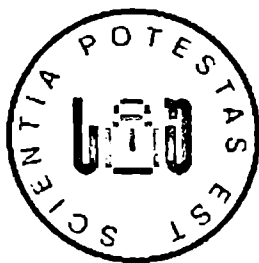


საქართველოს ტექნიკური უნივერსიტეტი

გია სურგულაძე, დავით გოგშელიძე,
გოჩა დალაქიშვილი

საინფორმაციო საზოგადოება:
საგამომცემლო მარკეტინგი და
მანქანური დასწავლა



დამტკიცებულია:

სტუ-ს „IT კონსალტინგის სამეც-
ნიერო ცენტრის“ სარედაქციო
კოლეგიის მიერ - ოქმი N8 3.07.23

თბილისი - 2023

უკ 004.5

განიხილება ინფორმაციული საზოგადოების ფორმირების მიზნით საგამომცემლო მარკეტინგის ბიზნესპროცესების ავტომატიზაციის როლი; იმიტაციური მოდელირების, ანალიზის და პროექტირების ტექნოლოგიების შემუშავება პეტრის ქსელების თეორიის და საწარმოო რესურსების მართვის (ERP) სისტემების საფუძველზე. ჩატარებულია ელ-ბიბლიოთეკების და ონლაინ მაღაზიების არსებული მრავალფუნქციური ტექსტური რედაქტორების ანალიზი. შესწავლილია მათი თვისებები, მწერალი-ავტორებისა და მკითხველთა კომფორტის გათვალისწინებით. შემოთავაზებულია ახალი კონცეფცია, პრინციპები და მეთოდები მათი სრულყოფის მიზნით. კერძოდ, სამეცნიერო ნაშრომების შექმნისათვის ისეთი პროგრამული პროდუქტის აგება, რომელიც სხვადასხვა სფეროს (ქანრის) მეცნიერს (მწერალს) მიაწვდის მასზე მორგებულ ინტერფეისს და ფუნქციონალს. მანქანური დასწავლის ალგორითმების დახმარებით იქმნება მოდელები, რომელთა გამოყენებითაც შესაძლებელია ტექსტური ინფორმაციის შექმნის და საგამომცემლო მარკეტინგის სასიცოცხლო ციკლის ბიზნესპროცესების გამარტივება. მონოგრაფია განკუთვნილია საგამომცემლო სისტემის, ლინგვისტიკის და პროგრამული დეველოპმენტის ავტომატიზაციით დაინტერესებული მკითხველისა და დოქტორანტებისათვის.

რეცენზენტები:

- პროფ. გ. გოგიჩაიშვილი - საქ. მეცნიერებათა ეროვნული აკადემიის წევრ-კორესპოდენტი, ტ.მ.დ.
- ასოც.პროფ. დ. მასხარაშვილი - სეუ, აკად.დოქტორი

რედკოლეგია:

ა. ფრანგიშვილი (თაემჯდომარე), ზ. აზმაიფარაშვილი, მ. ახოზაძე, გ. გოგიჩაიშვილი, ზ. ბოსიკაშვილი, ე. თურქია, რ. კაკუბავა, ვ. კვარაცხელია, თ. ლომინაძე, ნ. ლომინაძე, ჰ. მელაძე, ლ. პეტრიაშვილი, გ. სურგულაძე (რედაქტორი), ბ. შანშიაშვილი, ო. შონია, ზ. წვერაიძე

© სტუ-ს „IT კონსალტინგის სამეცნიერო ცენტრი“, 2023
ISBN 978-9941-8-5442-2



ყველა უფლება დაცულია, ამ წიგნის არც ერთი ნაწილი (იქნება ეს ტექსტი, ფოტო, ილუსტრაცია თუ სხვა) გამოყენება არც ერთი ფორმით და საშუალებით (ელექტრონული თუ მექანიკური), არ შეიძლება გამომცემლის წერილობითი ნებართვის გარეშე. საავტორო უფლებების დარღვევა ისჯება კანონით.

Georgian Technical University

**Surguladze Gia, Gogshelidze David,
Dalakishvili Gocha**

**INFORMATION SOCIETY:
PUBLISHING MARKETING AND
MACHINE LEARNING**



The role of automation of publishing marketing business processes for the purpose of forming the information society is discussed. In the research process, simulation modeling, analysis and design technology based on Coloured Petri Nets (CPN tools) and Enterprise Resource Planning (ERP) systems are used. Today's multifunctional text editors for e-libraries and online stores are analyzed. Their positive and negative qualities are studied, taking into account the comfort of writers-authors and readers. The concept, principles and methods for the development of such text editors are developed. Attention is paid to the creation of a software product for the creation of scientific papers, which will offer a scientist with different scientific directions an interface and functionality specially adapted to him. With the help of machine learning algorithms, models are created that can be used to simplify the business processes of textual information creation and publishing marketing. The monograph is intended for readers and doctoral students interested in modeling and programming publishing marketing processes.

© „IT-Consulting scientific center” of GTU, Tbilisi, 2023
ISBN 978-9941-8-5442-2



ედღვნება:

საქართველოს ტექნიკური უნივერსიტეტის
„ინფორმაციული საზოგადოების“ კათედრის დაარსების მე-20
წლისთავს (2003-2023) და მისი დამაარსებლის, აკადემიკოს
გოჩა ჩოგოვადის ნათელ ხსოვნას (1941-2022)

DEDICATED:

to the 20th Anniversary of the establishment "*Information Society*"
department of Georgian Technical University (2003-2023) and the bright
memory of its founder, academician *Gocha Chogovadze* (1941-2022)



საქართველოს ტექნიკური უნივერსიტეტის
რეფორმატორი რექტორი (1988-1994), იუნესკოს
(პარიზი) კულტურისა და განათლების დეპარ-
ტამენტის ღირუქტორი (1981-1988). დიდი
საზოგადო და სახელმწიფო მოღვაწე,

საქართველოს იუნესკოს საქმეთა ეროვნული
კომისიის ვიცე-პრეზიდენტი (2004-2012), ევრო-
პის და რუსეთის მრავალი უნივერსიტეტის
საპატიო დოქტორი და საერთაშორისო აკადემი-
ების წევრი, დიპლომატი, საქართველოს საგან-
გებო და სრულუფლებიანი ელჩი საფრანგეთსა

და ესპანეთის სამეფოში (1994-2004), სტუ-ს „მართვის ავტომატიზებული
სისტემების“ (1971) და სტუ-ს UNESCO-ს „ინფორმაციული საზოგადოების“
(2003) კათედრების დამაარსებელი, მათი პირველი გამგე.

საქართველოს მეცნიერებათა ეროვნული აკადემიის ნამდვილი წევრი
(1994-დან), ტექნიკის მეცნიერებათა დოქტორი (1975) - საპატიო პროფესორი
– *გოჩა გიორგის ძე ჩოგოვაძე* – მრავალი სამეცნიერო წიგნის, პროექტისა და
სტატიის ავტორი, მრავალი სტუდენტის აღმზრდელი და ახალგაზრდა
მეცნიერის „სამეცნიერო ნათლია“, ქართული და უცხოური ლიტერატურის,
პოეზიის, მუსიკისა და სიმღერის დიდი მოყვარული, საერთაშორისო
მეგობრობის დიდოსტატი, სუფრის ორიგინალური თამადა, მოსიყვარულე და
გულისხმიერი მეუღლე, მამა და ბაბუა, მუდამ სიკეთის მთესველი,
განუმეორებელი „ქართველი ფაუსტი“

შინაარსი

შესავალი	7
თავი 1. ინფორმაციული საზოგადოება და საგამომცემლო ინდუსტრია	14
1.1. ინფორმაციული საზოგადოების ფორმირება და საგამომცემლო ბიზნესი	15
1.2. საგამომცემლო მარკეტინგის სისტემის როლები და ფუნქციები	17
1.3. წიგნების გამოცემის პროცესის სასიცოცხლო ციკლი	20
1.4. ERP სიტემა საგამომცემლო ინდუსტრიაში	26
1.5. მარკეტინგული პროცესის იმიტაციური მოდელი პეტრის ფერადი ქსელით (CPN)	30
1.6. კვლევის მიზანი და შედეგების ინოვაციურობა	35
1.7. კვლევის მეთოდოლოგია, ინტერდისციპლინურობა და არეალი	41
1.8. საგამომცემლო ბიზნესპროცესების მოდელირება BPMN სტანდარტით	48
თავი 2. კვლევის მეთოდები და ინსტრუმენტები	53
2.1. ხელოვნური ინტელექტი და მანქანური დასწავლა	55
2.2. ბუნებრივი ენის დამუშავება (NLP)	63
2.3. თემის მოდელირება	69
2.4. დირიხლეს ლატენტური განაწილება	74
2.5. არაუარყოფითი მატრიცის ფაქტორიზაცია	77
2.6. კლასიფიკაცია	82
2.7. მონაცემთა განთავსება სერვერზე	85
2.7.1. მონაცემთა ბაზები	86
2.7.2. ინსტრუმენტები რეალური დროის ფუნქციების შესასრულებლად	86
2.8. ბიბლიოთეკები და პროგრამული ენები	88
2.9. პროგრესული ვებ აპლიკაცია (PWA)	90
თავი 3. მანქანური დასწავლა: კვლევა და შედეგები	91
3.1. მონაცემთა მოძიება	92
3.2. ნაშრომთა თემის მოდელირება	93

