentus indeksuoja vienas asmuo, o paiešką suindeksuotuose dokumentuose atlieka kitas asmuo. Tą patį dalyką kiekvienas žmogus supranta ir aprašo skirtingai, todėl ieškantis dokumentų žmogus gali jų nerasti, nes dokumento apibūdinimui buvo pavartoti kiti žodžiai ir frazės. Taipogi skirtingi žmonės tame pačiame dokumente gali ieškoti skirtingos informacijos. Raktiniai žodžiai negali aprašyti visos dokumente saugomos informacijos. Dalinis problemos sprendimo būdas – papildyti ir atnaujinti raktinius žodžius pagal vartotojų užklausas, tačiau ateityje vartotojas dažniausiai ieškos vis kitos informacijos.

Plačiausiai paplitęs indeksavimo metodas – indeksuoti kiekvieną žodį. Šiuo atveju suindeksuotos informacijos dydis siekia 180 – 350% pradinio dokumento. Šio metodo trūkumas yra tai, kad įvedus klaidingą užklausą (klaidingai įvedus bent vieną simbolį) sistema nerastų vartotojo ieškomos informacijos arba ras ne tai ko ieško. Indeksuojant kiekvieną žodį atsiranda dar viena problema. Skanuojant dokumentą ir po to naudojant optines teksto atpažinimo programas (OCR) gali dokumente įsivelti klaida. Jei ši klaida bus tame žodyje, kurio ieškos vartotojas, paieškos sistema šiuo atveju irgi nerastų ieškomos informacijos.

Infocap dokumento turinio indeksavimui siūlo naudoti adaptyvią šabloninę atpažinimo technologiją (Adaptive Pattern Recognition Processing). Ši technologija kiekvieną simbolį užrašo 8 bitų seka. Pavyzdžiui, žodžiai *dog* bei *dig* užrašomi taip:

0	1	1	0	0	1	0	0
0	1	1	0	1	1	1	1
0	1	1	0	0	1	1	1

dog

0	1	1	0	0	1	0	0
0	1	1	0	1	0	0	1
0	1	1	0	0	1	1	1

dig

Nesunku pastebėti, kad šių dviejų žodžių užrašymas skiriasi tik 2 bitais iš 24, taigi sutampa 91%. Vartotojui suklydus ir vietoj žodžio *dog* įvedus žodį *dig*, šiuo atveju sistema suras ir dokumentus ir su žodžiu *dog*, ir su žodžiu *dig*. Galima paieškos rezultatus surūšiuoti pagal atitikimą užklausiai. Taigi šiuo atveju paieškos patikimumas bus net 91%.

3.1. LAN paieškos apibūdinimas

Riboto paraiškų skaičiaus (limited assignment number) paieškos algoritmas dažniausiai remiasi grįžimo (backtracking) metodu.[3] Šio algoritmo rezultatas visuomet orientuotas į sprendinius. [8] Algoritmo pagrindinė užduotis yra identifikuoti paraiškų sritį, tenkinančią duotą sąlygą (apribojimus). Užduotis gali būti neįvykdoma, jei neįmanoma patenkinti visus apribojimus. Kartais klasikinis grįžimo algoritmas ne visuomet apima visą sprendinių aibę. Tam gali pasitarnauti apribojimų paskirstymo (constraint propagation) algoritmas. Kiekviename žingsnyje, yra parenkamas kintamasis, kuriam priskiriama reikšmė. Kiekviena paraiška yra paskirstoma per apribojimus į kitų kintamųjų sritis.

Su kiekvienu kintamuoju, LAN paieškos algoritmas atlieka skaičiavimus tiek laiko, kiek kartų reikšmė buvo priskirta jam. Algoritmas turi baigtinį iteracijų skaičių. Jei riba yra viršyta, kintamasis palieka nepaskirtas, ir paieška tęsiasi su kitais kintamaisiais. Nepaskirtų kintamųjų žymėjimas nėra galimas net naudojant grįžimo algoritmą. Šios algoritmo dėka, dalinės kintamųjų paraiškos yra gaunamos kartu su likusių nepaskirtų kintamųjų kompleksu.

LAN algoritmas realizuoja kintamųjų ir jų verčių euristinę paiešką, pagrįstą iteracinių skaičiavimų rezultatais:

- sėkmingai paskirtų ankstesniose iteracijose kintamųjų vertės, yra tikrinamos iš pradžių per kitą iteraciją, kai tinkama vertė kintamajam surandama, belieka tinkamiausią paraišką priskirti šiam kintamajam;
- kiekvienam kintamajam, kuris pasilieka nepaskirtas ankstesniose iteracijose, suteikiama reikšmė, kuri bus tikrinama vėliau;
- bet koks kintamasis, nepaskirtas ankstesniose iteracijose, yra pažymimas, ir jam suteikiama pirmumo teisė paskesniuose skaičiavimuose.

Atliekant pirminius skaičiavimus, dažniausiai remiamasi standartiniais euristiniais metodais.[8] Kiekviename iš kitų kartojimų, euristika, pagrįsta ankstesniu kartojimu, yra panaudota pirmiausiai. Vartotojas taip pat pats gali pakeisti paieškos kriterijus po kiekvieno kartojimo, jeigu jo netenkina paieškos rezultatas. Yra keletas atvejų, kai LAN paieška neduoda norimo rezultato. Tuomet visi kintamieji lieka nepaskirti. Taip gali nutikti tada, kai įvedama per

daug apribojimų. Čia vartotojas turi sumažinti apribojimų skaičių ir iš naujo pradėti algoritmą. Paieška gali taip pat nepasisiekti, jei kintamojo sritis yra ištuštinta anksčiau, negu viršyta jo riba ir, tai yra pirmas kintamasis, ar visi jo ankstesni kintamieji jau viršijo ribą. Šiuo atveju, algoritmas gali būti iš naujo pradėtas be šitų kintamųjų. [10]

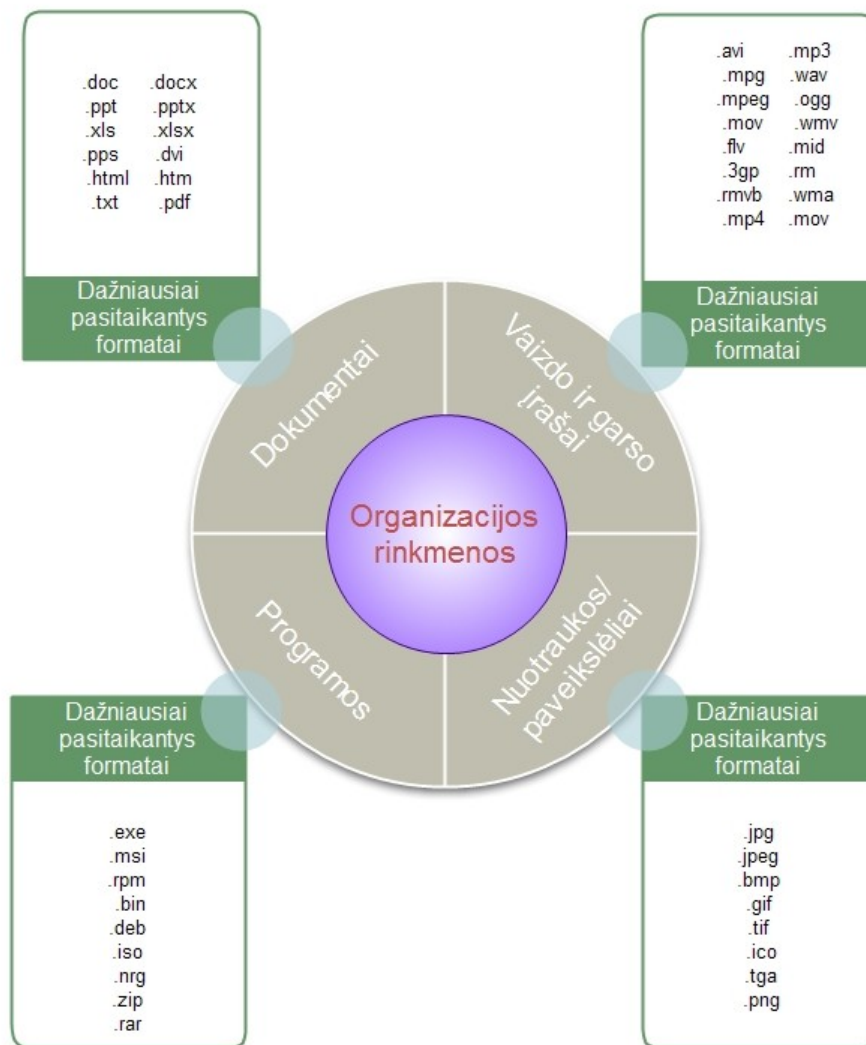
3.2. Reikalavimų serverio rinkmenų paieškos sistemai analizė

