

საქართველოს ტექნიკური უნივერსიტეტი

ხელნაწერის უფლებით

ლაშა გავაშელაშვილი

საძიებო სისტემაში ინფორმაციის ოპტიმალურად ძებნის მეთოდის
შემუშავება

დოქტორის აკადემიური ხარისხის მოსაპოვებლად

წარდგენილი დისერტაციის

ა ვ ტ ო რ ე ფ ე რ ა ტ ი

სადოქტორო პროგრამა „ინფორმატიკა“ შიფრი 0401

თბილისი

2015 წელი

სამუშაო შესრულებულია საქართველოს ტექნიკური უნივერსიტეტის ინფორმატიკისა და მართვის სისტემების ფაკულტეტის, კომპიუტერული ინჟინერიის დეპარტამენტის, კომპიუტერული სისტემებისა და ქსელების მიმართულებაზე

სამეცნიერო ხელმძღვანელები: პროფესორი გურამ ჩაჩანიძე
პროფესორი რომან სამხარაძე

რეცენზენტები:

დისერტაციის დაცვა შედგება 2015 წლის -----საათზე საქართველოს ტექნიკური უნივერსიტეტის ინფორმატიკისა და მართვის სისტემების ფაკულტეტის სადისერტაციო საბჭოს კოლეგიის სხდომაზე, კორპუსი -----, აუდიტორია -----.

მისამართი: 0175, თბილისი, მ. კოსტავას 77

დისერტაციის გაცნობა შესაძლებელია საქართველოს ტექნიკური უნივერსიტეტის ბიბლიოთეკაში (0175, თბილისი, მ. კოსტავას 77), ხოლო ავტორეფერატისა სტუ-ს ვებ-გვერდზე.

სადისერტაციო საბჭოს სწავლული მდივანი

სრული პროფესორი

თ. კაიშაური

შესავალი

აქტუალურობა. როგორც ცნობილია თანამედროვე კომპიუტერულ სისტემებში პროგრამების უდიდესი ნაწილი კავშირისთვის იყენებს ინტერნეტს. ინტერნეტში ინფორმაციის მოცულობა დღითიდღე იზრდება და შესაბამისად აქტუალური ხდება ინფორმაციის მოძიების ეფექტური ალგორითმების შემუშავება.

გლობალური ქსელის განვითარების შესაბამისად და მასში არსებული დოკუმენტების რაოდენობის ზრდის გამო წარმოიშვება ინტერნეტ ძიების პრობლემა. დღევანდელი მდგომარეობით ეს პრობლემა წყდება ინტერნეტ საძიებო სიტემების დახმარებით, რომლებიც განსაზღვრავენ იმ დოკუმენტების(ვებ-გვერდების) სიას, რომლებიც შეესაბამებიან მომხმარებლების მოთხოვნებს. ასეთი დოკუმენტები შეიძლება იყოს მილიონობით და საძიებო სისტემის ამოცანას წარმოადგენს მიაწოდოს მომხმარებელს მოპოვებული გვერდების სია, სადაც ისინი დალაგებულნი იქნებიან რელევანტურობის კლების მიხედვით.

რელევანტურობა განსაზღვრავს როგორც დოკუმენტის ძირეულ თვისებებს ასევე მომხმარებლის მოთხოვნას, რასაც ეწოდება რანჟირების ფაქტორები. მათმა რიცხვმა შეიძლება მიაღწიოს რამდენიმე ასეულს. თითოეული ფაქტორი შედის რანჟირების ფორმულაში, რომელიც განსაზღვრავს წარსადგენი ვებ-გვერდების სიას მომხმარებლის მოთხოვნის შესაბამისად.

მაგალითად, ეს ფაქტორები შეიძლება იყოს, მოთხოვნის ერთეული სიტყვები, რომლებიც გვხვდება დოკუმენტის სათაურში ან ტექსტში, მომხმარებლის გეოგრაფიული მდებარეობა(მაგალითად მომხმარებლის მოთხოვნაზე პასუხად შეიძლება მივიღოთ განსხვავებული სიები თუ ძიება ხდება ვთქვათ გერმანიიდან და სხვა სიები-საქართველოდან.) და სხვა.

ინფორმაციის ზრდის კვალდაკვალ ხდება საძიებო სისტემების ალგორითმების უწყვეტი განვითარება. უკანასკნელი წლის განმავლობაში შეიმჩნევა მრავალი სიახლე, მაგალითად: საძიებო სისტემებმა ისწავლეს განსაზღვრონ დოკუმენტის პირველწყარო(რომ გამოირიცხოს დუბლიკატების არსებობა), რეალიზებულია ძიება მომხმარებლის რეგიონის შესაბამისად და ა.შ. ყოველივე ეს ფაქტორები ამაღლებენ საძიებო სისტემის ხარისხს და საიმედოობას მათ შორის სანდოობას მათი საგანმანათლებლო დანიშნულების მხრივ.

ზემოთ თქმულიდან გამომდინარე, მეტად აქტუალურია აღნიშნული პრობლემის გადაწყვეტის გზებისა და საშუალებების შემუშავება.

კვლევის მიზანი. კვლევის მიზანს წარმოადგენს ინტერნეტ ძიების ისეთი მეთოდის შემუშავება, რომელიც იქნება მაქსიმალურად სასარგებლო და სწრაფი და ეფექტური

კვლევის ობიექტი. კვლევის ობიექტს თანამედროვე წარმოადგენს საძიებო სისტემები და მათ მიერ გამოყენებული ალგორითმები

კვლევის საგანი. კვლევის საგანს წარმოადგენს „ინფორმაციის ძიების პროცესი“

მეცნიერული სიახლე მდგომარეობს საძიებო სისტემის ოპტიმიზების მიზნით ძიების ახალი ალგორითმის შემუშავებაში. შემოთავაზებული ალგორითმი ინტეგრირდება მომხმარებლის ინტერნეტ ბრაუზერში, შეარულებს მის მიერ გამოყენებული ვებ-რესურსების გამოყენების სიხშირის გამოთვლას და ყოველივე ამის საფუძველზე მოახდენს ძიების პროცესის ოპტიმიზებას ისე, რომ კონკრეტულ შემთხვევაში მომხმარებელს წარედგინება მის მიერ გამოყენებული ვებ-გვერდების რეიტინგული სია. სიის სათავეში განლაგებული იქნება ის ვებ-გვერდები რომელიც მან ყველაზე მეტი სიხშირით გამოიყენა. ასეთი მოდგომა მკვეთრად ამაღლებს მოძებნილი ინფორმაციის სარგებლიანობას