3.2.1. მონაცემთა წინასწარი დამუშავება	94
3.2.2. LDA და NNMF მეთოდების გაშვება და მოდელების მიღება	96
3.3. სამეცნიერო სტატიების რეზიუმეთა კლასიფიკაცია	104
3.3.1. მონაცემთა წინასწარი დამუშავება	105
3.3.2. კლასიფიკაციის მოდელის მიღება	107
თავი 4. ტექსტური ინფორმაციის შექმნისათვის საჭირო ფუნქციონალის და ინტერფეისის კვლევა	111
4.1. გამოკითხვა და მისი ანალიზის შედეგები	112
4.1.1. ფუნქციონალი	112
4.1.2. ბიზნესპროცესების სრულყოფა	130
4.2. სხვადასხვა ტიპის სამეცნიერო ნაშრომებში არსებული ფუნქციონალის კვლევა და შედეგები	134
თავი 5. საგამომცემლო სფეროს ახალი პროგრამული პროდუქტის პროტოტიპის შექმნა	143
5.1. ტექსტური ინფორმაციის შექმნა	144
5.2. ტექსტის წამკითხველი	145
5.3. ახალი პროდუქტის შექმნის სისტემის პროტოტიპი	146
5.4. უსაფრთხოება	151
დასკვნა	155
გამოყენებული ლიტერატურა	157
დანართები	171
დ-1: მარკეტინგული მენეჯმენტის ბიზნეს პროცესები	172
დ-2: ინფორმაციული სისტემები (IS) და საწარმოო რესურ- სების მართვა (ERP, ისტორიულ ფაქტებით)	179
დ-3: პეტრის ქსელების თეორიული საფუძვლები	186
დ-4: პროგრამების ოპტიმიზაციის მანქანური დასწავლის ალგორითმები	201
დ-5: DevOps მეთოდოლოგია და ინსტრუმენტული საშუალებები.....	213
ავტორთა შესახებ	215

შესავალი

საქართველოს ტექნიკური უნივერსიტეტის ისტორიაში პირველად 2003 წლის 30 ივნისს დაარსდა სტუ-ის იუნესკოს „ინფორმაციული საზოგადოების“ კათედრა, აკადემიკოს გოჩა ჩოგოვაძის ინიციატივით (ამ დროისთვის იგი საქართველოს საგანგებო და სრულუფლებიანი ელჩი გახლდათ საფრანგეთსა და ესპანეთის სამეფოში), UNESCO-ს (პარიზი) გენერალური დირექტორის (კოიჩირო მაცუურას) და სტუ-ის რექტორის (აკად. არჩილ ფრანგიშვილის) და აკადემიური საბჭოს მხარდაჭერით, სამეცნიერო და საგანმანათლებლო სფეროს აქტუალურ საკითხებზე ერთობლივი თანამშრომლობის კოორდინაციის მიზნით. კათედრა ინფორმატიკისა და მართვის სისტემების ფაკულტეტის სტრუქტურული ერთეული გახდა, ხოლო აკად. გოჩა ჩოგოვაძე მისი პირველი გამგე [1-3]. ამჟამად მას ფაკულტეტის დეკანი, პროფ. თამარ ლომინაძე ხელმძღვანელობს.

ინფორმაციული საზოგადოება და მასთან დაკავშირებული სოციალური, კულტურულ-საგანმანათლებლო, ეკონომიკური, ტექნიკურ-ტექნოლოგიური და სხვა სფეროების პრობლემები აქტუალური გახდა XXI-საუკუნის დასაწყისიდან და განსაკუთრებული მნიშვნელობა შეიძინა Covid-19 პანდემიის პერიოდში [4]. უწყვეტი განათლების სისტემაში (სკოლა, კოლეჯი, უნივერსიტეტი), სახელმწიფო ორგანიზაციებსა და კერძო სტრუქტურებში, ოჯახებსა და თითქმის ყველა სამუშაო ადგილზე ინტენსიურად მოიმატა პერსონალური და მობილური ტექნიკის გამოყენების დონემ, ინფორმაციული და კომუნიკაციური ტექნოლოგიების (ICT) სფეროს დისციპლინების სწავლების ინტენსიფიკაციამ ციფრული ტექნოლოგიების ბაზაზე [3, 5]. ქენევის ყოველწლიური მსოფლიო სამიტის ფორუმი (World Summit on the Information Society – WSIS), ამის მკაფიო დადასტურებაა, სადაც

აქტიურად განიხილება ინფორმაციული საზოგადოების მდგრადი განვითარების აქტუალური საკითხები [6, 7].

ამგვარად, საინფორმაციო საზოგადოება ადამიანთა ის ერთობაა (მიუხედავად ეროვნების, რელიგიის, მსოფლმხედველობის, პროფესიის, სქესის, ასაკის და ა.შ.), სადაც ინფორმაციის შექმნა, გავრცელება, მანიპულირება, ინტეგრაცია და გამოყენება მათი დარგის, სფეროს ან ამოცანების გადაწყვეტაში მნიშვნელოვანი საქმიანობა ან გართობაა [4]. მისი მთავარი მამოძრავებელი ძალა საინფორმაციო და საკომუნიკაციო ტექნოლოგიებია (ICT).

ჩვენი წიგნის მიზანია საინფორმაციო საზოგადოების ფორმირების პროცესში საგამომცემლო ინდუსტრიის და მარკეტინგის როლის მნიშვნელობის გამოკვლევა და შესაბამისად, ახალი თაობის მხარდამჭერი ინფორმაციული სისტემის და ინსტრუმენტების შექმნა, რომლებიც მორგებული იქნება ამ დარგის (საგამომცემლო ინდუსტრიის) პროცესების ავტომატიზაციაზე [8-10].

გამომცემლობა (Publishing) არის საქმიანობა – ინფორმაციის, ლიტერატურის, მუსიკის, პროგრამული უზრუნველყოფის და სხვა შინაარსის (გამოსაცემი პროდუქტის) საზოგადოებისათვის გასაყიდად ან უფასოდ მიწოდების მიზნით. ტრადიციული (ბეჭდური ნაწარმოებების შექმნა და გავრცელება), თანამედროვე ციფრული ტექნოლოგიები, მათ შორის ელექტრონული ბიბლიოთეკები, წიგნების ინტერნეტ მაღაზიები და ა.შ. განსაკუთრებულ როლს ასრულებენ საინფორმაციო საზოგადოების ფორმირებასა და განვითარებაში [11, 12].

კაცობრიობის განვითარების უმთავრესი მიზეზი ადამიანთა ფანტაზიაა, აღმოჩენის, გამოცდილების მიღების უნარი და ამ ინფორმაციის გაზიარებისა და მიღების საშუალება. ინფორმაციის გაზიარების ერთ-ერთი ყველაზე მნიშვნელოვანი საშუალება მისი წერილობითი ასახვაა. ადამიანმა წერა ჯერ კიდევ ქვაზე დაიწყო და

ცივილიზაციასთან ერთად ფურცელზე გააგრძელა, შემდეგ კი უკვე კლავიატურაზე. თანამედროვე ტექნოლოგიური ევოლუციის პარალელურად, აუცილებელი გახდა ტექსტური ინფორმაციის ელექტრონულ ფორმატში გადატანა. წიგნებისა და სხვა ტექსტური ნაშრომების უმეტესობა ხელმისაწვდომია ელექტრონულ ბიბლიოთეკებსა და წიგნის მაღაზიებში. ისინი მკითხველებს საშუალებას აძლევენ მსოფლიოს ნებისმიერ წერტილში, მოხდეს წიგნის შესყიდვა ან მისი წაკითხვა სხვადასხვა ელ-მოწყობილობებზე. თავის მხრივ, ნაშრომების ელექტრონული კომერცია და გამოყენება ხელს უწყობს გარემოს დაცვას ნაკლები ქაღალდის გამოყენებით.

ასეთი ტექნოლოგიური და ავტომატიზებული სისტემების განვითარების ფონზე კი შეგვიძლია დარწმუნებით ვთქვათ, მისი განვითარების გზა ჯერ კიდევ წინაა. განვითარების ერთ-ერთ პოტენციურ მიმართულებად შეგვიძლია განვიხილოთ მდგომა, რომელიც მეცნიერებსა და ტექსტური ინფორმაციის სხვა შემქმნელების მუშაობის პროცესს სპეციალურად მათზე მორგებული ფუნქციონალისა და ინტერფეისის საშუალებით გაამარტივებს.

ამჟამად, არსებული ტექსტური რედაქტორები, მორგებულია ყველა ტიპის ინფორმაციის შექმნაზე ერთდროულად, მწერალს კი ნაკლები შანსი აქვს მიჰყვეს მისი წერის მეთოდოლოგიას.

მონოგრაფიის ერთ-ერთი ინოვაციურობა სწორედ ამ პრობლემის მოგვარებაში მდგომარეობს. დღეისათვის არსებობს მრავალი დოკუმენტური ფაილების შექმნის პროგრამა, მათ შორის მაიკროსოფტის ყველაზე ცნობილი საოფისე პაკეტის MsWord ინსტრუმენტი. თუმცა არ არსებობს სპეციალურად სამეცნიერო ნაშრომების, წიგნების შექმნისა და თარგმანისთვის განკუთვნილი პროგრამული პროდუქტი, რომელიც თითოეული ამ კატეგორიისათვის იქნება ხელშემწყობი.