Dokumentų paieškos sistemos pagrindinė užduotis – surasti organizacijos turimus dokumentus. Kaip paminėta įvade, dokumentu gali būti laikoma bet kokia rinkmena: tekstinė, audio, video, grafinė ir t.t. Sistema neturi būti priklausoma nuo turimų dokumentų. Tuo pačiu ji neturi būti priklausoma nuo organizacijos vykdomos veiklos. Sistemos veiklos sritis – organizacijos dokumentai. Pastarųjų kilmė gali būti bet koks organizacijoje vykstantis procesas. Dokumentų paieškos sistemos uždaviniai susiję su dokumentų valdymu, todėl analizuojama tik dokumentų naudojimo ir valdymo organizacijoje problemos ir sritys.

Kuriant rinkmenų paieškos sistemą, svarbu žinoti, kaip vartotojai kaupia savo rinkmenas failų sistemoje. Be to, svarbu numatyti paieškos sistemos elgesį vartotojo atžvilgiu.

Dar 1995 metais grupė mokslininkų atliko tyrimus su trimis skirtingomis operacinėmis sistemomis (Windows, DOS ir MacOS) ir nustatė, kad sukauptai informacijai kompiuteryje būdingos šios savybės:

1. Efemeriška: trumpalaikė informacija (įvairios užduotys, tekstiniai užrašai). Šio tipo informacija paprastai laikoma viršutinio lygio katalogose ir naudojama labai dažnai.
2. Darbinė: dažnai naudojami darbiniai projektai (tekstiniai failai), kurie gali būti naudojami savaitių ar mėnesių intervalu. Šios informacijos kaupimas dažniausiai vykdomas kataloguose pagal kategorijas.
3. Archyvinė: tai informacija, kuri saugoma mėnesius, metus ir daugiau. Ši informacija yra retai naudojama.[1]



1 pav. Organizacijos rinkmenų struktūra

Detalesniems ir aiškesniems poreikiams ir reikalavimams duomenų valdymo sistemai sudaryti reikalingas virtualios organizacijos struktūros, jos tikslų, procesų modeliai. Organizacija dokumentus kuria ir naudoja sprendama ir vykdydama jai iškeltus uždavinius, vykdydama tam tikrus procesus, todėl šie procesai ir užduoda reikalavimus ir poreikius dokumentų valdymo sistemai. Pagal organizacijos poreikius bus keliami tikslūs funkciniai reikalavimai dokumentų valdymo sistemai.

3.3. Vartotojų ir jų poreikių analizė

Atlikus Tauragės rajono savivaldybės administracijos dokumentų valdymo sistemos analizę, ir ištyrus organizacijos resursus, paaiškėjo, kad poreikis dalintis elektroniniais dokumentais tarpusavyje yra tikrai didelis. Norint surasti reikiamą dokumentą, pirmiausia reikia išsiaiškinti, kurio darbuotojo kompiuteryje tas dokumentas saugomas. Tada reikia kreiptis į

būtent tą darbuotoją, kad jis atsiųstų surastą dokumentą. Šie veiksmai atima nemažai laiko sanaudų.

Dokumentų paieškos sistemos vartotojas – asmuo, kuris naudoja sistemą dokumentų saugojimui, jų paieškai, apsaugai nuo nesankcionuoto priėjimo ir pan.

Vartotojų poreikiai dokumentų paieškos sistemai – tai reikalavimai ir norai sistemos funkcionalumui. Vartotojas nori, kad sistema atliktų bei išspręstų jam aktualius uždavinius ir procesus. Vartotojų poreikiai formuluojami pasinaudojant analizės etapo modeliais – tikslų, procesų, veiklos panaudojimo atvejų.

Sistemos vartotojo poreikiai:

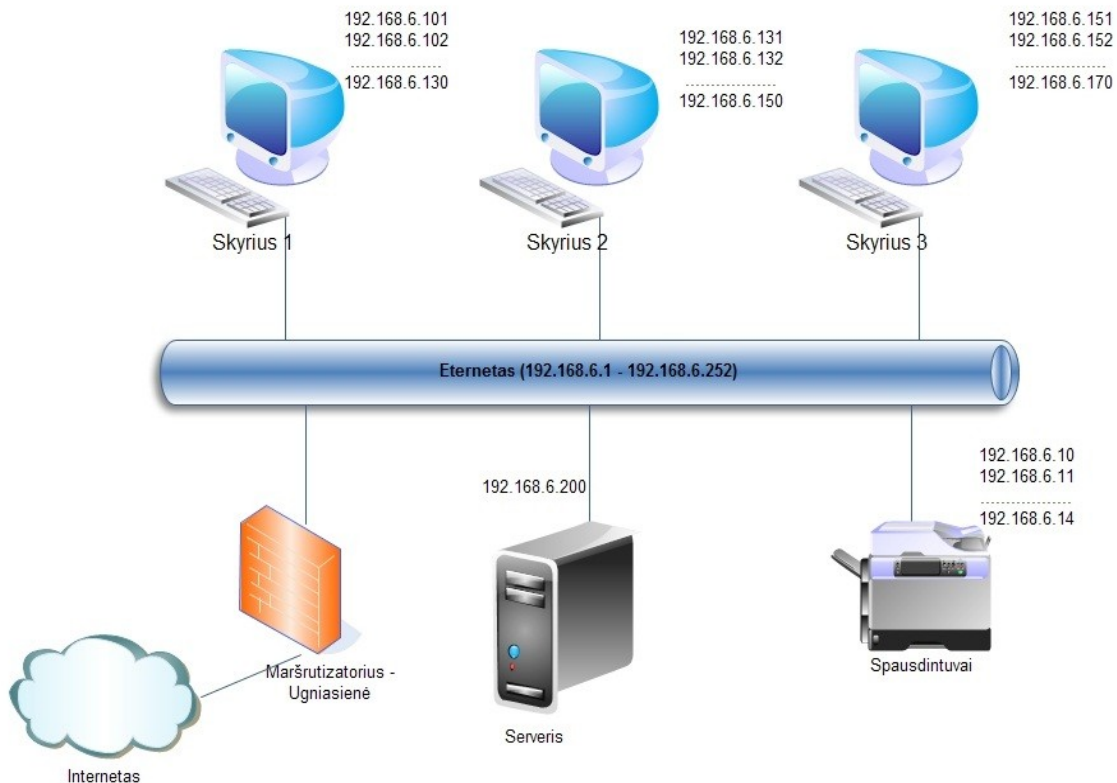
- Naujo ar pakeisto dokumento saugojimas sistemoje
- Dokumento paieška sistemoje
- Priėjimas prie dokumentų per LAN tinklą
- Informacijos personalizavimas
- Dokumentų klasifikavimas pagal įvairius kriterijus
- Pilnas dokumento gavybos ciklas

Šie išvardinti poreikiai reikalingi pilnos ir išbaigtos dokumentų paieškos sistemos sukūrimui. Visi šie poreikiai yra svarbūs, nes vartotojų poreikiai organizacijoje yra įvairūs, ir jiems visiems realizuoti reikia įgyvendinti šiuos uždavinius.