კვლევის მეთოდები. ინტერნეტ ძიების პროცესის ოპტიმიზაციის კონცეფციის დამუშავების მიზნით დასმული ამოცანების გადაწყვეტაში მეთოდური მიდგომა გულისხმობს ინტერნეტ ძიების ალგორითმების მახასიათებლების ზოგადის თეორიის განსაზღვრას. ოპტიმიზაციის მეთოდების, სისტემური მიდგომისა და ანალიზის ოპტიმალური მართვის თეორიების გამოყენებას. განხილულია Fillipo Menczer, Gautam Pant, Nicholas Xiaoguang Qi, Brian D. Davison, Mr.Ramesh Prajapati , S V Kashmir Raja², Felix K Akorli³ and Jawahar M⁴ ნაშრომები.

პრაქტიკული ღირებულება. პრაქტიკულ ღირებულებას წარმოადგენს ის რომ, მივიღებთ ინტერნეტ ძიების მაღალ ეფექტურობას, ძიებაზე შედეგის დაბრუნების დროულობას და საიმედოობას. ამგვარი ოპტიმიზაციით მიიღწევა ვებ-რესურსების გაუმჯობესებული სიის წარდგენა მიზნობრივი ძიებისას.

ნაშრომის აპრობაცია. სადისერტაციო თემის გარკვეული ნაწილი მოხსენებულია საერთაშორისო სამეცნიერო კონფერენციებზე:

1. ქუთაისის სასწავლო უნივერსიტეტი ლამპარი. პროფესორ მასწავლებელთა III რესპუბლიკური სამეცნიერო კონფერენცია. თანამედროვე მეცნიერების აქტუალური საკითხები. შრომები 2010. „ინფორმაციული საზოგადოება და გეოინფორმაციული სისტემები“ მანანა ქარქაშაძე, ლაშა გავაშელაშვილი.
2. აკაკი წერეთლის სახელმწიფო უნივერსიტეტი. IV საერთაშორისო სამეცნიერო-მეთოდური კონფერენცია. სწავლებისა და აღზრდის აქტუალური პრობლემები. შრომები. ქუთაისი 2013. „ინფორმაციის ძიების ოპტიმალური მეთოდების შემუშავების საკითხები“ ლაშა გავაშელაშვილი.
3. პედაგოგთა კვალიფიკაციის ამაღლების ქუთაისის ზონალური ინსტიტუტი. I საერთაშორისო სამეცნიერო-მეთოდური კონფერენცია. განათლებისა და მეცნიერების მაღალი

- ინტელექტუალური ტექნოლოგიები. შრომები. ქუთაისი 2013. „ინტერნეტში ძიების თანამედროვე საშუალებების ანალიზი“ ლაშა გავაშელაშვილი.
4. პედაგოგთა კვალიფიკაციის ამაღლების ქუთაისის ზონალური ინსტიტუტი. რესპუბლიკური სამეცნიერო მეთოდური კონფერენცია II. ინოვაციები და ტრადიციები თანამედროვე განათლებაში. შრომები 2014. „ინტერნეტ ძიების ალგორითმები“ ლაშა გავაშელაშვილი.
 5. მეცნიერებისა და საზოგადოების განვითარების ფონდი „ინტელექტი“ III საერთაშორისო სამეცნიერო კონფერენცია. მოხსენებათა თეზისები. კომპიუტინგი, განათლების მეცნიერებები, მასწავლებლის სწავლება. ბათუმი 2014. „ვებ-გვერდების კლასიფიკაცია“ ლაშა გავაშელაშვილი.
 6. საერთაშორისო სამეცნიერო ჟურნალი ინტელექტი 1(48). 2014წ. „ეფექტური ინტერნეტ ძიების ალგორითმები(PageRank,MartxNET) მიმოხილვა“. ლაშა გავაშელაშვილი.
 7. მეცნიერებისა და საზოგადოების განვითარების ფონდი „ინტელექტი“ III საერთაშორისო სამეცნიერო კონფერენცია. მოხსენებათა თეზისები. კომპიუტინგი, განათლების მეცნიერებები, მასწავლებლის სწავლება. ბათუმი 2014. ინტერნეტ ძიების ალგორითმი და Google-ს არქიტექტურა ლაშა გავაშელაშვილი.
 8. მეცნიერებისა და საზოგადოების განვითარების ფონდი „ინტელექტი“ III საერთაშორისო სამეცნიერო კონფერენცია. მოხსენებათა შრომები. კომპიუტინგი, განათლების მეცნიერებები, მასწავლებლის სწავლება. თბილისი 2015. „ვებ-გვერდების კლასიფიკაცია“ ლაშა გავაშელაშვილი.
 9. მეცნიერებისა და საზოგადოების განვითარების ფონდი „ინტელექტი“ III საერთაშორისო სამეცნიერო კონფერენცია. მოხსენებათა შრომები. კომპიუტინგი, განათლების მეცნიერებები,

მასწავლებლის სწავლება. თბილისი 2015. ინტერნეტ ძიების ალგორითმი და Google-ს“ არქიტექტურა ლაშა გავაშელაშვილი, გურამ ჩაჩანიძე, რომან სამხარაძე.

პუბლიკაციები. დისერტაციის ირგვლის გამოქვეყნებულია 9 ნაშრომი 7 სამეცნიერო სტატიას და 2 მოხსენებათა თეზისები. ნაშრომის ძირითადი დებულებები მოხსენებულია სამეცნიერო კონფერენციებზე(იხ. შრომების სია).

დისერტაციის სტრუქტურა და მოცულობა. სადისერტაციო ნაშრომი შედგება შესავლის, სამი თავის, დასკვნების, გამოყენებული ლიტერატურის სიისა და დანართებისაგან. იგი მოიცავს კომპიუტერულ ნაბეჭდ 111 გვერდს, 24 ნახაზს, 1 სურათს, 3 დანართს და ლიტერატურის ჩამონათვალს 50 დასახელებით.