წიგნში ასახულია ავტორების მიერ ჩატარებული სამეცნიერო პროექტის კვლევები სხვადასხვა ტიპის მწერლებისა და მათი

მუშაობის მეთოდოლოგიის, სხვადასხვა კატეგორიის სამეცნიერო ნაშრომებში გამოყენებული ფორმების და შესაბამისად, წერის დროს საჭირო ფუნქციონალის შესახებ.

მონოგრაფია ხუთი თავისაგან შედგება.

პირველ თავში გადმოცემულია ინფორმაციული საზოგადოების ფორმირების პროცესის არსი, მისი მიზანი, პრობლემები და ამოცანები. ამ პროცესებზე მოქმედი ფაქტორები. განსაკუთრებით, საგანმანათლებლო სფეროს ევოლუციური, ინტენსიური განვითარება ინფორმაციული და საკომუნიკაციო ტექნოლოგიების ბაზაზე. განხილულია ტრადიციული და ციფრული საგამომცემლო ტექნოლოგიების როლი საინფორმაციო საზოგადოების ფორმირებაზე. გაანალიზებულია საგამომცემლო მარკეტინგის პროცესები, მისი სასიცოცხლო ციკლი, შესაბამისი პროდუქციის დაგეგმვის, წარმოებისა და გავრცელებისათვის. წარმოდგენილია საგამომცემლო ინდუსტრიის რესურსების მართვის მხარდამჭერი ERP-სისტემის კონცეფცია. აგებულია ამ პროცესების იმიტაციური მოდელი პეტრის ფერადი ქსელების გამოყენებით.

მეორე თავი ეხება საგამომცემლო ინდუსტრიის კომპანიების მხარდამჭერი საინფორმაციო სისტემების ასაგებად შესაბამისი ინფრასტრუქტურის განსაზღვრას და კვლევის მეთოდების განხილვას. კერძოდ, ყურადღება გამახვილებულია ხელოვნური ინტელექტისა და მანქანური დასწავლის მეთოდებისა და ალგორითმების გამოყენებაზე. აქ წარმოდგენილია ბუნებრივი ენის დამუშავების, თემის მოდელირების, ტექსტის კლასიფიკაციის და სხვ., მანქანური დასწავლისა და ლინგვისტური მეთოდების და ალგორითმების მუშაობის პრინციპები. გამოყენებულია ბაიესის პირობითი ალბათობის, დირიხლეს ლატენტური და არაუარყოფითი მატრიცის ფაქტორიზაციის მეთოდები. ამასთანავე, მეორე თავში განხილულია კვლევის პროექტისთვის მონაცემთა რელაციური ბაზა MySQL, რეალურ დროში განახლებადი მონაცემთა ბაზა Google Firebase და

არარელაციური NoSQL-ის ბაზა და მათი უსაფრთხოდ განთავსება ღრუბლოვან სერვერებზე. პროექტში გათვალისწინებულია Python პროგრამული ენის, მისი ბიბლიოთეკების და პროგრესული ვებ აპლიკაციის PWA და სხვ. გამოყენება. კლიენტისა და სერვერის აპლიკაციებისთვის (Frontend, Backend) გათვალისწინებულია Angular, TypeScript და Node.js.

მესამე თავში განხილულია მანქანური დასწავლის მეთოდების და ალგორითმების გამოყენების საფუძველზე საგამომცემლო მარკეტინგის ბიზნესპროცესებისთვის ჩატარებული კვლევებით მიღებული შედეგები. კერძოდ, პროცესების თანამიმდევრობა შესაბამება მარკეტინგის სასიცოცხლო ციკლის ეტაპებს: მოძიებული და შეკრებილია მონაცემები (მაგალითად, Arxiv.org-ის მონაცემთა ბაზა, რომელიც ორ მილიონამდე სამეცნიერო ნაშრომს მოიცავს. ამასთან დამატებული იქნა Pubmed-ის მონაცემთა ბაზა, რომელიც 130000-მდე სამეცნიერო სტატიისგან შედგება).

ჩატარებულია მონაცემთა წინასწარი დამუშავება (ფილტრაცია) და შემდეგ დირიხლეს და არაუარყოფითი მატრიცის ფაქტორიზაციის მეთოდებით დამუშავება. მიღებულია თემის მოდელირების პროცესის დამაკმაყოფილებელი შედეგები. შედარებულია მეთოდების შედეგები და გამოვლენილია მათ შორის უკეთესი კონკრეტული კლასტერების შესაბამისად.

მესამე თავში ჩატარებულია აგრეთვე ტექსტების (სამეცნიერო სტატიების რეზიუმეების) კლასიფიკაცია. იგი კონტროლირებადი (მართვადი) მანქანური დასწავლის მეთოდია. შესაბამისად, მას ორი შემავალი პარამეტრი აქვს: თავად შესასწავლი ტექსტი და მისი კატეგორია. მრავალკატეგორიიანი სტატიების შემთხვევაში არჩეულ იქნა პირველი კატეგორია. ამ მიზეზით არსებული ხარვეზები, გადაჭრილ იქნა მონაცემთა სიმრავლის მეშვეობით. სტატიების რეზიუმეთა კლასიფიკაციისათვის გამოყენებულ იქნა ორი მეთოდი, ბაიესის მიაშიტი ალგორითმი, და მეორე, წრფივი

მხარდაჭერის ვექტორული კლასიფიკატორი. გამოკვლეულ იქნა მათი შედეგები სხვადასხვა პირობებისათვის.

მეოთხე თავში წარმოდგენილია ჩვენ მიერ განხორციელებული კონტექსტუალური გამოკითხვები და სხვადასხვა ტიპის ტექსტებში (მაგალითად, სამეცნიერო ნაშრომებში) არსებული ფორმების კვლევა და მისი შედეგები. ასევე, ჩატარდა ტექსტურ ინფორმაციაზე მომუშავე რედაქტორის ფუნქციონალის ნაკლოვანებებისა და მისი არასრულყოფილების გამოვლენა. გამოკითხვისას დასმული შეკითხვები ეხებოდა რესპოდენტთა მიერ გამოყენებულ ტექსტურ რედაქტორებს. კერძოდ, რა მოსწონდათ და რა არა მათ ამ რედაქტორებში, რას სრულყოფდნენ და რის დანაკლისს განიცდიდნენ, რა დამატებით ინსტრუმენტებს იყენებდნენ წერის პროცესში ტექსტური რედაქტორის გარდა და რაში. რესპოდენტებმა პასუხი გასცეს შეკითხვებს – თუ როგორ ახდენდნენ ისინი ნაშრომთა შენახვასა და გამოქვეყნებას, როგორ ახორციელებენ თანავტორებთან ურთიერთობას და მათ მიერ ნაშრომის განხილვის პროცესის მენეჯმენტს.

ამასთანავე, გამოკვლეულ იქნა სხვადასხვა ტიპის ტექსტებში (მაგალითად, სამეცნიერო ნაშრომებში) არსებული ფორმები. ყურადღება გამახვილდა ისეთი ფორმების არსებობაზე, რომელთა გამოყენებისთვისაც რაიმე სპეციალური ფუნქციების არსებობაა საჭირო. ასეთ ფორმებად აღებულია: დიაგრამები, გრაფიკები, ცხრილები, ფორმულები, სამედიცინო სურათები, მოლეკულური სტრუქტურები, პროგრამული კოდები და რუკები. სხვადასხვა კატეგორიის სამეცნიერო ნაშრომებში გამოყენებული ფუნქციონალი საკმაოდ მრავალფეროვანია. თუ განვიხილავთ თითოეული მიმართლების გამოყენებულ ფორმებს, აღმოვაჩენთ, რომ სხვადასხვა ტიპის ნაშრომში სხვადასხვა ფორმებია პრიორიტეტული. მაგალითად, ლოგიკასა და მათემატიკაში, ფორმების

უმრავლესობა ფორმულებია, როდესაც სტატისტიკასა და მონაცემთა მეცნიერებაში გრაფიკები ჭარბობს.

მეხუთე თავი, ერთგვარად შემუშავებული შედეგების შემაჯამებელი ნაწილია. აქ კიდევ ერთხელ, კომპაქტურად არის წარმოდგენილი საგამომცემლო სფეროში არსებული პროდუქტები (რედაქტორები) თავისი პლიუს-მინუსებით. მათი ავტორებზე, მეცნიერებსა თუ მთარგმნელებზე მოსარგებად, სამუშაო გარემოს სრულყოფის მიზნით, განსაზღვრულია სამუშაო სპეციფიკიდან გამომდინარე, აუცილებელი და საკმარისი ფუნქციონალების განთავსებით ინტერფეისზე. ეს კი სპეციალიზებული რედაქტორის ახალი პროტოტიპის შექმნის ტოლფასია.

მეხუთე თავში განხილულია ასევე საპროექტო სისტემის უსაფრთხოების საკითხები, რეგისტრაციის, იდენტიფიკაციისა და ავტორიზაციის საფუძველზე. პროექტში მოდელირებული სისტემის უსაფრთხოება უზრუნველყოფილია Node.js ბიბლიოთეკის მეშვეობითა.

დანართში მოცემულია მოკლე საინფორმაციო მასალა მარკეტინგის და პეტრის ქსელების საფუძვლების, ERP-ინფორმაციული სისტემის და მანქანური დასწავლის კლასიკური ალგორითმების შესახებ. წიგნის ეს ნაწილი განკუთვნილია იმ მკითხველისათვის, რომელსაც სჭირდება საწყის, აუცილებელ ტერმინებსა და ცნებებში გარკვევა.