Mano tiriamoje organizacijoje – Tauragės rajono savivaldybėje yra 98 kompiuterinės darbo vietos. Iš jų 94-iose Windows XP Professional programinė įranga. Vidutinis įstaigos kompiuterio procesoriaus taktinis dažnis 2,7 Ghz. Vidutinis operatyviosios atminties kiekis - 2 GB. Minėti kompiuteriai, taip pat ir aptarnaujantis serveris per komutatorius yra prijungti prie maršrutizatoriaus, taip sudarydami vieną bendrą potinklį (2 pav.)

Serverio parametrai:

- 4 Intel Xeon procesoriaus branduoliai;
- 10 Mbps tinklo pralaidumas;
- 4 GB operatyviosios atminties;
- *MySQL, PostgreSQL* duomenų bazės;
- Windows Server 2.0.03 Standart Edition operacinė sistema



2 pav. Savivaldybės vietinio tinklo modelis

3.4. Nefunkciniai reikalavimai sistemai

Sistemos pagrindinė paskirtis – palengvinti organizacijos uždavinį tvarkyti ir valdyti dokumentų srautą, greitai ir nesudėtingai rasti reikiamus dokumentus. Labiausiai tikėtinas sistemos vartotojas – neturintis profesionalių žinių informacinių technologijų srityje.

Atsižvelgiant į tai, sudaromi nefunkciniai reikalavimai sistemai. Nefunkcinius reikalavimus galima suskaidyti į detalesnius. Sistemai keliami tokie nefunkciniai reikalavimai:

Paprastumas – darbas su sistema turi būti nesudėtingas, nereikalaujantis iš vartotojo gilių žinių informacinių technologijų srityje.

Patikimumas – sistema turi būti patikima, t.y. korektiškai atlikti jai pavestas užduotis ir funkcijas.

Efektyvumas – sistema turi atlikti jos vykdomas funkcijas greičiau, negu tai būtų atliekama be sistemos, ir kad jos atliekamos funkcijos ir jų atlikimo sparta būtų realiai naudingi sistemos vartotojams.

Nesudėtingas eksploatavimas – sistemos naudojimas neturi reikalauti didesnių papildomų žmogiškųjų išteklių bei materialinių resursų. Grafinės sąsaja turi būti paprasta, patogi, logiška.

4. SERVERIO RINKMENŲ PAIEŠKOS IR ANALIZĖS SISTEMOS MODELIS

4.1. Techninė užduotis

Pagal analizės dalyje išnagrinėtus ir apibrėžtus vartotojų reikalavimus numatoma realizuoti paieškos sistemą, kuri įvykdytų pagrindinį uždavinį, t.y. atliktų serverio bei vietiniame tinkle sujungtuose kompiuteriuose rinkmenų paiešką, pagal šiuos kriterijus:

- pagal ieškomo failo vardą arba vardo fragmentą;
- pagal failo tipą;
- pagal failo dydį.

Vykdydama paiešką pagal failo tipą, programa turi rasti šių tipų failus:

- Vaizdo įrašai - *.avi, *.mpg, *.mpeg, *.mov, *.mp4, *.flv;
- Garso įrašai - *.mp3, *.wav, *.mid, *.wma;
- Paveikslėliai - *.jpg, *.jpeg, *.bmp, *.gif, *.ico, *.tga, *.png;
- Dokumentai - *.doc, *.rtf, *.txt, *.ppt, *.xl, *.pdf, *.htm, *.html;
- Programos - *.exe, *.msi, *.rpm, *.bin, *.def, *.iso, *.zip, *.rar;
- Programinis kodas - *.c, *.cpp, *.java, *.jar, *.vbproj;

Gauti rezultatai būtų pateikti sąrašė, kuriame matytųsi:

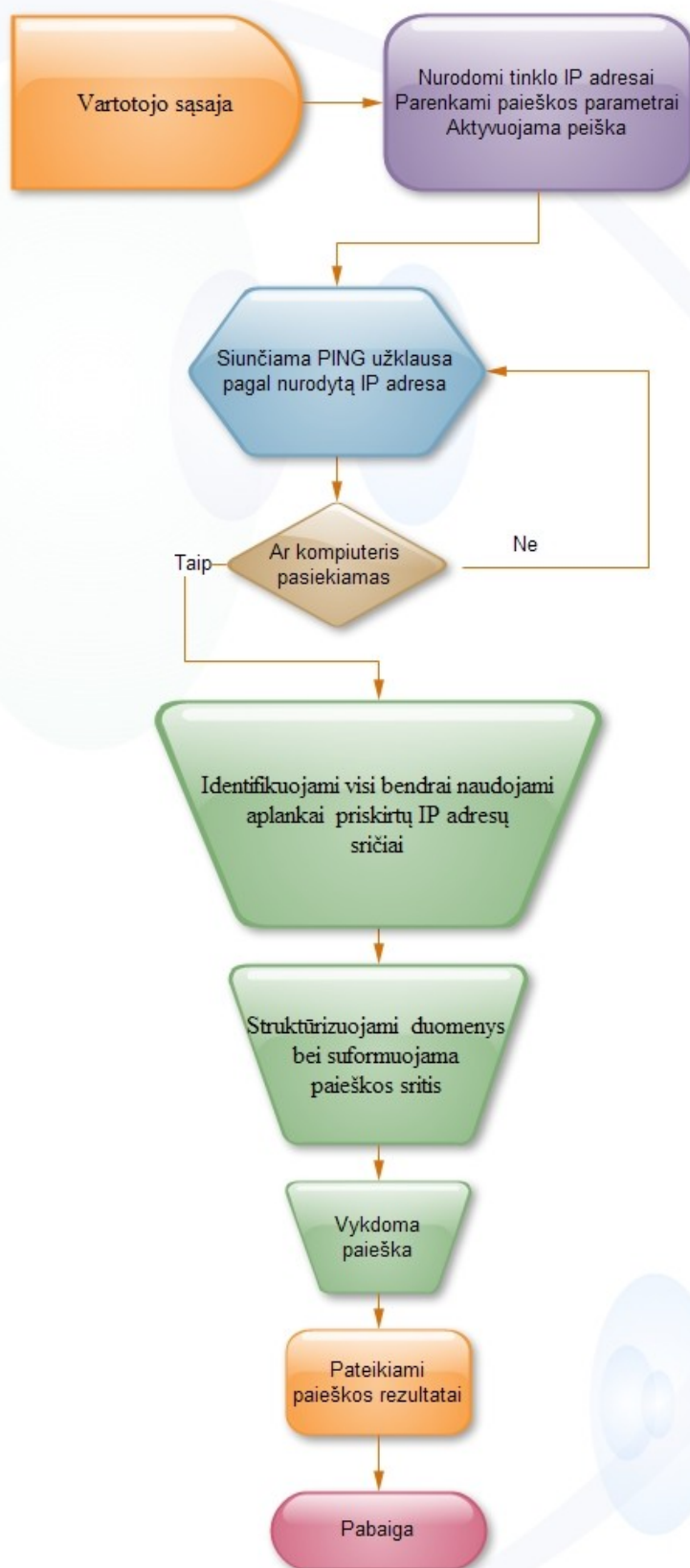
- failo vardas;
- dydis;
- failo vieta.