ნაშრომის შინაარსი

შესავალში დასაბუთებულია პრობლემის აქტუალურობა და ნაჩვენებია მეცნიერული კვლევების გააქტიურების აუცილებლობა აღნიშნული პრობლემების გადასაჭრელად.

დისერტაციის პირველ თავში ჩატარებულია ინტერნეტ ძიების პროცესების მიმოხილვა და ანალიზი. განზოგადებულია ინფორმაციულ საზოგადოებაზე გადასვლის აუცილებლობა. ნაჩვენებია ინტერნეტ-ძიების მუშაობის პრინციპი. დასაბუთებულია, რომ დღეისათვის არსებული მეთოდები აღარ არის ეფექტური, ძიების პროცესში გამოვლენილი პრობლემების გადასაწყვეტად. დასაბუთებულია ინტერნეტ-ძიების ახალი მეთოდების გამოყენების აუცილებლობა. ჩამოყალიბებულია მათი

გამოყენების აუცილებლობა.

განხილულია კლასიკური ძიების არსებული მეთოდი, კერძოდ Seo-ს ოპტიმიზაციის მაგალითზე, გამოვლენილია ნაკლოვანებები და დადებითი მხარეები. ჩატარებულია წესებზე დაფუძნებული ძიების პრინციპების მიმოხილვა და კრიტიკული ანალიზი. დასაბუთებულია, რომ არსებული მეთოდები არ პასუხობენ ძიების მაღალ ხარისხს და შესაბამისად მეტად მნიშვნელოვანია ამ პრობლემის გადასაწყვეტი გზების ძიება და შესაბამისი მოდელებისა და ალგორითმების შემუშავება.

პირველ პარაგრაფი-ინტერნეტის შექმნის მოკლე ისტორია, გადმოგვცემს თუ როგორ იქმნებოდა ინტერნეტი და რა სახით არის ის დღეისათვის წამოდგენილი.

მეორე პარაგრაფი-სადიებო სისტემების შექმნის ასპექტები, წარმოდგენილია სადიებო სისტემების შექმნის მოკლე ისტორია, და მისი მუშაობის პრინციპები. დღეისათვის არსებული მდგომარეობა.

მესამე პარაგრაფი- მოცემულია სადიებო სისტემის ოპტიმიზაციის მეთოდები.

- თეთრი ოპტიმიზაცია – ოპტიმიზირებული სამუშაო რესურსებთან, რომლებიც ოფიციალურად აკრძალულია ყოველი სადიებო სისტემის მიერ – საიტის ალგორითმებზე ზეგავლენის მოხდენის გარეშე.ეს მოიცავს საიტთან მუშაობას, მის შიგა ნავიგაციასთან და შინაარსთან, ასევე გარე სამუშაოებს, საიტის ოპტიმიზირება გამოკითხვების მეშვეობით, პრეს – რელიზების, სოციალურ ქსელებში რეგისტრირების, პარტნიორულ პროგრამებსა და ა.შ.
- ნაცრისფერი ოპტიმიზაცია. სადიებო სისტემის ოპტიმიზაცია, მოიცავს საკვანძო სიტყვების დიდი რაოდენობით ჩასმას საიტის ტექსტში, ხშირად ადამიანებისთვის ყველაზე გამოსაჩენ ადგილებში. ამასთან, ხშირად გამოიყენება HTML – ის სათაურის ტეგში, h1, საკვანძო სიტყვების ატრიბუტი.

- შავ ოპტიმიზაციაში შედის მეთოდები, რომლებიც საძიებო სისტემას ეწინააღმდეგებიან. მათ შორის შემდეგის გამოყოფა შეიძლება:
 - “Doorway” - ს (გვერდები და რესურსები, სპეციალურად შექმნილი საძიებო სისტემების სამუშაოდ, ხშირად დიდი რაოდენობის საკვანძო სიტყვები საიტზე) გამოყენება.
 - “Cloaking” (მომხმარებელს მიეცემა ერთი გვერდი, საძიებო სისტემას – მეორე, ოპტიმიზირებული გარკვეული მოთხოვნით) გამოყენება.
 - საიტის დაფარული ტექსტის გამოყენება.
 - „ერთ პიქსელიანი“ ლინკის გამოყენება.

არსებობს ნორმალური ინტერნეტ ძიებისათვის რამდენიმე ხელისშემშლელი ფაქტორი

- დიდი მოცულობა.
- ძალიან სწრაფი ცვლილებები.
- დინამიური გვერდის გენერირება.

Web Crawler შეიძლება მრავალი სახეობის იყოს:

- ავტომატური ინდექსირები;
- ვებ-ობობები;
- ვებ-რობოტები;
- ბოტები;
- სკუტერები და ა.შ.

“მცოცავები” კოპირებას უკეთებენ ყველა მის მიერ მოძებნილ ვებ-გვერდს, რადგან იმ გვერდზე შეიძლება ყველანაირი ინფორმაცია იყოს, რაც მომხმარებელს შეიძლება დასჭირდეს. არავინ იცის კონკრეტულად რისი მოძებნა სჭირდება ამა თუ იმ მომხმარებელს, ასე რომ Crawler „თავს იზღვევს“. ის საფუძვლიანად იკვლევს საიტის ყველა გვერდის შიგთავსს და იმახსოვრებს ყველა ბმულს/URL-ს.

Spamdexing (ასევე ცნობილი, როგორც საძიებო სისტემის სპამერი, ძებნის სპამი, ვებ სპამი და ა.შ) არის იმ მეთოდების ნაკრები, რომელთა მიზანი ერთია, არალეგალური გზით, წარმატებული ოპტიმიზაციის მიღწევა. არალეგალურ გზაში იგულისხმება: არასაკურო/არასწორი ფრაზების გამოყენება.

მეთოდე პარაგრაფი- წარმოდგენილია ძიების სტრუქტურა Google-ს მაგალითზე. გუგლის საძიებო სისტემას ინტერნეტში ძიებისათვის გააჩნია სამი სახის საშუალება:

- გუგლბოტი (googlebot).
- ინდექსატორი (indexer).
- შეკითხვის, მოთხოვნის პროცესორი (query processor).