წიგნის ავტორები მადლობას უხდიან წინასწარ გულისხმიერ მკითხველებს, რომლებიც მოგვაწვდიან შენიშვნებს და რეკომენდაციებს ნაშრომის მომავალი სრულყოფის მიზნით:

გია სურგულაძე - g.surguladze@gtu.ge

დავით გოგშელიძე - davitil@mail.com

გოჩა დალაქიშვილი - g.dalakishvili@gtu.ge

თავი 1

ინფორმაციული საზოგადოება და საგამომცემლო ინდუსტრია

1.1. ინფორმაციული საზოგადოების ფორმირება და საგამომცემლო ბიზნესი	15
1.2. საგამომცემლო მარკეტინგის სისტემის როლები და ფუნქციები	17
1.3. წიგნების გამოცემის პროცესის სასიცოცხლო ციკლი	20
1.4. ERP სიტემა საგამომცემლო ინდუსტრიაში	26
1.5. მარკეტინგული პროცესის იმიტაციური მოდელი პეტრის ფერადი ქსელით	30
1.6. კვლევის მიზანი და შედეგების ინოვაციურობა	35
1.7. კვლევის მეთოდოლოგია, ინტერდიციპლინურობა და არეალი	41
1.8. საგამომცემლო ბიზნესპროცესების მოდელირება BPMN სტანდარტით	48

1.1. ინფორმაციული საზოგადოების ფორმირება და საგამომცემლო ბიზნესი

ტერმინი „საინფორმაციო (ან ინფორმაციული) საზოგადოება“ (Information Society) XIX საუკუნის შუა წლებიდან იღებს სათავეს, როგორც პოსტინდუსტრიული საზოგადოების მემკვიდრე. მისი ერთ-ერთი ყველაზე დამსახურებული ინიციატორია ავსტრიელი წარმოშობის ამერიკელი ეკონომისტი ფრიც მაჩლუპი (Fritz Machlup), რომელიც იკვლევდა თემას „ცოდნა – როგორც ეკონომიკის რესურსი“. იგი ფართო პოპულარიზაციას უწევდა ინფორმაციული საზოგადოების კონცეფციას [13].

ინფორმაციული საზოგადოება მსოფლიოში აქტუალური გახდა XXI-საუკუნის დასაწყისიდან და განსაკუთრებული მნიშვნელობა შეიძინა მისი ფორმირების პრობლემის გადაწყვეტამ Covid-19 პანდემიის პერიოდში [4]. უწყვეტი განათლების სისტემაში (სკოლა, კოლეჯი, უნივერსიტეტი), სახელმწიფო ორგანიზაციებსა და კერძო სტრუქტურებში, ოჯახებსა და თითქმის ყველა სამუშაო ადგილზე ინტენსიურად მოიმატა პერსონალური და მობილური ტექნიკის გამოყენების დონემ, ინფორმაციული და კომუნიკაციური ტექნოლოგიების (ICT) სფეროს დისციპლინების სწავლების ინტენსიფიკაციამ ციფრული ტექნოლოგიების ბაზაზე [3, 5].

განსაკუთრებით მაღალ დონეზეა ასახული ეს გამოწვევა ჟენევის ყოველწლიურ მსოფლიო სამიტის ფორუმებზე (World Summit on the Information Society – WSIS), სადაც აქტიურად განიხილებოდა ინფორმაციული საზოგადოების მდგრადი განვითარების აქტუალური საკითხები [6].

ამგვარად, საინფორმაციო საზოგადოება ადამიანთა ის ერთობაა, სადაც ინფორმაციის შექმნა, გავრცელება, მანიპულირება,

ინტეგრაცია და გამოყენება მათი დარგის, სფეროს ან ამოცანების გადაწყვეტაში მნიშვნელოვანი საქმიანობაა [7]. მისი მთავარი ძალაა საინფორმაციო და საკომუნიკაციო ტექნოლოგიები (ICT).

ჩვენი წიგნის მიზანია საინფორმაციო საზოგადოების ფორმირების პროცესში საგამომცემლო ინდუსტრიის და მარკეტინგის როლის მნიშვნელობის გამოკვლევა და შესაბამისად, ახალი თაობის მხარდამჭერი ინფორმაციული სისტემის და ინსტრუმენტების შექმნა, რომლებიც მორგებული იქნება ამ დარგის (საგამომცემლო ინდუსტრიის) პროცესების ავტომატიზაციაზე.

გამომცემლობა (Publishing) არის საქმიანობა – ინფორმაციის, ლიტერატურის, მუსიკის, პროგრამული უზრუნველყოფის და სხვა შინაარსის (გამოსაქვეყნებელი პროდუქტის) საზოგადოებისათვის გასაცემად ან უფასოდ მიწოდების მიზნით. ტრადიციულად, ტერმინი „გამოცემა“ ეხება ბეჭდური ნაწარმოებების შექმნას და გავრცელებას, როგორცაა წიგნები, გაზეთები და ჟურნალები [8, 9].

ტრადიციული და ციფრული საგამომცემლო კომპანია არის *ბიზნესი*, რომელიც ქმნის და ავრცელებს ნაწარმოების წერილობით ან ციფრულ ასლებს (საბავშვო მხატვრული ბროშურებით დაწყებული და კლასიკური რომანებით დამთავრებული), სხვადასხვა სფეროს და ჟანრის წიგნებს, გაზეთებს, ჟურნალებს და ა.შ.

საგამომცემლო ინდუსტრია (Publishing Industry) – ფართო გაგებით მოიცავს გაზეთების, ჟურნალების, მუსიკალური ნაწარმოებების, რუქების, სამთავრობო ინფორმაციის, კომიქსების და სხვადასხვა ჟანრის ლიტერატურული, სამეცნიერო-ტექნიკური, პოპულარული, საცნობარო წიგნების და სხვ. გამოცემას [10-12].

1.2. საგამომცემლო მარკეტინგის სისტემის როლები და ფუნქციები

საგამომცემლო მარკეტინგი (*Publishing Marketing*) – არის გამომცემლობის პროდუქტებთან დაკავშირებული ყველა საკითხის გამოკვლევის (შესწავლის) აქტივობათა ერთობლიობა [14-16].

ეს პროცესებია:

- მომხმარებელთა კვლევა;
- ბაზარზე მისი ქცევის მოტივების კვლევა;
- გამომცემლობის საკუთარი ბაზრის ანალიზი;
- პროდუქციის რეალიზაციის ფორმების და არხების ანალიზი;
- საქონელბრუნვის მოცულობის ანალიზი;
- კონკურენტის შესწავლა, კონკურენციის ფორმებისა და

დონის განსაზღვრა;

- სარეკლამო საქმიანობის შესწავლა;
- წიგნების ბაზარზე პოპულარიზაციის ყველაზე ეფექტური

გზების განსაზღვრა;

- საკუთარი <ნიშის> ბაზრის შესწავლა, სადაც გამომცემლობას კონკურენტებთან შედარებით საუკეთესო შესაძლებლობები აქვს თავისი პროდუქციის გასაყიდად.

ამერიკელი მეცნიერის, ფილიპ კოტლერის (Philip Kotler) განმარტებით, მარკეტინგი ნებისმიერი პროდუქციის საწარმოო ფირმის ბიზნესის ფილოსოფიაა [14], ხოლო მარკეტინგული პროცესები საბაზრო მოთხოვნილებათა გამოკვლევის, წარმოების დაგეგმვის, პროდუქციის წარმოებისა და რეალიზაციის, საწარმოო რესურსების, მათ შორის კადრების მართვის და სხვა ორგანიზაციული და ტექნოლოგიური პროცესების ერთობლიობაა. ასეთი ბიზნესპროცესების ავტომატიზაცია, მათი წარმოებაში დანერგვა და ეფექტური გამოყენება კი მეტად აქტუალური საკითხია როგორც საზღვარგარეთ, ასევე ჩვენ ქვეყანაში [16, 17].

საგამომცემლო მარკეტინგი რთული ორგანიზაციული სისტემაა სხვადასხვა სახის ტექსტური, გრაფიკული თუ ელექტრონული პროდუქციის გამოსაცემად. მრავალფეროვანია ამ სისტემაში ავტორსა და მკითხველს შორის მოღვაწე ადამიანთა საქმიანობა, დაწყებული წიგნის, სტატიის თუ თარგმანის შექმნის კონცეფციიდან, მის საბოლოო გამოცემამდე და დაინტერესებულ პირთან (მკითხველთან) მიტანამდე.

1.1 ნახაზზე წარმოდგენილია საგამომცემლო მარკეტინგის ბიზნესპროცესებში მონაწილე როლები და მათი ფუნქციები. ასეთი ზოგადი სქემა აგებულია პროცესების უნიფიცირებული მოდელირების ენის (UML) საფუძველზე Visual Studio .NET პლატფორმის გამოყენებით [18,19]. იგი ემყარება სისტემების ობიექტ- და პროცეს-ორიენტირებული ანალიზისა და პროექტირების მეთოდებს, რეალიზაცია კი სერვის-ორიენტირებულ არქიტექტურას.