Programa turi leisti kopijuoti failą iš sąrašo, į pasirinktą kompiuterio aplanką. Programą numatoma realizuoti C# programavimo kalba, Microsoft Visual C# aplinkoje. Sistemos vartotojai, naudodamiesi programa *Paieška* gali atlikti tam tikrus veiksmus.

Programos įvykių srautas:

- 1 žingsnis. Vartotojas paleidžia programą;
- 2 žingsnis. Vartotojas meniu juostoje pasirenka komandą *Nustatymai*;
- 3 žingsnis. Nustatymų kortelėje vartotojas suveda tinklo kompiuterių IP adresus;
- 4 žingsnis. Vartotojas patvirtina įvestus IP adresus (sekantį kartą paleidus programą, IP adresų suvedinėti nebereikia);
- 5 žingsnis. Vartotojas paieškos laukelyje įrašo ieškomo failo vardą, vardo dalį arba failo plėtinį;
- 6 žingsnis. Vartotojas patikslina paieškos parametrus, (šį žingsnį galima praleisti);

- 7 žingsnis. Vartotojas aktyvuoja paiešką;
- 8 žingsnis. Sistema vykdo paiešką;
- 9 žingsnis. Sistema sudaro rastų rinkmenų sąrašą.



3 pav. Serverio rinkmenų paieškos programos „Paieška“ architektūros modelis

4.2. Programos „Paieška“ užklausų savybės

Paprastai, prieš pradėdami paiešką, vartotojai apsisprendžia ko jie ieškos ir pasirinkę failo tipą (pvz., el. paštas, tekstinis dokumentas, paveikslėlis ir t.t.) paieškos laukelyje įveda raktinį žodį.

Yra nemažai sukurta paieškos metodų. Vienas iš jų yra Microsoft Research laboratorijose sukurtas metodas *Matyta Medžiaga* (angl. Stuff I've Seen). [4]

Viena iš pagrindinių šio metodo prielaidų yra tai, kad dauguma rinkmenų, kurias vartotojas nori rasti buvo anksčiau matytos. [4] *Matyta Medžiaga* indeksuoja visus duomenis, esančius kompiuteryje, todėl vartotojui pateikiama medžiaga iš įvairių šaltinių (pvz., el. pašto laišakai, tekstiniai dokumentai, paveikslėliai ir t.t.). Sistemos grafinė vartotojo sąsaja pateikia rezultatus pagal datą, turinį, pavadinimą, autorių, šaltinį ir t.t.

Kreipimasi į failą pavadinime f , o į failų rinkinį F . Kiekvienam failui f būdingi šie trys požymiai:

kelias (f) visas failo kelias $c:f\backslash g\backslash \text{dienynas.txt}$, kur **kelias** (f) elementų intervalas $\{c, f, g, \text{dienynas}, \text{txt}\}$;

pavadinimas (f) pilnas failo pavadinimas $c:\backslash f\backslash g\backslash \text{dienynas.txt}$, kur **pavadinimas** (f) intervalas $\{\text{dienynas}, \text{txt}\}$;

turinys (f) jei failas neturi tekstinio turinio, pavyzdžiui, failas yra paveikslėlis arba garso failas, tada **turinys** (f) = 0 .

Tegul q bus užklausa, t.y. paieškos frazė, o F - failų rinkinys. Paieškos rezultatai $\text{Rezult}(q, F)$. Visiems q iš F galioja lygybė $j \in F$ jei yra tenkinamas bent vienas iš aukščiau minėtų požymių. Mano sistemoje (aprašyta 4 skyriuje), kai vartotojas pasiunčia užklausa q , failų sistema susidedanti iš bylų F , vartotojui pateikia rezultatų sąrašas $\text{Rezult}(q, F)$. Vartotojas gali pasirinkti vieną iš pateiktų rezultatų $\text{Rezult}(q, F)$. Iš to išplaukia dvi svarbios prielaidos:

Failų atrinkimas. Įvykdžius užklausa q pateikiami rezultatai. Išskiriami failai $f_{q|F} \in F$

Failo išskyrimas. Kai vartotojas įvykdo užklausa q ir pasirenka failą iš pateikto rezultatų sąrašo, tada turime $f = f_{q|F}$

Įprastinėje failų sistemoje, kiekvienas failas turi savo metaduomenis: failo pavadinimas, tipas, dydis, sukūrimo data, vieta, ir t.t.

A priori visi šie atributai yra svarbūs formuojant paieškos užklausas. Mano sistema pateikia failo vardą, failo vietą ir dydį.

Užklausa q ir failų rinkinį F , kur kiekvienas failas $f \in F$ galima išreikšti vektoriumi $f(q) \in R$. Kiekviena funkcija atitinka vieną ar kitą failo savybę (dydį, tipą, vardą).

Individualios užklausų ypatybės

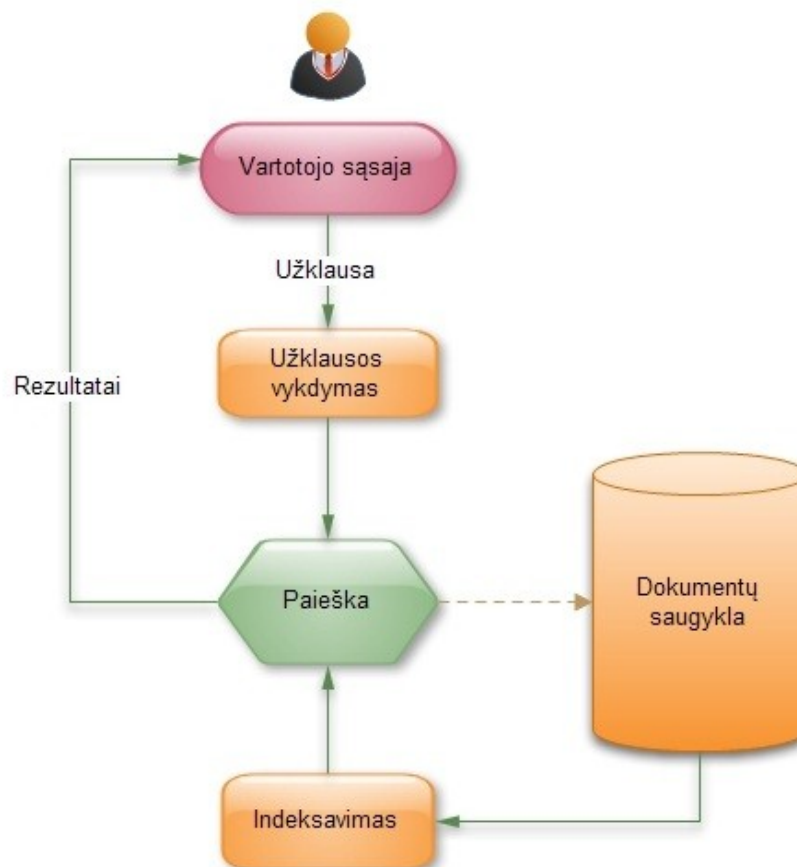
Visų pirma paminėsime į failo savybes, kurių reikšmės nustatomos nepriklausomai nuo konkretaus vartotojo užklausų.

- $D_{YDIS}(f)$ fiksuojamas santykinis dydis f . Tarkime, kad $D_{YDIS}(f) = 1$ (atitinkamai gali būti 0,8; 0,6; 0,4; 0,2)

- $G_{YLIS}(f)$ yra failo (f) atstumas iki pirminės direktorijos. Pavyzdžiui $G_{YLIS}(c:\backslash zoo.txt) = 1$. Tada nesunku pastebėti, kad $G_{YLIS}(c:\backslash fg\backslash zoo.txt) = 1/3$

- $TIPASX$ naudojamas failo formatui aprašyti, kur X yra viena iš galimų reikšmių (doc,txt, pdf, ppt, html ir t.t.) $TIPASPDF(f) = 1$ jeigu (f) yra PDF dokumentas.