გუგლბოტი წარმოადგენს რობოტს, რომელიც მოგზაურობს ინტერნეტ სივრცეში, პოულობს ვებ-გვერდებს და გადასცემს მათ ინდექსატორს შემდგომი დამუშავებისათვის. ეს შეიძლება შევადაროთ კიბერსივრცეში „გახლართულ“ ობობას. გუგლბოტი შედგება უამრავი კომპიუტერისაგან. მას უმოკლეს დროში შეუძლია ათასობით განსხვავებული ვებ-გვერდის ერთდროული ჩატვირთვა.

გუგლბოტი ვებ-გვერდებს ორი განსხვავებული მეთოდით პოულობს:

- url დამატების ფორმა ([www.google.com /addurl.html](http://www.google.com/addurl.html)).
- გვერდების ჩამოცოცებით.

ინდექსი არის მონაცემთა ბაზა, რომელშიც ინფორმაცია ინახება ალფავიტის მიხედვით. იგი შეიცავს ინფორმაციას სიტყვებისა და მათი ადგილმდებარეობის შესახებ. მონაცემთა სტრუქტურა იძლევა საშუალებას, მივიღოთ სწრაფი დაშვება დოკუმენტებზე, რომელიც შეესაბამება ჩვენს მოთხოვნას. ძებნის გაუმჯობესების მიზნით, ინდექსატორი იგნორირებას უკეთებს მაკავშირებელ სიტყვებს როგორცაა („ან“, „და“) ერთ ასობგერას,

ერთ ციფრს.

შეკითხვის მოთხოვნის პროცესორი შედგება რამდენიმე ნაწილისაგან. მის ინტერფეისს წარმოადგენს „ძებნის ყუთი“, „ძრავი“, რომელიც მოთხოვნებს შეუსაბამებს სათანადო დოკუმენტაციას და შედეგების ფორმატირების საშუალება.

გვერდის ხარისხი (Pagerank) არის გუგლის სიტემა, რომლის დანიშნულებაც ვებ-გვერდების შეფასება. ფასდება ვებ-გვერდების ხარისხიანობა. გვერდი, რომელსაც მაღალი PageRank გააჩნია, ითვლება, რომ უფრო მოთხოვნადია.

დისერტაციის მეორე თავში, აგებულია ინტერნეტ ძიების ეფექტური მოდელი. განხილულია მომხმარებლის ხარისხის მახასიათებლების კონტროლის მეთოდები. დადგენილია, რომ აღნიშნული მეთოდები ოპტიმალური ძიების პროცესში არიან შეზღუდულნი. გამოვლენილია დღევანდელი გლობალური ქსელების პრობლემები და შემუშავებულია ამ პრობლემების გადაწყვეტის მეთოდები. გაანალიზებულია თითოეული მათგანი.

პირველი პარაგრაფი- წარმოდგენილია ინტერნეტ ძიების არქიტექტურა გრაფიკული სახით და ჩამოყალიბებულია მისი მუშაობის არსი.

მეორე პარაგრაფი- წარმოდგენილია შეზღუდვები რომლებიც წარმოიქმნება ძიებისა და ალგორითმების შედარება. Crawler-ები მოიხმარენ შემდეგ რესურსებს:

- ✓ ქსელის სიჩქარე გვერდის ჩამოსატვირთად.
- ✓ მეხსიერება, რომ შეინარჩუნოს მონაცემთა სტრუქტურები.
- ✓ საკუთარი ალგორითმი.
- ✓ CPU.
- ✓ დისკის მოცულობა.

ცხადია, რაც უფრო რთულია ალგორითმი ის მით უფრო მეტ რესურსს მოითხოვს. იმისათვის რომ მოხდეს სამართლიანი შედარება ძიების ალგორითმებისა, არსებობს შემდეგი ორი გზა:

1. თვალყური ვადევნოთ CPU-ს, დრო რომელიც დაჭირდება თითოეულ crawler თითოეული გვერდისათვის და მოვახდინოთ იგნორირება მოდულის ჩამოტვირთვის დროზე, რომელიც საერთოა ყველა crawler-ისათვის. დროის ნიშნული საჭიროა მხოლოდ crawler სპეციფიკურ ოპერაციებში. მონიტორინგი CPU-ზე და მის მიერ გამოყენებული დრო საჭიროა სხვადასხვა სირთულის ძიების ალგორითმის შედარებისათვის.
2. შევუზღუდოთ მეხსიერება, თითოეულ crawler და მივცეთ ბუფერის განსაზღვრული ზომა. ბუფერი გამოყენებულ იქნება crawler-ის მიერ დროებითი ბმულებისათვის. თითოეულ crawler-საქვს უფლება აკონტროლოს მისი MAX_BUFFER ბმულები. თუ ბუფერი შეივსება მაშინ crawler-ს შეუძლია გადაწყვიტოს რომელი ბმული შეცვალოს და მის ნაცვლად ჩასვას ახალი. ბუფერის ზომა შეიძლება სხვადასხვანაირი 2^8 -დან 2^{11} -მდე, რომ შეფასდეს, რამდენად იყენებს მეხსიერების რესურსს ძიების ალგორითმი.

მესამე პარაგრაფი- წარმოდგენილია ძიების ალგორითმები. მათი მუშაობის პრინციპები და მაგალითები.

მეოთხე პარაგრაფი- წარმოდგენილია რანჟირების ალგორითმების კვლევა. ინტერნეტში განთავსებული ინფორმაცია სწრაფად იცვლება: ყოველწამიერად ჩნდება და ქრება ინტერნეტრესურსები.

მიუხედავად ამისა, არსებობს რამდენიმე მნიშვნელოვანი რამ, რაც ინტერნეტში განთავსებულ ვებგვერდებს აერთიანებს: ისინი დაწერილია ერთსა და იმავე ენაზე ან მასთან დაკავშირებულ „დიალექტზე“ და ერთმანეთთან ჰიპერბმულების საშუალებით არის დაკავშირებული. ჰიპერბმულების ერთ-ერთი მნიშვნელოვანი ნაწილია მისამართი,

რომლითაც იგი სხვა რესურსზე მიუთითებს. ეს მისამართები მომხმარებლისთვის უმეტესწილად უხილავია. მომხმარებელი ერთი რესურსიდან მეორეზე მხოლოდ ჰიპერბმულის შესაბამის ვიზუალურ კომპონენტზე (მაგ., ტექსტი, დილაკი, ნახატი) დაჭერით გადადის.