როლები და ფუნქციები ასე შეიძლება წარმოვადგინოთ:

ავტორ(ებ)ი: ნაშრომის კონტენტის (შინაარსის) შექმნა;

გამომცემელი: კონტრაქტის გაფორმება ავტორთან; ნაშრომის მომზადება გამოსაცემად; სარედაქციო საქმიანობის დაყოფა სხვადასხვა სახის რედაქტორზე. მუშაობა ტიპოგრაფიებთან წიგნის ეგზემპლარის შესაქმნელად; მარკეტინგის ბიზნესპროცესების კონტროლი; გამომცემლობის სტრატეგიის წარმართვა და სხვ.;

რედაქტორ(ებ)ი: მუშაობა ავტორებთან ნაშრომის გასაუმჯობესებლად; ტექსტის, დიზაინის, წარმოების, რეკლამის, გაყიდვების, დისტრიბუციის, კონტრაქტების, უფლებების (მაგალითად, თარგმანი, ლიცენზირება) და ადმინისტრირების საკითხებზე;

აგენტ(ებ)ი: ავტორების და მათ ნაშრომთა ხელნაწერების ძებნა; ნაშრომების შეფასება მომგებიანობაზე გამომცემლებისათვის; ავტორებთან მუშაობა მათი წარმოჩენის და ინტერესების დასაცავად;

იურიტი: ავტორთან და გამომცემელთან მუშაობა იურიდიული განხრით კონტრაქტის შესათანხმებლად;

ინფორმაციის დამმუშავებელი: ნამუშევრების ელექტრონულად მომზადება და მიწოდება გამომცემელზე;

დისტრიბუტორ(ებ)ი: მზა პროდუქტის შეფუთვა და გაგზავნა მომხმარებლებთან (მაგალითად, წიგნის გამყიდველებთან, ბიბლიოთეკებთან, სასკოლო სისტემებთან), რომლებიც ყიდულობენ კონტენტს (შემდგომი გაყიდვის უფლებით მოგების მისაღებად ან გასაცემი სერვისის სახით).

და ა.შ. (შეიძლება იყოს სხვა დამხმარე როლებიც, რაც დამოკიდებულია კონკრეტული გამომცემლობის ორგანიზაციულ სტრუქტურასა და მათ ამოცანებზე).

დასასრულ, შეიძლება აღინიშნოს, რომ დღეისათვის საგამომცემლო მარკეტინგი თითქმის მთლიანად ციფრულია. უფრო მეტიც, ის თამაშობს მნიშვნელოვან როლს ქაღალდისა და ციფრული ტექნოლოგიის დასაბალანსებლად.

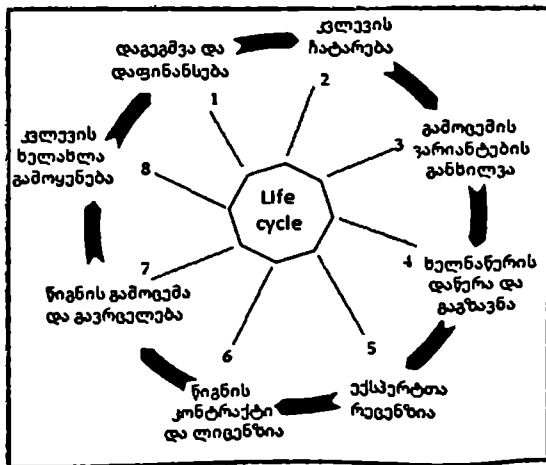
1.3. წიგნების გამოცემის პროცესის სასიცოცხლო ციკლი

დღეისათვის ძალზე აქტუალური და ფართო სპექტრით გამოყენებადია ტერმინი „ციფრული გამოცემა“ (ციფრული ტექნოლოგიების ბაზაზე). შექმნილია და ფუნქციონირებს არაერთი ასეთი სისტემა, რომელთაგან შეიძლება გამოვყოთ მაგალითად, ღია წვდომის წიგნების ინსტრუმენტარიუმი (Open Access Books Toolkit OAPEN) [20]. იგი მოიცავს კონკრეტულ თემებს, რომლებიც დაკავშირებულია ღია წვდომის წიგნებთან. თითოეული სტატია გვთავაზობს სწრაფ და მოკლე შესავალს ღია ხელმისაწვდომობის წიგნის გამოცემის კონკრეტულ ასპექტზე. ინსტრუმენტთა ნაკრებს

აქვს აგრეთვე მიმთითებლის ფუნქციონალი ინდექსირებისათვის: ასრულებს სტატიაში მითითებული წყაროების სიის მოპოვებას, მათ წაკითხვას და ბმულებს საკვანძო ტერმინების განმარტებებთან.

განვიხილოთ იგი დეტალურად.

რას წარმოადგენს წიგნის შინაარსის გამოცემის პროცესი (Content Publishing Workflow) ? იგი არის ყველა ქმედების ერთობლიობა, რომელიც უნდა განხორციელდეს იმისათვის, რომ შინაარსის ნაწილი იდეის სტადიიდან გადავიდეს ნაბეჭდ ეგზემპლარში ან ციფრულ პროდუქტში (წიგნის მფლობელის ვებსაიტზე). 1.2 ნახაზზე მოცემულია წიგნის გამოცემის სასიცოცხლო ციკლის ტიპური მოდელი, რომელიც რვა ეტაპისგან შედგება [20].



ნახ.1.2. წიგნის გამოცემის სასიცოცხლო ციკლი

1) დაგეგმვა და დაფინანსება.

სასიცოცხლო ციკლის ეს ეტაპი მოიცავს სტატიებს, რომლებიც შეიძლება იყოს შესაბამისი ავტორისთვის. იგი გეგმავს თავის

კვლევას და კვლევის შესაძლო შედეგებს. გარდა ამისა, გაირკვევა დაინტერესებული პირები ან ორგანიზაციები, დაფინანსების წყაროები. გამოცემის რომელი ბიზნესმოდელით არის მისაღები ამ ღია ხელმისაწვდომობის წიგნის გამოშვება.

იმის გამო, რომ ღია წვდომის ელექტრონული წიგნები თავისუფლად ხელმისაწვდომია ყველასთვის წასაკითხად, საჭიროა დამატებითი შემოსავლის წყაროები გამოქვეყნების ხარჯების დასაფარად. ამისათვის გამომცემლები იყენებენ სხვადასხვა ბიზნესმოდელს წიგნების მხარდასაჭერად. ზოგიერთი მოდელი მოითხოვს საფასურის გადახდას ავტორის, დამფინანსებლის ან დაწესებულების მიერ, მაგრამ ბევრი არა. ავტორმა უნდა იცოდეს, დაერიცხება თუ არა საფასური, გარანტირებულია თუ არა წიგნი ღია წვდომისათვის, იქნება თუ არა წიგნის ღია წვდომა გამოქვეყნების-თანავე ან მოგვიანებით.

ღია წვდომის წიგნებისთვის არსებობს *მწვანე, ოქროს და ბრილიანტის მოდელები*. ისინი განსხვავდება ერთმანეთისგან გამოყენების გარკვეული პირობებით და შეზღუდვებით. კერძოდ:

- *მწვანე* პუბლიკაცია არქივდება ონლაინ, მაგალითად, საცავში, არ შეიცავს გამომცემლის მიერ შესრულებულ სამუშაოებს (ასლების რედაქტირება, კორექტირება, ტიპაჟირება, ინდექსირება, მეტამონაცემების თეგირება, მარკეტინგი ან განაწილება). იგი შეიძლება არ იყოს დაფიქსირებული გამომცემლის ვებგვერდზე. მასზე წვდომა თავისუფალია, მაგრამ ზოგჯერ მხოლოდ გარკვეული გადასახადის შემდეგ. არ ქვეყნდება Creative Commons ლიცენზიით. მწვანე ღია წვდომას ასევე მოიხსენიებენ, როგორც „თვითარქივირებას“;

- *ოქროს* მოდელი - დაუყოვნებელი ღია წვდომის პუბლიკაციაა გამომცემლის მიერ შექმნილ გამოცემაში. იგი ხელმისაწვდომია გამომცემლის ონლაინ გამოცემის პლატფორმის მეშვეობით. წიგნი გამოქვეყნებულია Creative Commons ლიცენზიით, რომელიც

იძლევა ხელახლა გამოყენების ნებართვას, ზოგიერთ შემთხვევაში ის იბეგრება საკომისიოთი (მას უწოდებენ „წიგნის დამუშავების საფასურს -BPC");

ბრილიანტის (ან პლატინის) მოდელი – მსგავსა ოქროს მოდელის, ოღონდ ღია წვდომაა *საფასურის გარეშე*. მას იყენებენ გამომცემლები, რომლებიც ავტორებს არ უხდიან საფასურს ღია წვდომის გამოქვეყნებისთვის. არსებობს უფასო (Libre) წაკითხვა, მაგრამ შეზღუდვით ხელახალი გამოყენებისათვის. ან Creative Commons ლიცენზიის გამოყენება – წაკითხვა უფასოა და ხელახალი გამოყენებაც ნებადართულია;

არსებობს აგრეთვე სხვა სახის მოდელებიც, არა ღია წვდომის წიგნებისთვისაც (ბრინჯაო, ნაცრისფერი, შავი და ა.შ.) [20].