Vartotojas gali pateikti užklausą pagal failo vardą (vardo dalį), failo tipą, dydį. Gali susiaurinti arba padidinti paieškos sritį nurodydamas tinklo kompiuterių IP adresus. Grafinė vartotojo sąsaja pateikia paieškos rezultatų sąrašą, suformuotą funkcijos $Rezult(q, F)$, kuriame nurodytas failo vardas, tipas, vieta kokiam kompiuteryje ir kataloge yra failas. Vartotojas gali iš sąrašo išsirinkti pageidaujamą failą ir jį nusikopijuoti, arba iš karto atidaryti.



4 pav. Serverio rinkmenų paieškos programos „Paieška“ procesų modelis

4.3. „Paieška“ programos kodo aprašymas

Šioje dalyje aprašoma, kaip veikia programa serverio rinkmenų programa „Paieška“. Visų pirma vartotojas turi apibrėžti tinklo kompiuterių IP adresų diapazoną. Sistema tinkle ieško aktyvių kompiuterių. (3 pav.) Jeigu norime paiešką vykdyti tik tame kompiuteryje kuriame yra paleista programa, nurodome savo kompiuterio IP adresą. Programa, nustačiusi tinkle bendrai naudojamų aplankų sąrašą, ieško rinkmenų pagal reikšminį žodį, failo tipą, dydį ir kt.

Pirmas programos uždavinys – nustatyti, ar nurodytu IP adresu kompiuteris yra pasiekiamas. Tam naudojama *Ping* klasė, aprašyta *System.Net.NetworkingInformation* vardų srityje. *Ping* funkcija gali būti realizuojama dviem būdais: sinchroniniu ir asinchroniniu metodais. Siekdamas išvengti paketų praradimo, aš naudoju asinchroninį metodą.

```
Ping pingSender = new Ping();
    // iškviečiamas PingCompletedCallback metodas
pingSender.PingCompleted += new PingCompletedEventHandler(AddHost);
    // kartojama kas 1000 milisekundžių
int timeout = 1000 ;
    // sukuriamas 32 baitų failų mainų buferis
byte[] buffer = new byte[] { 100 };
    //paieškos diapazonas 64 maršrutizatoriai arba komutatoriai
PingOptions options = new PingOptions(64, true);
pingSender.SendAsync("192.168.209.178", timeout, buffer, options, null);
```

Kai *Ping* operacija įvykdyta, aktyvuojama funkcija *AddHost*. Čia matome *Ping* operacijos rezultatų tikrinimą:

```
private void AddHost(object sender, PingCompletedEventArgs e)
{
    PingReply reply = e.Reply;
    if (reply == null)
        return;
    if (reply.Status == IPStatus.Success)
    }
```

Sekantis žingsnis – kaip sudaryti nutolusių kompiuterių ar darbo stočių bendrai naudojamų aplankų sąrašą, kuriame bus vykdoma paieška?

Šiam tikslui aš pasinaudojau Robert Deeming sukurta biblioteka, kurioje svarbiausios yra dvi klasės: *ShareCollection* ir *Share*.

Klasė *ShareCollection* identifikuoja visus bendrai naudojamus aplankus priskirtų IP adresų kompiuterių, o klasė *Share* šiuos duomenis struktūrizuoja bei suformuoja paieškos sritį:

```
ShareCollection sc = new ShareCollection(reply.Address.ToString());
foreach (Share s in sc)
{
    // s.NetName parodo bendrai naudojamą aplanko vardą
    // s.ShareType parodo bendrai naudojamą aplanko tipą
}
```

Identifikavus paieškos sritį, galima pradėti vykdyti rinkmenų paiešką. Failų paieškai naudosime rekursijos metodą. Tuo pačiu principu vykdysime paiešką visuose bendrai naudojamuose aplankuose.

Kai failų kiekis yra pakankamai didelis (apie 1000 GB) ir aplankų struktūra ypač gili, susiduriama su problema – paieška gali užsitęsti gan ilgai. Kad to išvengtume, naudosime paieškos gylio apribojimus. Paieškos gylio kintamasis yra sveikasis skaičius, kurio reikšmė kiekvieną kartą padidinama vienetu kai paieška persikelia į gilesnį aplanką aplankų medyje. Kai kintamasis viršija tam tikrą vertę, paprastai tai nedidelis sveikasis skaičius (nuo 4 iki 7) rekursija sustoja. Žemiau matome rekursijos funkcijos kodo fragmentą:

```
ShareCollection sc = new ShareCollection(reply.Address.ToString());
foreach (Share share in sc)
{
    RecursiveSearch(share.ToString(), 0);
}
```

```
private void RecursiveSearch(string path,int depth)
```

```
{
    DirectoryInfo di = new DirectoryInfo(path);
    if (!di.Exists)
        return;
    depth++;
    // searchDepth yra pastovus kintamasis, aprašytas Settings klasėje
    if(depth>Settings.searchDepth)
```

```

return;
    // paieškos reikšminis žodis
string keyword=Settings.keyword;
if (di.Name.ToLower().Contains(keyword.ToLower()))
{
    AddRow(name,parent, "Folder");
}
foreach (FileInfo file in di.GetFiles())
{
    if (file.Name.ToLower().Contains(keyword.ToLower()))
    {
        AddRow(file.Name, di.FullName,size);
    }
}
foreach (DirectoryInfo subdir in di.GetDirectories())
{
    RecursiveSearch(subdir.FullName, depth);
}
}

```

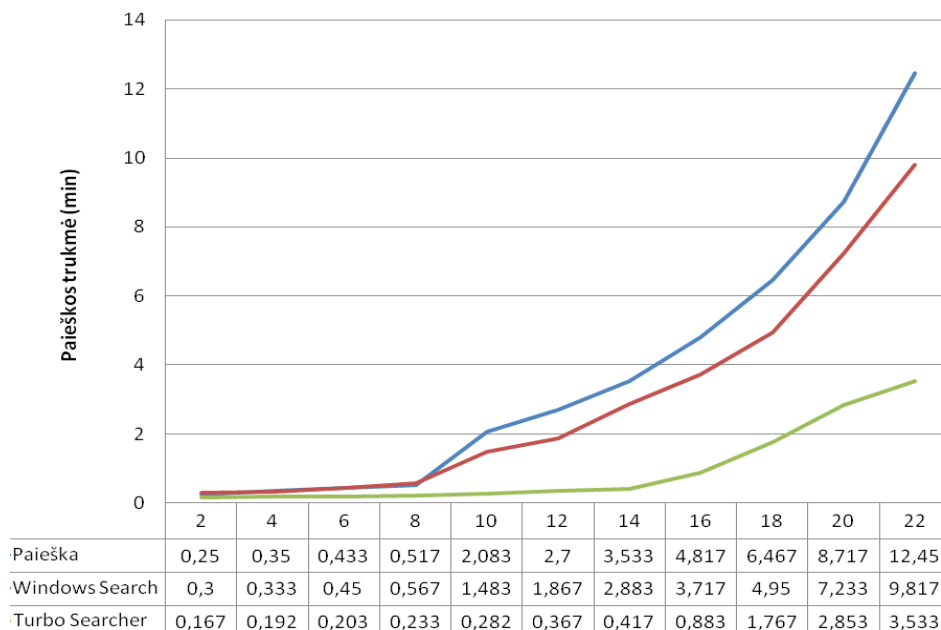
4.4. Programos „Paieška“ testavimas

Programa *Paieška* buvo testuojama aukščiau aprašytame organizacijos serveryje. Buvo testuojami skirtingų dydžių aplankai nuo 2 GB iki 22 GB talpos. Testai atliekami vykdant paveikslėlių paiešką, tekstinių dokumentų paiešką, taip pat tekstinio dokumento pagal reikšminį žodį paiešką. Analogiški testai buvo atliekami *Microsoft Windows Search* ir *Turbo Searcher* paieškos priemonėmis.