ყოველი რანჟირების ალგორითმი ეფუძნება ბმულების ანალიზს. ისინი შეიძლება დაიყოს ორ ქვეკატეგორიად

1. კითხვაზე დამოკიდებული.
2. კითხვაზე დამოუკიდებელი.

ვებ-გვერდების რანჟირება ნიშნავს მათ დალაგებას იმის მიხედვით, თუ რომელ მათგანში მოხვედრის ალბათობა არის უფრო მეტი ინტერნეტში ნავიგაციის პროცესის უსასრულოდ გაგრძელების შედეგად.

მეხუთე პარაგრაფი- წარმოდგენილია კლასიფიკაციის დეტალები. კლასიფიკაცია მნიშვნელოვან როლს ასრულებს ბევრი ინფორმაციის მოძიებისა და მართვის საქმეში. ეს მნიშვნელოვანია, რადგან მოხდეს ორიენტირებული ძიება და დაეხმაროს ვებ-დირექტორიებს განვითარებაში. მოხდეს თემატური ბმულის ანალიზი და ანალიზის აქტუალური სტრუქტურის ჩამოყალიბება ქსელში. ვებ-გვერდების კლასიფიკაცია ასევე აუმჯობესებს ინტერნეტ ძიების ხარისხს.

ვებ-გვერდების კლასიფიკაციის დროს, როგორც წესი ხდება ვებ-გვერდის გამოყენების სპეციფიკის დეტალური განხილვა. ყურადღება ეთმობა შემდეგ დეტალებს.

- მოხდეს დეტალური განხილვა თუ როგორ უნდა გამოისახოს სასარგებლო ვებ-გვერდი კლასიფიკაციისას.
- გაკეთდეს სია იმ აპლიკაციების რომლების მოახდენენ კლასიფიკაციას.
- განხილულ იქნას სამომავლო კვლევები და მიმართულებები.

მეექვსე პარაგრაფი- ძიების სტრუქტურული მიმოხილვა. ინტერნეტში ძიება ეფუძნება შემდეგი ამოცანების გადაჭრას:

- რესურსების მოძიება.
- ინფორმაციის დახარისხება.
- გენერალიზაცია.
- ანალიზი.

ძიების პროცესი იყოფა სამ ნაწილად ესენია:

- შინაარსის მოპოვება.
 - გამოყენების სფერო.
 - სტრუქტურის მოპოვება.
- ✓ **შინაარსის მოპოვება-** ეს არის ტექნიკა, რომელიც იყენებს ფაქტიურ ინფორმაციას, რომელიც ინახება ინტერნეტში და ახდენს მისი შაბლონის ძიებას. იგი თავის თავში ახდენს ტექსტის ანალიზს, რომელიც ინახება ფაილებში.
- ✓ **გამოყენების სფერო-** აქცენტი გაკეთებულია იმაზე, თუ როგორ იყენებენ მომხმარებლები ინტერნეტს. ის იყენებს დაშვების ჟურნალს ვებ-სერვერზე, ცდილობს აღმოაჩინოს ნიმუშები, რომლებიც იქმნება მომხმარებელთა მიერ.
- ✓ **სტრუქტურის მოპოვება-** გულისხმობს იპოვოს გზა, რომლის დახმარებითაც ვებ-გვერდები უკავშირდება ერთმანეთს .
- მას შემდეგ რაც ყველა შემადგენელი ობიექტი მოიპოვება, შეიძლება გვერდი თავისუფლად მივაკუთნოთ იმ გვერდების სიას რომელსაც გააჩნია ატრიბუტები. უნდა აღინიშნოს, რომ ვებ-კლასიფიკაცია არ არის მხოლოდ კონკრეტული ატრიბუტების ძიება, მას გააჩნია სირთულეები და მნიშვნელოვანი განსხვავებები.

დისერტაციის მესამე თავში შემუშავებულია ალგორითმის კონცეფცია, რომელსაც საფუძვლად უდევს ძირითადი მოთხოვნა: ალგორითმებს უნდა შეეძლოთ ძიების ხარისხის ამაღლება და ხასიათდებოდეს ძიების

მაღალეფექტურობით!

შემუშავებული ალგორითმი იყენებს ძიების ინტელექტუალურ მეთოდს, რომელიც ახდენს საკუთარი თავის „სწავლებას“ და ეს დინამური პროცესია. მას შეუძლია კონკრეტული მომხმარებლისათვის ძიების შედეგის შეთავაზება. ის აუმჯობესებს უკვე არსებულ ძიების საშუალებებს, ვებ-რესურსის გამოყენების სიხშირის კონტროლით. განხილულია შემუშავებული ალგორითმის ბრაუზერებსა და ოპერაციულ სისტემებზე ინტეგრირება და მისი მუშაობის პრინციპის ანალიზი. დასაბუთებულია ძიების ალგორითმის შესაძლებლობები და მათი გამოყენების ასპექტები ძიების ფაზაში.

გაანალიზებულია ძიების ხარისხის მახასიათებლების კონტროლის ალგორითმების მუშაობის შედეგები.