2) **კვლევის ჩატარება** – ამ ეტაპზე ხდება ავტორის მიერ თემის შესაბამისი (რელავანტური) სხვა სტატიების ან მასალის განხილვა, რომლებიც მოიპოვება ინტერნეტ წყაროებშიიგი შეისწავლია მათ ანალიზის საშუალებით;

3) **გამოქვეყნების ვარიანტების განხილვა**. ამ ეტაპზე ავტორი იღებს გადაწყვეტილებას – რატომ გამოაქვეყნოს ღია წვდომის წიგნი; როგორ მოძებნოს ღია წვდომის წიგნის გამომცემელი; აირჩიოს თუ არა Creative Commons ლიცენზია გამოსაქვეყნებლად და ა.შ.;

4) **კონტენტის (შინაარსის) დაწერა და გაგზავნა** ღია წვდომის წიგნების გამომცემელთან. ამ ეტაპზე იწერება კონტრაქტი მათ შორის და საავტორო უფლებები;

5) **ექსპერტთა რეგენზიები** – კოლეგების მიერ წიგნის მიმოხილვა და შეფასება, ხარისხის კონტროლის მიზნით. ზოგიერთი ღია წვდომის გამომცემელი იყენებს ხარისხის კონტროლის უმაღლეს სტანდარტებს, ზოგიერთი არაღია წვდომის გამომცემელი აქვეყნებს დაბალი ხარისხის ნამუშევრებს და პირიქით. საჭიროა ავტორმა

გაითვალისწინოს წიგნის ღია წვდომის გამოქვეყნებისას გამომცემლის სტატუსი;

6) წიგნის კონტრაქტი და ლიცენზია. Creative Commons ლიცენზიის არჩევა; კონტრაქტის და საავტორო უფლებების დაფიქსირება; მესამე მხარის ნებართვების მიღება და ა.შ.;

7) წიგნის გამოცემა და გავრცელება. კვლევის სასიცოცხლო ციკლის ეს ეტაპი მოიცავს არსებული წესების საფუძველზე ღია ხელმისაწვდომობის წიგნების ბეჭდვითი და ციფრული ვერსიების გამოცემას და ელექტრონული ვერსის განთავსებას ვებგვერდზე, მათი გავრცელების საკითხის მოგვარებას. კონკრეტულად ასეთი საკითხებია გადასაწყვეტი:

- ციფრული და ბეჭდური გამოცემა;
- გავრცელება და აღმოჩენადობა (როგორ იძებნება იგი ინტერნეტში);

გამოცემული ღია ხელმისაწვდომობის წიგნის ხელახლა გამოიყენება სხვა მკვლევარების მიერ და ამ ინფორმაციის დაფიქსირება (მაგალითად, რეგისტრაციის საშუალებით წიგნის მისაღებად);

- **მარკეტინგი** - ესაა გამომცემლის მიერ მწერლის საქმიანობის წარმოდგენა ბაზარზე და მისი პოპულარიზაცია. ეს პროცესი ზრდის ხელმისაწვდომობას და გაზიარების შესაძლებლობას, რაც აძლიერებს მწერლის/მეცნიერის ნამუშევრის აღმოჩენას და გავრცელებას. იმის გასაგებად, თუ როგორ ხდება წიგნის ხელმისაწვდომობა, მნიშვნელოვანია მწერლის კომუნიკაცია გამომცემელთან, ციფრული ფორმატების და პლატფორმების შესახებ, სადაც წიგნი განთავსდება. ციფრულ ეპოქაში გამომცემლები უფრო მეტად ეყრდნობიან ონლაინ არხებს, რათა მიაღწიონ უფრო ფართო გლობალურ მკითხველს და გამოიყენონ მიზნობრივი მარკეტინგული ტექნიკა. გამომცემლის გლობალური წვდომა შეიძლება გაძლიერდეს ვებსაიტის ეფექტური საძიებო სისტემის ოპტიმიზაციით

გამომცემლები ქმნიან მეტამონაცემებს წიგნის ბიბლიოგრაფიული ინფორმაციით, როგორიცაა სათაური, რეზიუმე, გამოცემის ადგილი, თარიღი, მოცულობა, გამომცემლობა და ა.შ., ასევე მონაცემები ავტორის შესახებ. მეტამონაცემები გადაეცემა დაინტერესებულ მხარეებს წიგნის გამოცემის ჯაჭვში, როგორიცაა სერვისის პროვაიდერები, პლატფორმები და ინსტრუმენტები, მკვლევარები, დამფინანსებლები, ბიბლიოთეკარები, მონაცემთა კურატორები და საცავები, რათა სხვებმა აღმოაჩინონ ავტორის ნამუშევარი..

წიგნის აღმოჩენის გასაუმჯობესებლად გამომცემელი ავტორის ნაშრომს მიანიჭებს PID-ს (Publisher Identification). მაგალითად, DOI და ORCID.

ციფრული ობიექტების იდენტიფიკატორები (DOI) არის უნიკალური იდენტიფიკატორი, რომელიც მიეცემა წიგნს, ჟურნალს, სტატიას და ა.შ., იგი აუცილებელია, თუ წიგნის აღმოჩენა ხდება ბიბლიოთეკის აღმოჩენის სერვისებითა და ინტერნეტის საძიებო სისტემებით;

ღია მკვლევარის და კონტრიბუტორის იდენტიფიკატორი (ORCID) ეხმარება ავტორს გაურკვევლობაში. მათ შეუძლიათ დარეგისტრირდნენ ORCID-ზე და შემდეგ მოითხოვონ თავიანთი ნამუშევრები Crossref-ის საშუალებით.

Crossref არის არაკომერციული ღია ციფრული ინფრასტრუქტურის ორგანიზაცია (ასოციაცია) გლობალური მეცნიერული კვლევითი საზოგადოებისთვის [21]. ცოდნის უნიკალური და მუდმივი ჩაწერა და დაკავშირება ღია მეტამონაცემების და იდენტიფიკატორების (DOI, ORCID) მეშვეობით ყველა კვლევის ობიექტისთვის, გამოიყენება Scopus, Google Scholar-ში და სხვ.

8) *კვლევის ხელახლა გამოყენება* - სხვა მკვლევარების მიერ შესაძლებელია თუ წიგნს (ან სხვა ნაშრომს) ექნება Creative Commons (CC) ლიცენზია. ზოგიერთი გამომცემელი იყენებს საკუთარ

ლიცენზიას. ამ შემთხვევაში მკითხველმა უნდა შეამოწმოს რა არის მასში ნებადართული. ლიცენზიის გარეშე წიგნი არ განიხილება როგორც „ღია წვდომის“. მისი ჩამოტვირთვა უფასოა, მაგრამ არავის აქვს უფლება მისი ასლების გაკეთების, გაყიდვის ან სხვა გზით გავრცელების, საავტორო უფლებების მფლობელის წინასწარი ნებართვის გარეშე. ესაა რაც გულისხმობს ფრაზას „ყველა უფლება დაცულია“.

1.4. ERP სისტემა საგამომცემლო ინდუსტრიაში

წინა პარაგრაფებში განვიხილეთ საგამომცემლო ინდუსტრიის მიზანი (საგამომცემლო პროდუქციის წარმოება), მისი სტრუქტურა, ბიზნესპროცესები და გამომცემლობის მარკეტინგის სასიცოცხლო ციკლი. იგი ნამდვილად მიეკუთვნება რთული და დიდი სისტემების კლასს, რაც მოითხოვს მისი მართვის პროცესების შემდგომ სრულყოფას, რესურსების ოპტიმიზაციას და კორპორაციული შემოსავლების გაზრდას.

ამ თვალსაზრისით საგამომცემლო ინდუსტრიის კომპანიებისთვის ძალზე მნიშვნელოვანია ERP (Enterprise Resource Planning - საწარმოს რესურსების დაგეგმვის) სისტემების გამოყენება [15, 22].

ERP სისტემა შეიძლება გამოყენებულ იქნას ძირითადი ბიზნესპროცესების რეალურ დროში ინტეგრირებული მართვისათვის, პროგრამული უზრუნველყოფისა და ტექნოლოგიების დახმარებით. იგი ბიზნესის მართვის პროგრამული აპლიკაციის კატეგორია, რომელიც ორგანიზაციას შეუძლია გამოიყენოს სამუშაო ნაკადების მონაცემთა შეგროვების, შენახვის, მართვისა და ინტერპრეტაციისთვის. ERP სისტემები შეიძლება იყოს ლოკალური ან ღრუბელზე (Cloud-ზე) დაფუძნებული. ასეთი აპლიკაციები აქტუალური გახდა ბოლო წლებში გაზრდილი ეფექტიანობით, ინფორმაციის ადვილად წვდომის გამო ინტერნეტში.