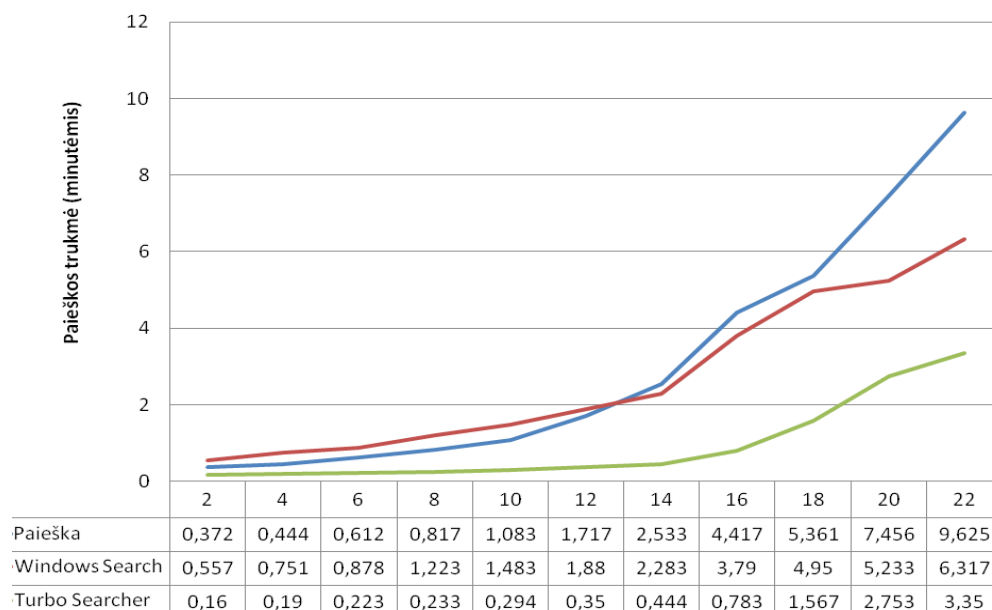
Atlikus paveikslėlių paieškos testą, su *Windows Search* paieškos įrankiu, ypač sparčia rinkmenų paieškos programa *Turbo Searcher* bei mano sukurtu paieškos įrankiu *Paieška*, matyti, kad 2-8 GB ribose paieškos rezultatai beveik vienodi visų trijų testuojamų programų. Vykdant paiešką 2GB apimties kataloge, visos trys programos vidutiniškai užtrukdavo 15 sekundžių, kas sudaro beveik 0,23 minutės. 8 GB talpos kataloge *Windows Search* ir *Paieška* paveikslėliui rasti užtrukdavo pusę minutės. Tik *Turbo Searcher* įvykdydavo paiešką per 15 sekundžių. Aplankuose, kurių dydis 10 GB ir daugiau paveikslėlių paiešką greičiausiai vykdė

Turbo Searcher įrankis, nuo jo gerokai atsiliko *Windows Search* ir *Paieška*. Kaip matome, pagal testo rezultatus, dokumentų paieška greičiausiai vyko naudojant *Turbo Searcher*. Antroje vietoje liko *Windows Search* paieškos įrankis. Nuo jo visai nedaug teatsiliko mano sukurta paieškos programa *Paieška*.

2 Lentelė. Paveikslėlių paieškos rezultatai



3 Lentelė. Tekstinių dokumentų paieškos rezultatai



IŠVADOS

Išanalizavus elektroninių dokumentų savybes, paaiškėjo, kad pastaruoju metu vis didesnė dalis elektroninės informacijos, dėl sparčiai augančios jos apimties, tampa sunkiau prieinama nei anksčiau. Elektroninė informacija kuriama, valdoma, perduodama ir saugoma naudojant dinamiškas, greitai kintančias informacinių technologijų priemones. Dėl šių priežasčių informacija prarandama arba tampa sunkiai prieinama, ir dėl to jos naudojimas tampa ribotas.

Apžvelgus populiarias rinkmenų paieškos sistemas, paaiškėjo, kad šiuolaikinės paieškos sistemos ne visuomet gali patenkinti nuolat augančius vartotojų poreikius. Tai įrodo trumpa vartotojų poreikių analizė. Vis didėjantis elektroninių dokumentų kiekis reikalauja priimti tam tikrus sprendimus: didinti arba optimizuoti organizacijos žmogiškuosius išteklius, naudoti pažangias, atitinkančias organizacijos poreikius, dokumentų valdymo bei paieškos sistemas. Vienareikšmiškai atsakyti kokia yra geriausia dokumentų paieškos sistema negalima. Nėra paieškos sistemos, kuri visiškai atitiktų kiekvieno vartotojo poreikius.

Išanalizavus elektroninių dokumentų paieškos metodus, buvo sukurta elektroninių dokumentų paieškos sistema „Paieška“, kuri buvo praktiškai pritaikyta Tauragės rajono savivaldybės administracijoje, dokumentų valdymo veikloje. Pagal užsibrėžtus reikalavimus funkcionalumui ir parinktus metodus iškeltoms problemoms spręsti sukurta elektroninių dokumentų paieškos ir analizės sistema, kaip parodė testų rezultatai, yra pakankamai konkurencinga komercinėms panašaus pobūdžio programoms.