პირველი პარაგრაფი- წარმოდგენილია ინტერნეტ ძიების ხერხები. როგორც ჩვენთვის ცნობილია, არსებობს ინტერნეტში ძიების გარკვეული წესები, რომლებსაც უნდა დავემორჩილოთ, რომ მივიღოთ შესაბამისი შედეგი, მოთხოვნილ რესურსზე. ესენია:

- “+/-“ - მაგალითად: თუ ჩავწერთ ასეთ რამეს „ხელოვნური ინტელექტი+ რობოტები-ფილმი“. ეს ნიშნავს იმას, რომ მოიძებნება რესურსები სადაც აუცილებელად უნდა იყოს „ხელოვნური ინტელექტი“ და „რობოტები“ და არ უნდა იყოს „ფილმი“.
- “|“ - მაგალითად: თუ ვეძებთ ინფორმაციას ვებ-გვერდებზე, სადაც გვინდა შეგვხვდეს ერთ-ერთი სიტვა მაინც. მაგალითად: „ხელოვნური ინტელექტი|რობოტი“
- „-“ ზუსტი ფრაზის ძიება. მაგალითად: „ხელოვნური ინტელექტი“. მივიღებთ იმ რესურსებს, სადაც ზუსტად ეს ფრაზა გვხვდება.
- „!“-სიტყვის სხვადასხვა ფორმით ძიება. ეს ნიშნავს იმას, რომ მაგალითად: შეიძლება ერთ ვებ-გვერდზე იყოს სიტყვა „ინტელექტი“, მეორეზე „ინტელექტის“, „ინტელექტმა“ და ა.შ. შესაბამისად, რომ მოხდეს მისი გარჩევა ძიებისას, გამოვიყენებთ შემდეგ ფორმას

„ინტელექტი“ და მივიღებთ ყველა იმ ვებ-რესურსს, სადაც წარმოდგენილი არის სიტყვა „ინტელექტი“ მისი სხვადასხვა ბრუნვით.

- „/n“- როდესაც გვსურს ვიპოვოთ ფრაზის ყველა განმარტება, გამოიყენება სიმბოლო „/“ და ციფრი რომელიც გამოხატავს ფრაზის მნიშვნელობათა რაოდენობას.
- „#url“-მისამართის ძიება, მაგალითად: „#url=www.“
- „#link“- ბმულის ძიება, მაგალითად: „#link= www.file.ge“ აჩვენებს ყველა ვებ-გვერდს, სადაც არის ლინკი მოცემულ ვებ-გვერდზე.

მეორე პარაგრაფი- მოცემულია პრობლემა, რომელიც ხელს უშლის ეფექტური ძიების განხორციელებას ინტერნეტ სივრცეში. რადგან მსოფლიოში არსებობს სხვადასხვა ენა და დამწერლობა, არსებობს ინფორმაციის აღქმის სხვადასხვაობა. გერმანელი როგორც ეძებს ინფორმაციას, ისე არ მოძებნის ქართული და ქართული როგორც ეძებს, ისე არ მოძებნის ინგლისელი და ა.შ. არსებობს სხვადასხვა სახის სირთულეები, რომელიც დღემდე ვერ წყდება და შეუძლებელი ხდება ძიების პროცესის დახვეწა. საძიებო სისტემები ამ პრობლემებს გარკვეულწილად წყვეტენ ესენია:

- ✓ გეოგრაფიული ადგილდებარეობა
- ✓ მომხმარებლის კვალიფიკაცია

ხოლო პრობლემა, რომელიც ჯერ-ჯერობით ყველაზე დიდ თავსატეხს უჩენს საძიებო სისტემებს არის:

- ენობრივი ბარიერი.

მესამე პარაგრაფი- შემუშავებული კონცეფციის ალგორითმის მუშაობის არსი. კონკრეტული ძიების პროცესში ალგორითმი მოახდენს იმ ვებ-გვერდის მისამართის ჩანიშვნას, რომელზე გადასვლის შემდეგაც დაფიქსირდება შეყოვნება გარკვეული დროით. ანუ, შეიქმნება ვებ-

გვერდების მონაცემთა ბაზა, სადაც კონკრეტული მოთხოვნის შესაბამისად, მოხდება მასზე პასუხის გამცემი ვებ-გვერდების მისამართების ჩანიშვნა.

მეოთხე პარაგრაფი- დანერგვა/ინტეგრაცია. ალგორითმის რეალიზაცია შესაძლებელია სხვადასხვანაირად, კერძოდ შესაძლებელია მოხდეს მისი ინტეგრაცია ინტერნეტ ბრაუზერებში, ან გაკეთდეს სულაც ახალი, რომელშიც ჩაშენებული იქნება ძიების ეს ალგორითმი და ის იქნება ხელმისაწვდომი ნებისმიერ ოპერაციულ სისტემაზე. ისეთი ინტერნეტ ბრაუზერები, როგორცაა Mozilla firefox, Google chrome, იძლევიან საშუალებას, მოხდეს მათ აპლიკაციაში ე.წ. აპლეტის ჩაშენება, რაც აიოლებს საქმეს თუ კი გადავწყვეტთ, რომ ალგორითმი ჩაშენებულ იქნას ინტერნეტ ბრაუზერში.

მეხუთე პარაგრაფი- კონცეფციის პრაქტიკული რეალიზაცია. Windows გარემოში სამუშაოდ. windows აპლიკაციების შესაქმნელად გამოიყენება პროგრამული გარემო Microsoft visual studio და ობიექტ ორიენტირებული ენა მაგალითად: ეს შეიძლება იყოს C++ ან C#, რომლის დახმარებითაც შესაძლებელია ვებ-ბრაუზერის გაკეთება.

მეექვსე პარაგრაფი- მოცემულია შემუშავებული კონცეფციის პრაქტიკული რეალიზაციის დეტალური აღწერილობა.

1. Windows ფორმის ჩატვირთვაზე იქმენა web ობიექტი, ამის შემდეგ ხდება მისი შევსება იმ მისამართის შესაბამისი ვებ-გვერდით, რომელიც ჩაიწერება სამისამართო გამყოფილებაში.
2. ვებ-გვერდების მისამართების დამახსოვრება იწყება მაშინ, როდესაც სამისამართო ველში მოხდება ერთ-ერთი შემდეგი საძებო სისტემის გამოძახება. იქნება ეს Google, Yandex, Bing ან Yahoo.
3. იმ შემთხვევაში, თუ მოხდა ვებ-ბმულზე განმეორებით გადასვლა, მისი მნიშვნელობა შესაბამის მონაცემთა ბაზაში გაიზრდება ერთით.

4. იმისათვის, რომ მონაცემთა ბაზა დავაკავშიროთ აპლიკაციასთან აუცილებელია მაკავშირებელი ობიექტის შექნმა და შესაბამისი დრაივერის მითითება.
5. ძიებისას ეშვება მოთხოვნა, რომელიც უზრუნველყოფს ბაზაში არსებული 10 ყველაზე მეტი მნიშვნელობის(intensivity) ვებ-ბმულის ამოტანას. და ეს სია ამოიტანება ექსპერტული ძიების კლავიშის გააქტიურებით.