ზოგადად, ERP-ის ძირითადი მახასიათებლები შეიძლება საგამომცემლო ინდუსტრიისთვისაც ასე აისახოს [22]:

- *ინტეგრაცია*: ERP პროგრამული უზრუნველყოფის გამოყენებით შესაძლებელია მონაცემების საფუძველზე ამოცანების შესრულება, მონაცემთა ანალიზი და მონიტორინგის განხორციელება. მონაცემთა შეგროვება, შენახვა და ანალიზი ხდება ERP სისტემებით ორგანიზაციის ყველა დეპარტამენტში, რათა უზრუნველყოფილი იყოს შეუფერხებელი კომუნიკაცია;

- *ავტომატიზაცია*: ERP საფუძვლიანად ასრულებს შრომატევადი სამუშაოების ავტომატიზაციას. ეს ხელს უწყობს მონაცემების უშეცდომოდ გაგზავნას სისტემის ერთი მოდულიდან მეორეში;

- *აღრიცხვა (ბუღალტერია)*: ბიზნესის მნიშვნელოვანი ფუნქციაა ფინანსური მენეჯმენტი. ყველა მნიშვნელოვანი ფინანსური ასპექტი, როგორცაა საკრედიტო გადასახდელები და დებიტორული ანგარიშები, ძირითადი საშუალებები, რისკები და გადასახადები, ყველა იმართება ERP სისტემებით;

- *CRM (Customer Relationship Management)*: მოდულის გამოყენებით შესაძლებელია წვდომა შეკვეთების ისტორიაზე, კლიენტის საკონტაქტო ინფორმაციით, შესყიდვების შეკვეთის ინფორმაციაზე, პოტენციალური კლიენტების მენეჯმენტზე და მარკეტინგის ავტომატიზაციაზე. ამ ყველაფერმა შეიძლება ხელი შეუწყოს მომხმარებელთა ლოიალობის გაზრდას და კლიენტთა შენარჩუნებას;

- *HRM (Human Resource Management)*: ERP სისტემის გამოყენებით, HR მოდულებს შეუძლია მართოს თანამშრომელთა დაქირავება, სახელფასო ანგარიშები, მათი ადაპტირება, თანამშრომლების შეღავათები და სხვ. ასეთი გადაწყვეტა ასრულებს პერსონალის მართვის რთული ფუნქციების ავტომატიზაციას, როგორცაა მაგალითად, საგადასახადო დაქვითვები, რათა დაიზოგოს დრო და რესურსები და შემცირდეს შეცდომების ალბათობა;

- *SCM (Supply Chain Management)*: როგორც მიწოდების ჯაჭვის მოდულის ნაწილი, ხდება ნედლეულის შესყიდვა და მარკეტინგული არხების შერჩევა მზა პროდუქციის პოპულარიზაციისთვის. როგორც მარაგების მენეჯმენტის ნაწილი, ის განსაზღვრავს სასაწყობო რაოდენობას, უზრუნველყოფს შევსებას, ასრულებს წარმოების პროცესების ოპტიმიზაციას მოთხოვნებისა და მიწოდების ბალანსის უზრუნველსაყოფად, აგრეთვე უზრუნველყოფს მზა პროდუქციის ტრანსპორტირებას და დისტრიბუციას;

- *გაყიდვები და მარკეტინგი*: გაყიდვებისა და მარკეტინგის მოდული საშუალებას იძლევა განახორციელონ კლიენტებთან დაკავშირება, შეკვეთებზე ყურადღების მიქცევა, პროდუქციისა და ინვოისების მიწოდება, გადასახდების მიღება და დაბრუნებული შემოსავლების დამუშავება.

- *თვალყურის დევნება და ბილვადობა*: ეს არის ERP-ის ერთ-ერთი საუკეთესო შესაძლებლობა – ხელი შეუწყოს მიწოდების ჯაჭვის სიცოცხლისუნარიანობას,

განხილული მახასიათებლების გათვალისწინებით ERP სისტემას შეუძლია საგამომცემლო ბიზნესის და მარკეტინგის ფუნქციების სწრაფად და მოქნილად განხორციელება.

სარედაქციო მენეჯმენტი (Editorial Management -EM) საგამომცემლო პროცესების სამუშაო ნაკადების (Workflows) მართვის სისტემაა [19]. ეს პროცესები წიგნის გამოცემის სასიცოცხლო ციკლის პარაგრაფში განვიხილეთ. ახლა მაგალითისთვის, გამოვყოთ ის ამოცანები, რომელთა გადაწყვეტა ავტომატიზებულ რეჟიმში შეუძლია ERP სისტემის მოდულებს [22]:

- *ჰონორარი და გამოთვლების ანგარიში* : თანხის გამოთვლა, რომელსაც გამომცემელი უხდის ავტორს თავისი წიგნის გამოქვეყნების უფლების სანაცვლოდ;

– *სათაურის სასიცოცხლო ციკლის მართვა*: საგამომცემლო ინდუსტრიის შემთხვევაში, *წიგნის სათაური* არის ინტელექტუალური საკუთრება და მათი მართვა უკიდურესი სიფრთხილით უნდა მოხდეს. სათაურის მენეჯერი შეიძლება იყოს საწყისი ნაბიჯი, რომელიც დაეხმარება ბიუჯეტის დაგეგმვის პროცესს არასრული ინფორმაციის პირობებში სათაურთან დაკავშირებით;

– *გაყიდვების ფასის მენეჯმენტი*: ERP აადვილებს ფასების რთული სტრუქტურების დაშლას. ინტელექტუალური, ბრაუზერზე დაფუძნებული პროგრამული უზრუნველყოფა ამარტივებს ფასების რთული სტრუქტურების აგებას და გვთავაზობს ყველა პროდუქტის ფასის მკაფიო, საფუძვლიან მიმოხილვას;

– *წარმოების მენეჯმენტი*: წარმოება პირდაპირ დამოკიდებულია ბაზარზე პროდუქციის მოთხოვნაზე. პროდუქტები დროულად უნდა მიეწოდოს, რათა სათაურები განხორციელდეს გრაფიკის მიხედვით. საჭიროა დარწმუნება, რომ ყველაფერი მიწოდებულია გრაფიკით. ასევე უნდა დაინერგოს პროცედურები, რათა თავიდან იქნას აცილებული გადაზიდვის შეფერხების ან მიწოდების პროცესის რაიმე შეცდომა;

– *ხელმოწერების მართვა*: ნივთს ან ნივთების ჯგუფს, რომელიც შეიძლება გამოყენებულ იქნას გამოწერისთვის, როგორიცაა რეგულარული სერვისები, გამოყენების საფასური და ერთჯერადი ხარჯები, ეწოდება სააბონენტო გეგმა. გამოწერა შეიძლება შეიცავდეს რამდენიმე გეგმას. ERP პროგრამული უზრუნველყოფის გამოყენებით შესაძლებელია ხელმოწერების მარტივად მართვა;

– *გამოქვეყნების სამუშაო ნაკადის მართვა (Publishing Workflow Management)*: სამუშაო პროცესი ეხება ყველა პროცედურას, რომელიც უნდა დავიცვათ სასურველი შედეგის მისაღებად. ანალოგიურად, საგამომცემლო კომპანიებმა უნდა უზრუნველყონ, რომ მათი სამუშაო პროცესი ან ოპერაციები მხარს უჭერდეს მათ საბოლოო დანიშნულებას, რაც არის სტანდარტული და ხარისხიანი

შინაარსის გამოქვეყნება. შესაბამისად, საგამომცემლო სამუშაო პროცესი შედგება ყველაფრისგან, რაც უნდა მოხდეს გამოქვეყნებამდე.

ადამიანებსა და მონაცემებს შორის დამაკავშირებელი მოქმედებით, საწარმოს სისტემა, როგორიცაა ERP, შეუძლია მნიშვნელოვნად გააუმჯობესოს სამუშაო ნაკადის მართვა. თუმცა, ასეთი გადაწყვეტილებების წარმატებით გამოყენება მოითხოვს ფრთხილად დაგეგმვასა და მომზადებას. პროცესების დასაპროექტებლად და სამუშაო ნაკადების მოდელების შესაცვლელად ERP პროგრამულ უზრუნველყოფაში საჭიროა საფუძვლიანად იქნას შესწავლილი საკუთარი ბიზნესოპერაციები.

დამატებითი ინფორმაცია ERP სისტემაზე გადმოცემულია წიგნის *დანართში-2*, სადაც ზოგიერთ ძირითად ცნებებთან ერთად ასახულია სტუ-ს ინფორმატიკისა და მართვის სისტემების ფაკულტეტზე ჩვენ მიერ ამ მიმართულებით შემუშავებული ინოვაციური საკითხები [15, 43, 105].

1.5. მარკეტინგული პროცესის იმიტაციური მოდელი პეტრის ფერადი ქსელით (CPN)

დანიელი მეცნიერ-ინჟინრების მიერ შექმნილი პეტრის ფერადი ქსელების გრაფო-ანალიზური CPN-ინსტრუმენტი იყენებს ობიექტ-ორიენტირებული, ვიზუალური დაპროგრამების პრინციპებს [23-25]. მისი ენა CPN ML საშუალებას იძლევა აღიწეროს ქსელის ფერადი კომპონენტები (მარკერები), ცვლადები, კონსტანტები და თვით პოზიციების, გადასასვლელებისა და რკალების ტექსტური აღწერები, რაც ერთგვარ კომფორტს ქმნის ქსელის წასაკითხად და გასაგებად [17].

1.3 ნახაზებზე წარმოდგენილია ჩვენ მიერ აგებული პეტრის ფერადი ქსელის ფრაგმენტი (მოდელირების ეტაპი). საგამომცემლო საწარმოო ფირმის მარკეტინგული პროცესების იმიტაციური მოდელირების შესაძლებლობით.

მოდელი აგებულია CPN Tools რედაქტორში, რომელიც სტოქასტიკური პროცესების იმიტაციის საშუალებას იძლევა (მასში არსებული შემთხვევით რიცხვთა გენერატორის დახმარებით). ექსპერიმენტები ტარდება (ანალიზის ეტაპი) საწყისი მონაცემების სხვადასხვა ვარიანტისთვის ოპტიმალური შედეგების განსაზღვრის მიზნით (დეტალურად იხ. დანართი-2).