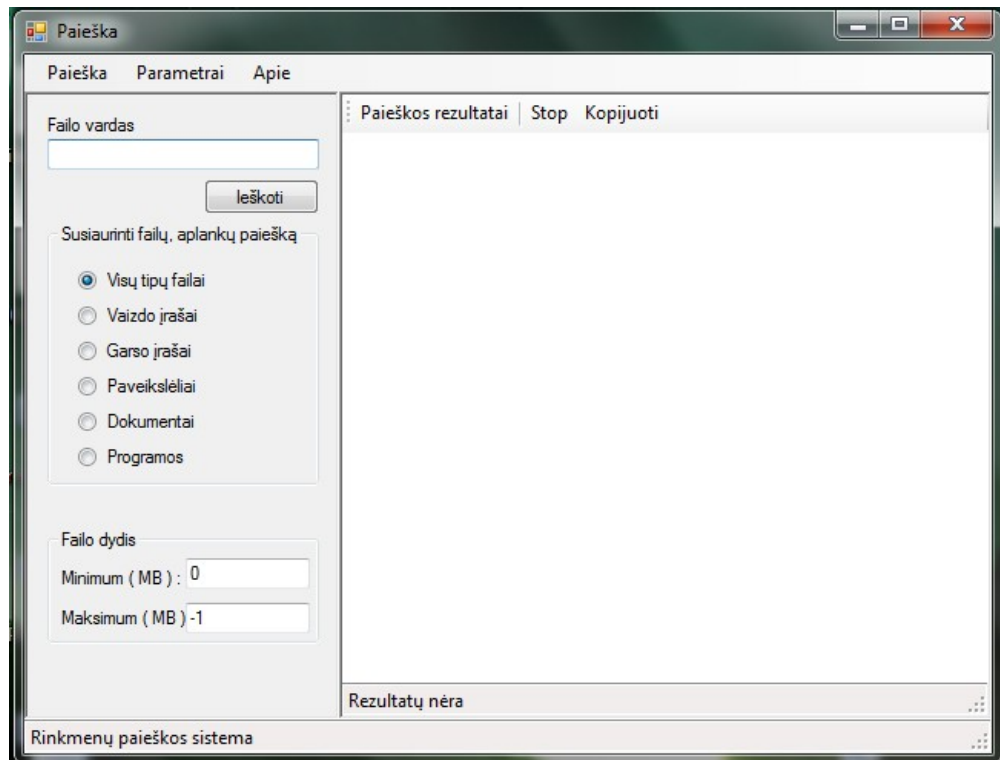
LITERATŪROS SĄRAŠAS

1. Barreau D. and Nardi B. A. Finding and reminding - File organization from the desktop. ACM SIGCHI Bulletin, 27(3):39–43, 1995.
2. Cai, M.; Frank, M.; Pan, B.; and MacGregor, R. 2004. A Subscribable Peer-to-Peer RDF Repository for Distributed Metadata Management. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* 2(2):109–130.
3. Carrier B. (2005, March 17). *File System Forensic Analysis*. Addison Wesley Professional. [interaktyvus]. [žiūrėta 2010-03-23] Prieiga per internetą: <<http://dubeiko.com/development/FileSystems/BOOKS/FileSystemAnalysis.pdf>>
4. Dumais S., Cutrell E., Cadiz J.J., Jancke G., Sarin R., Robbins D. C. Stuff I've seen: a system for personal information retrieval and re-use. *SIGIR '03*, 2003, 72–79 psl.
5. Čėsna R., Štītīlis D. Kompiuterinės informacijos ir elektroninių dokumentų apsauga viešajame administravime, LTU, 2000. 44-47 psl.
6. Elektroninių dokumentų valdymo taisyklės, patvirtintos Lietuvos archyvų departamento prie Lietuvos Respublikos Vyriausybės generalinio direktoriaus 2006 m. sausio 11 d. įsakymu Nr. V-12 (Žin., 2006, Nr. 7-268, 2006-01-28 Nr. 10)
7. LST ISO 15489-1. Informacija ir dokumentai – Dokumentų vadyba. 1 dalis: Bendrieji dalykai (tapatus ISO 15489-1: 2001). Vilnius: Lietuvos standartizacijos departamentas, 2004. 19 p.
8. Mikhailov D. *NTFS file system*. [interaktyvus]. [žiūrėta 2010-03-29] Prieiga per internetą: <<http://www.digitlife.com/articles/ntfs/>>
9. Petravičiūtė I. Elektroninių dokumentų autentiškumas: ilgalaikio išsaugojimo principai// Knygotyra. 2005, t. 45, p. 168–185.
10. Verroirovsky K, H. Rudova *Limited Assignment Number Search Algorithm* [interaktyvus]. [žiūrėta 2010-04-28] Prieiga per internetą: <<http://www.fi.muni.cz/~hanka/publ/sofsem02.pdf>>
11. Widrow, B.; Rumelhart, D. E.; Lehr, M. A. 1994. Neural networks: Applications in industry, business and science, *Comm. ACM* 37: 93–105.
12. Williams, Caroline. *Managing archives: foundations, principles and practice*. 2006. 248 p
13. Witten I. H., Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition. 2005. 83-97 psl.
14. Produkto tinklapis. [interaktyvus]. [žiūrėta 2010-04-15] Prieiga per internetą: <<http://desktop.google.com/features.html>>
15. Produkto tinklapis. [interaktyvus]. [žiūrėta 2010-04-18] Prieiga per internetą: <<http://www.voidtools.com/>>

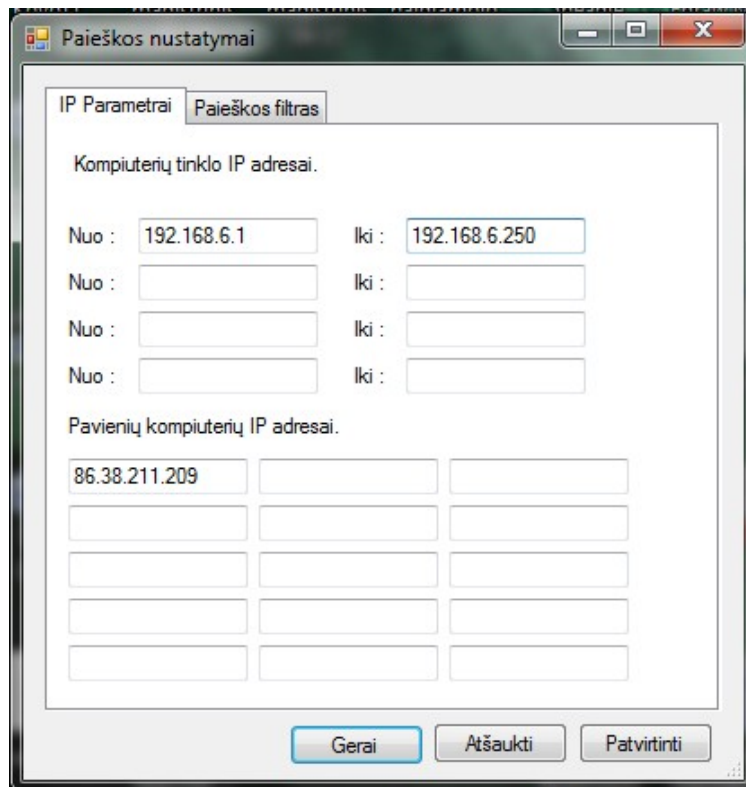
16. Produkto tinklapis. [interaktyvus]. [žiūrėta 2010-04-18] Prieiga per internetą:
<http://beagle-project.org/Main_Page>
17. Produkto tinklapis. [interaktyvus]. [žiūrėta 2010-04-18] Prieiga per internetą:
<<http://www.vandenoever.info/software/strigi/>>

1 PRIEDAS

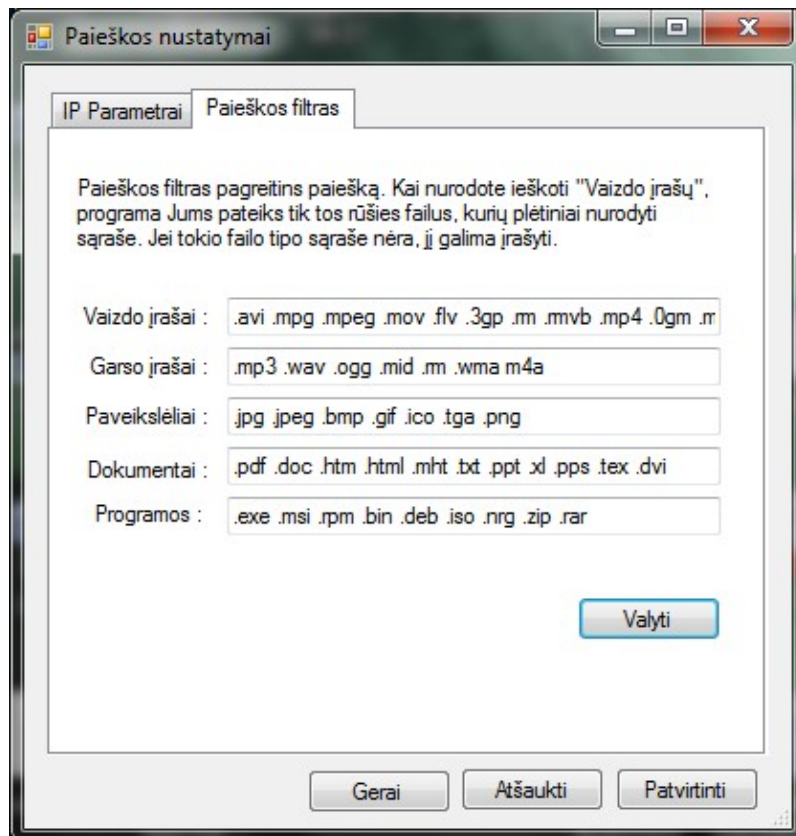
Serverio rinkmenų paieškos ir analizės sistemos *Paieška* grafinė vartotojo sąsaja