უნდა აღინიშნოს, რომ ეს პროცესი დინამიური და ის დამოკიდებულია გამოყენების ინტენსივობაზე. აღნიშნული აპლიკაცია მუშავდება და ვფიქრობთ ძალიან მალე იხილავს დღის სინათლეს.

დასკვნების ნაწილში, შეჯამებულია სადისეტაციო ნაშრომის ძირითადი მიზნები და მიღებული შედეგები; თემის აქტუალურობა და შემუშავებული მეთოდების მეცნიერული სიახლე.

დასკვნები

ძირითადი თეორიული და პრაქტიკული შედეგები

სადისერტაციო ნაშრომში მიღებული შედეგების მიმართ შეიძლება გაკეთდეს შემდეგი დასკვნები:

1. თანამედროვე ინტერნეტ-ძიების დროს წამოჭრილი პრობლემების გადასაჭრელად მიძღვნილი ლიტერატურის მიმოხილვა და ანალიზი გვიჩვენებს, რომ ძიების ალგორითმები ირჩევენ ე.წ. გვერდის რანჟირების მეთოდს მისი ინდექსირებისათვის, იყენებენ კატალოგებისა და შეკითხვის დასმის სისტემებს.
2. განხილულია საძიებო სისტემის ოპტიმიზაციის საკითხები. სადაც აღწერილია Seo-ს ოპტიმიზაციის მეთოდები, არსებობს ოპტიმიზაციის სამი სხვადასხვა მეთოდი: თეთრი ოპტიმიზაცია, შავი ოპტიმიზაცია და ნაცრისფერი ოპტიმიზაცია.
3. ჩატარებულია კლასიკური ძიების ალგორითმების დეტალური ანალიზი, რომელიც გვიჩვენებს, რომ საძიებო სისტემები ვებ-რესურსების საიმედოობას განსაზღვრავენ ციფრული კალკულაციის მეთოდით.
4. ნაჩვენებია ის პრობლემები რაც წამოიჭრება ძიებისას. კერძოდ ის, რომ ვერ ხერხდება ინფორმაციის აზრობრივი მოძიება, საძიებო სისტემა ვერ გებულობს თუ რა სურს მომხმარებელს. ძიება მიმდინარეობს მხოლოდ სინტაქსურ დონეზე, ეს გულისხმობს საძიებო სისტემისათვის მიწოდებული საძიებო სიტყვის ზუსტი ასლის ძიებას ინტერნეტში.
5. გაანალიზებულია Pagerank-ძიების ალგორითმის მუშაობის პრინციპები, ის ითვალისწინებს გვერდის მნიშვნელობის შეფასებას მასზე მომმართველი ბმულების რაოდენობისა და „წონის“ მიხედვით.

6. განხილულია ინტერნეტ ძიების კონკრეტული ამოცანა რანჟირების ალგორითმების შეფასებისათვის ე.წ. მარკოვის ჯაჭვის გამოყენებით, რომელიც გვიჩვენებს, რომ ვებ-გვერდების რანჟირება ნიშნავს მათ დალაგებას იმის მიხედვით, თუ რომელ მათგანში მოხვედრის ალბათობა არის უფრო მეტი ინტერნეტში ნავიგაციის ში
7. ნაჩვენებია კლასიფიკაციის მეთოდები, რომლებიც ხელს უწყობენ ძიების პროცესის ამაღლებას. კლასიფიკაციისას ხდება დეტალური განხილვა თუ როგორ უნდა გამოისახოს სასარგებლო ვებ-გვერდი და კეთდება იმ აპლიკაციების სია რომლებიც მონაწილეობას მიიღებენ ამ პროცესში.
8. ჩატარებულია ექსპერიმენტი, რომელიც ძიების შედეგს წარადგენს სამიუბო სისტემის მიერ ამ დროისათვის არსებული ალგორითმით.
9. ნაჩვენებია, რომ აღნიშნული ძიების შედეგი არ არის მაქსიმალურად სასარგებლო. სამიუბო სიტყვის სხვადასხვაგვარი ფორმით ჩაწერისას მიიღება შედეგების სხვადასხვანაირად დალაგებული სია
10. შემუშავებულია კონცეფცია, რომელსაც საფუძვლად უდევს, მომხმარებლის მიერ გამოყენებული ვებ-გვერდების სიხშირეთა დაანაგარიშება, ძიების პროცესში ალგორითმი მომხმარებელს სთავაზობს მის მიერ ყველაზე ხშირად გამოყენებადი ვებ-გვერდების თემატურად დალაგებულ ინდექსირებულ სიას.
11. შემუშავებული ალგორითმი იყენებს მომხმარებლის ინდივიდუალურ მიდგომას სამიუბო სისტემის მიმართ. მის მიერ მოძიებული ინფორმაციის საფუძველზე ქმნის ძიების ინდექსს და ეფუძნება დინამიურ „სწავლებას“
12. წარმოდგენილი ალგორითმი აუმჯობესებს ინტერნეტ-ძიების პროცესს, მომხმარებლის მიერ გამოყენებული ვებ-გვერდების სიხშირული კონტროლით.
13. განხილულია ალგორითმის შესაბამისი პროგრამული უზრუნველყოფის შექმნის საკითხები. შესაძლებელია აპლიკაციის

შექმნა, რომელიც ინტეგრირდება ინტერნეტ ბრაუზერებში, ან ცალკეული ვებ-ბრაუზერის შექმნა, რომელიც დაფუძნებული იქნება ე.წ. ღია პლათმორმულობის პრინციპზე და ის მორგებადი იქნება ნებისმიერ ოპერაციულ სისტემაზე.

14. მოცემულია ალგორითმის პრაქტიკული რეალიზების მაგალითი. რომელიც შექმნის პროცესშია და წარმოადგენს ჩამოყალიბებული კონცეფციის ხორცშესხმას.
15. მოცემულია კონცეფციის მუშაობის სქემა, რომელიც გრაფიკულად ზუსტად გამოსახავს ალგორითმის მუშაობის პრინციპებს.
16. წარმოდგენილია კონცეფციის დადებითი და უარყოფითი მხარეები.
17. კონცეფციის უარყოფითი მხარეები დღეისათვის წარმოადგენს გადაუჭრელ პრობლემას და მის მოგვარებას მივყავართ ხელოვნური ინტელექტის შექმნის საკითხებზე.
18. კონცეფციის დადებითი მხარეები სრულად პასუხობს, ძიების ეფექტურობის ამაღლების საკითხებს და ის უზრუნველყოფს ძიების მაღალ ეფექტურობას.