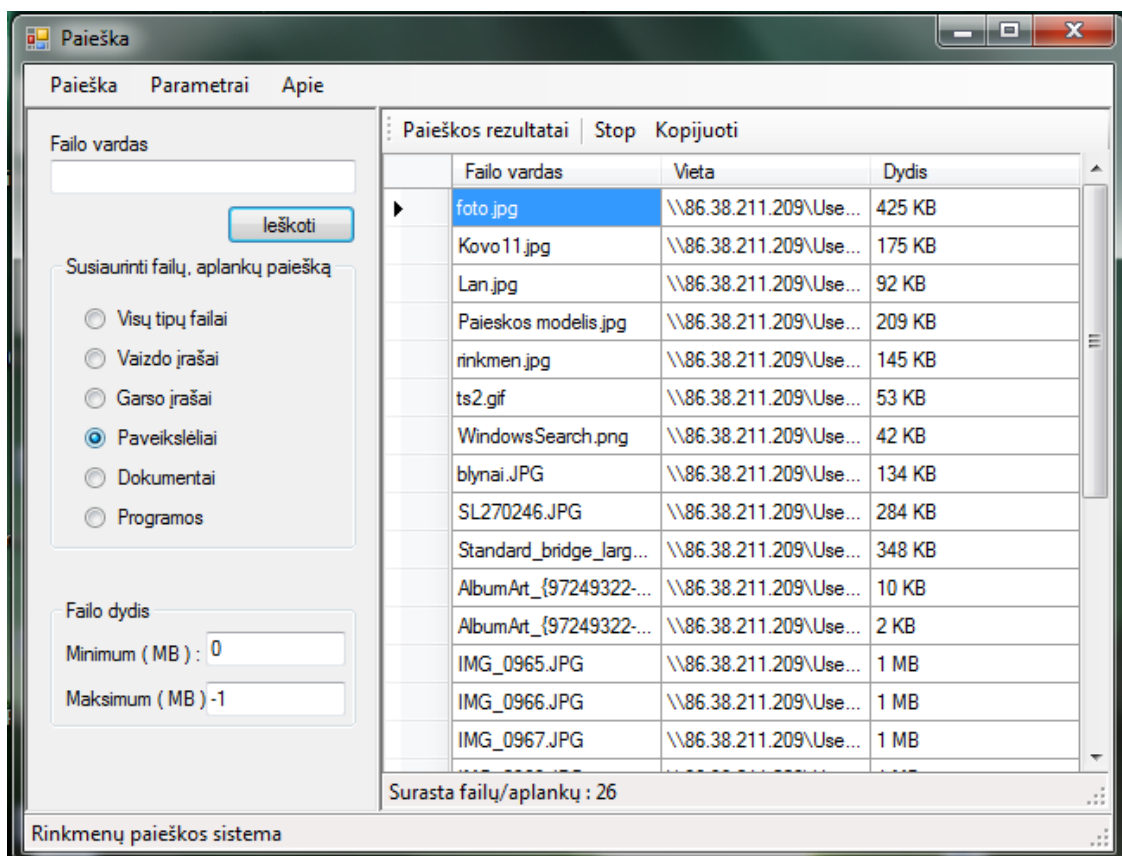
1 pav. Pagrindinis programos langas



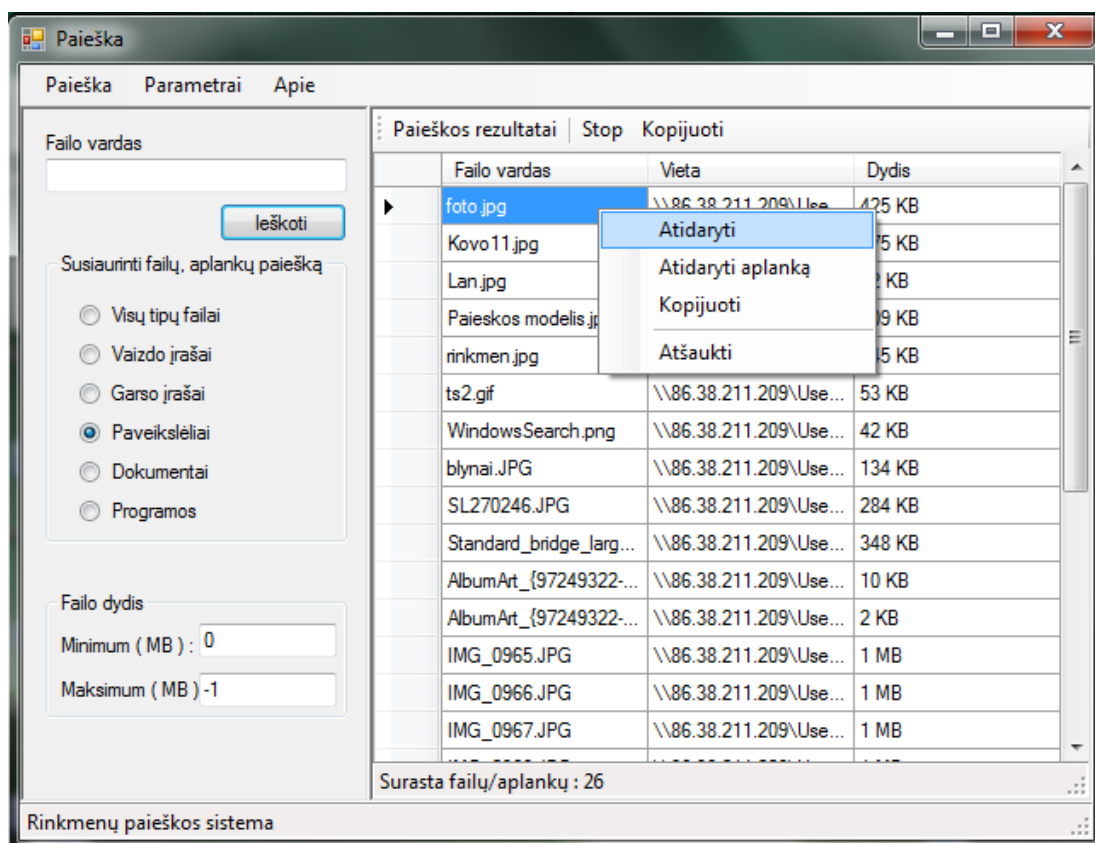
2 pav. IP adresų įvedimo kortelė



3 pav. Rinkmenų paieškos filtro kortelė



4 pav. Paveikslėlių paieškos rezultatai



5 pav. Surasto failo atvėrimo kortelė

2 PRIEDAS



TAURAGĖS RAJONO SAVIVALDYBĖS ADMINISTRACIJA

Savivaldybės biudžetinė įstaiga, Respublikos g. 2, 72255 Tauragė, tel. (8 446) 62 813, faks. (8 446) 70 801, el. p. savivald@taurage.lt. Duomenys kaupiami ir saugomi Juridinių asmenų registre, kodas 188737457

Šiaulių universitetui
Matematikos ir Informatikos fakultetui
Informatikos katedrai

2010-05-18 Nr.19- 2252

DĖL MAGISTRO DARBO TECHNOLOGINIŲ SPRENDIMŲ ĮGYVENDINIMO

Patvirtinu, kad Tauragės rajono savivaldybės administracijoje yra sėkmingai integruota ir naudojama Šiaulių universiteto Matematikos ir informatikos fakulteto informatikos specialybės magistro studijų studento Vaido Gargaso rinkmenų paieškos sistema „Paieška“. Šioje programoje realizuoti technologiniai sprendimai padeda efektyviau valdyti organizacijos elektroninius išteklius, operatyviau surasti duomenis įstaigos vietiniame tinkle.

Administracijos direktorius

Algirdas Mosėjus

Informatikos skyriaus vedėjas
Mindaugas Maraulis, (8 446) 61 872,
el. p. it@taurage.lt