The basic theoretical and practical results

We can make the following conclusions about the results of dissertation:

1. The overview and analysis of the literature on solving the problems which arose during modern Internet search shows that search algorithms choose the so-called method of page arranging for its indexing; they also use the system of catalogues and questions.
2. The search engine optimisation discussed in the dissertation describes the method of Seo optimisation. There are three different methods of optimisation: white optimisation, black optimisation and grey optimisation.
3. Algorithms of classic search have been analysed in detail. The analysis shows that search engines define the reliability of web-resources using digital calculation method.
4. There are problems revealed during the search. Namely, it is not possible to search for mental information, the search system can't identify what the user wants. The search is only performed at a syntactic level, which means that the search engine is given an exact copy of the searched word.
5. The working principles of Pagerank-search algorithm are analysed. They consider the assessment of the importance of the page according to the quantity and "weight" of the links connected to it.
6. There is a review of a specific task of the Internet search to assess arranging algorithm by using the so-called Markov's chain, which shows that arranging web-pages means their arrangement according to which of them they are most likely to belong to while navigating the Internet.
7. The thesis has shown the methods of classification which help improve the process of search. While classifying, there is a detailed overview of how a

useful web-page should be presented and there is a list of applications which will take part in this process.

8. We have carried out an experiment which presents the search result by search system with the existing algorithm.
9. It is shown that the given search result is not useful enough. While writing the search word in different forms, we get a differently arranged list of results.
10. We have worked out a conception which is based on the frequency of web-pages used by users. In the search process, the algorithm offers users the thematic index list of the most frequently used web-pages.
11. The algorithm uses individual approach of the users to search system. Based on the acquired information, it creates a search index and bases it on dynamic “studies”.
12. The given algorithm improves the process of Internet search by controlling the frequency of usage of a web-page by users.
13. The thesis has reviewed the issues of creation of relevant programming provision of algorithm. It is possible to create an application which is integrated in the Internet browsers, or to create separate web-browsers which will be based on the so-called open platforming principle and will be adjustable to any operational system.
14. There is an example of practical realisation of algorithm. It is in the process of creation and is the realisation of an established conception.
15. There is a working scheme of the conception which presents the working principles of algorithm in a precise graphic way.
16. Positive and negative sides of the conception have been revealed.
17. Nowadays, the negative sides of the conception are an unsolved problem and its solution leads to the creation of artificial intelligence.
18. The positive sides of the conception fully correspond to the issues of improving search efficiency which provides the high quality of search.

დისერტაციის ძირითადი შინაარსი გამოქვეყნებულია

შემდეგ სამეცნიერო სტატიებში:

1. ქუთაისის სასწავლო უნივერსიტეტი ლამპარი. პროფესორ მასწავლებელთა III რესპუბლიკური სამეცნიერო კონფერენცია. თანამედროვე მეცნიერების აქტუალური საკითხები. შრომები 2010. „ინფორმაციული საზოგადოება და გეოინფორმაციული სისტემები“ მანანა ქარქაშაძე, ლაშა გავაშელაშვილი.
2. საერთაშორისო სამეცნიერო ჟურნალი ინტელექტი 1(48). 2014წ. „ეფექტური ინტერნეტ ძიების ალგორითმები(PageRank,MartxNET) მიმოხილვა“ გვ 63. ლაშა გავაშელაშვილი.
3. აკაკი წერეთლის სახელმწიფო უნივერსიტეტი IV საერთაშორისო სამეცნიერო მეთოდური კონფერენცია. სწავლებისა და აღზრდის აქტუალური საკითხები. 2013 შრომები. „ინფორმაციის ძიების ოპტიმალური მეთოდების შემუშავების საკითხები“ გვ.191. ლაშა გავაშელაშვილი.
4. პედაგოგთა კვალიფიკაციის ამაღლების ქუთაისის ზონალური ინსტიტუტი.
I საერთაშორისო სამეცნიერო-მეთოდური კონფერენცია. განათლებისა და მეცნიერების მაღალი ინტელექტუალური ტექნოლოგიები 2013.შრომები „ინტერნეტში ძიების თანამედროვე საშუალებების ანალიზი“ გვ. 16. ლაშა გავაშელაშვილი.
5. მეცნიერებისა და საზოგადოების განვითარების ფონდი "ინტელექტი"-საქართველო III საერთაშორისო სამეცნიერო კონფერენცია. თეზისები. 2014 „ინტერნეტ ძიების ალგორითმი და GOOGLE"-ს არქიტექტურა“ გვ.37. ლაშა გავაშელაშვილი.
6. მეცნიერებისა და საზოგადოების განვითარების ფონდი "ინტელექტი"-საქართველო III საერთაშორისო სამეცნიერო კონფერენცია.

თეზისები.2014. „ვებ-გვერდების კლასიფიკაცია“ გვ.38. ლაშა გავაშელაშვილი.

7. პედაგოგთა კვალიფიკაციის ამაღლების ქუთაისის ზონალური ინსტიტუტი.
რესპუბლიკური სამეცნიერო-მეთოდური კონფერენცია II შრომები „ინტერნეტ ძიების ალგორითმები“ გვ.27. ლაშა გავაშელაშვილი.
8. მეცნიერებისა და საზოგადოების განვითარების ფონდი "ინტელექტი"-საქართველო III საერთაშორისო სამეცნიერო კონფერენცია.
შრომები „ვებ-გვერდების კლასიფიკაცია“ ლაშა გავაშელაშვილი გურამ ჩაჩანიძე რომან სამხარაძე.
9. მეცნიერებისა და საზოგადოების განვითარების ფონდი "ინტელექტი"-საქართველო III საერთაშორისო სამეცნიერო კონფერენცია.
შრომები „ინტერნეტ ძიების ალგორითმი და GOOGLE"-ს არქიტექტურა“ ლაშა გავაშელაშვილი.