მოკლედ გავანალიზოთ 1.3 ნახაზზე მოცემული პეტრის ქსელის მუშაობის პრინციპი:

- საწყის მდგომარეობაში (ნახ.1.3-ა) პოზიცია „Kontraqtebi“ შეიცავს INTxDATA ტიპის ფერად მარკერთა 5-ელემენტთან სიმრავლეს (საინიციალიზაციო მარკირება): [1'(1,"kontraqti_1"), 1'(2, „ kontraqti_2"), 1'(3, „ kontraqti_3"), 1'(4, „ kontraqti_4"), 1'(5, „#### ")]. აქ ბოლო, მე-5 ელემენტი შეესაბამება დასასრულის იდენტიფიკაციას - stop; კოეფიციენტი „1“-იანი ყოველი ელემენტის დასაწყისში, მიუთითებს, რომ პოზიციაშია არაუმეტეს 1 ცალი მოცემული ფერის მონაცემი (ანუ არსებობს მხოლოდ ერთი კონტრაქტი ნომრით „კონტრაქტი_1“, რომლის ფერია - რიგითი ნომერი 1). ამ შემთხვევაში გვაქვს მონაცემთა ელემენტების სიმრავლე (დანართი-2);

- ქვემოთ, მეორე პოზიცია „Mza produkcia“ (ნახ.1.3-ა) ან Mza Tsignebi (ნახ.1.3-ბ, მაგალითად ნაბეჭდი წიგნები) შედგება 19 ელემენტისგან ($5+1+2+1+4+3+2+1$), რომლებიც 7 სხვადასხვა (მარკერების ფერის) დამზადებული პროდუქტთა პაკეტების რაოდენობაა, ანუ იგი ასახავს *მულტისიმრავლეს* (დანართი-2). პროცესების შესრულების დრო გადასავლელთან აისახება დაყოვნების სიმბოლოს და დროის ერთეულის (მაგალითად, @+12, @+wait) მითითებით, სადაც wait წინასწარ განსაზღვრული კონსტანტაა. ამავე ნახაზზე ასახულია არადეტერმინირებული ლოგიკური გამოსახულება (პირობის ბლოკი) ფერადი პეტრის

ქსელის რკალებზე, რომელიც გადასასვლელთა გაშვების სხვადასხვა პირობებს და შედეგებს ასახავს, ანუ ლოგიკური პირობის ჭეშმარიტებისას (true) გადასასვლელს განსხვავებული მნიშვნელობა მიეწოდება (ან გადასასვლელიდან განსხვავებული მნიშვნელობა გამოვა), მცდარობისას (false) კი – განსხვავებული;

მაგალითად, გადასასვლელს „Tsignebis Partiiis gadagzavna“ გამოსასვლელ რკალზე აქვს ლოგიკური პირობა - თუ გაგზავნილი პროდუქციის ნომერი (n) ემთხვევა კლიენტის (მაგალითად, წიგნების მაღაზია) კონტრაქტით განსაზღვრულ მისაღებ პროდუქციის პარტიის ნომერს (k), მაშინ გვაქვს "true" (ანუ if (n=k) then ...), წინააღმდეგ შემთხვევაში "false", რაც იმას ნიშნავს, რომ საჭირო პროდუქცია არაა მიღებული. თუ ყველაფერი წესრიგშია, მაშინ მიმღები (წიგნების მაღაზია) უგზავნის მწარმოებელს (გამომცემლობას) შეტყობინებას გადასასვლელით "Dasturis gamogzavna".

პროდუქციის და შეტყობინების გადაცემათა ქსელში შემთხვევითი პროცესის არსებობა განპირობებულია დაყოვნების ცვლადი დროით, რაც აისახება მაგალითად, colset NetDelay=int with 25..75, fun DEL() =NetDelay.ran() random-ფუნქციით.

ლოგიკური პირობის მნიშვნელობა სხვადასხვა შემთხვევებში სხვადასხვანაირად განისაზღვრება. ინტერაქტიულ სიმულატორებში ჭეშმარიტება-მცდარობას თავად მომხმარებელი განსაზღვრავს, ავტომატური სიმულაციისას კი – შემთხვევით რიცხვთა გენერატორი. 1.3-ბ ნახაზზე ნაჩვენებია პროდუქციის პარტიის გადაგზავნის გადასასვლელის აქტიური მდგომარეობა. აქ მარკერები არის A და SP პოზიციებშიც. ამ გადასასვლელიდან B-პოზიციაში შემავალი რკალი ლოგიკურ პირობას აკონტროლებს, ანუ დასაშვებია ორი შემთხვევა:

TP+=(Tsignebis_partiis_gadagzavna, <n=1,p="Prod1", success=true>),

TP -= (Tsignebis _partiis_gadagzavna, <n=1,p="Prod1", success=false>).

შეიძლება დავასკვნათ, რომ პეტრის ფერადი ქსელებისა და ობიექტ-ორიენტირებული დაპროგრამების თეორიის საკითხები (მაგალითად, იერარქიულობა, მოდულურობა – დიდი სისტემების მოდელირებისთვის), იზომორფიზმის თვალსაზრისით, კარგადაა შერწყმული ერთმანეთთან. ეს კი განაპირობებს განხილული კვლევის მეთოდოლოგიის გამოყენების დიდ პრაქტიკულ ღირებულებას მრავალ საპრობლემო სფეროში, განსაკუთრებით ბიზნესისა და მარკეტინგის მენეჯმენტის ამოცანებისათვის.

დანართში-2 გადმოცემულია პეტრის ქსელის იმიტაციური მოდელირების CPN Tools-ის დამატებითი ინფორმაცია.

1.6. კვლევის მიზანი და შედეგების ინოვაციურობა

➤ **მიზანი.** მონოგრაფიის კვლევის მთავარი მიზანი საგამომცემლო მარკეტინგის, კერძოდ, ტექსტური ინფორმაციის შექმნის პროცესების ხელშეწყობაა [18]. პროექტი მწერლებისთვის, მათი წერის სტილის გათვალისწინებით, წერასა და ბიზნესპროცესების წარმართვაში უნდა იყოს ხელშემწყობი. ჩვენი ერთ-ერთი მიზანია, რომ პროექტში მოდელირებულმა სისტემამ მუშაობის პროცესი გაუმარტივოს მთარგმნელებსაც. ამასთან, პროექტის ინოვაციურობიდან გამომდინარე, კვლევათა შედეგები სასარგებლო უნდა იყოს საგამომცემლო სფეროსათვის.

ამასთანავე საჭიროა შეიქმნას პროექტის პროგრამული რეალიზაციის შედეგების საფუძელზე ჩამოყალიბებული, სრულყოფილი დოკუმენტირებული ინფორმაცია სისტემის არქიტექტურის შესახებ, ინფრასტრუქტურა მოიცავს პროგრამული აპლიკაციის არქიტექტურას, მისი დიზაინის, კლასებისა და მოდელების სახით. უნდა გაანალიზდეს ის თუ რა კომპიუტერული რესურსები,

პროგრამული ენები, მონაცემთა ბაზები, თუ სხვა ტექნიკური და ტექნოლოგიური საშუალებებია საჭირო მიზნის მისაღწევად.

საპროექტო ნაშრომი სამი ძირითადი ეტაპითაა მოცემული:

1) *პირველი ეტაპი* ეხება იმის განსაზღვრას, თუ რა *ტექნიკური საშუალებებია* საჭირო კვლევისათვის. ჩატარდა ანალიზი, თუ *მონაცემთა მეცნიერების* (Data science) რა საკითხებია გამოსაყენებელი იმისთვის, რომ შეიქმნას ინოვაციური მიდგომა მოცემული პრობლემების გადასაჭრელად. მანქანური დასწავლის რა ალგორითმები და მიდგომებია საჭირო დასახული მიზნების მისაღწევად. რომლებსაც ექნება როგორც თეორიული სამეცნიერო მნიშვნელობა, ასევე პრაქტიკული ღირებულებაც. რომელ პლატფორმაზე უნდა იყოს შექმნილი ჩვენი პროგრამის Back-End (რომელი სტრუქტურა თუ პროგრამული ენა უნდა იყოს გამოყენებული) ან Front-End (რომელი თანამედროვე Javascript-ის ბიბლიოთეკა შეესაბამება ჩვენს მოთხოვნებს ყველაზე მეტად) კომპონენტები.

ჩატარდა ანალიზი იმის შესახებ, თუ ინფორმაციის რა ნაწილი უნდა ინახებოდეს რელაციურ მონაცემთა ბაზებში (მაგალითად, Ms SQL Server, MySQL, SQLite) და რა ნაწილი NoSQL, ფაილურ ბაზებში (მაგალითად, MongoDB, Redis ან სხვ.). შემუშავებულ იქნა სისტემის მუშაობის დროს ინფორმაციის უსაფრთხოების გეგმა;

2) *მეორე ეტაპზე* ჩამოყალიბდა ინოვაციური მოდელები მანქანური დასწავლის მეთოდების გამოყენებით. განისაზღვრა ბიზნესმოთხოვნები, თუ რა ფუნქციონალები და შესაძლებლობებია საჭირო კვლევით პროექტში დასახული მიზნების მისაღწევად. ამ ეტაპზე ერთ-ერთი მნიშვნელოვანი სიახლე იყო სხვადასხვა კატეგორიის მწერალთა და მეცნიერთა მუშაობის სტილისა და ცოდნის კლასიფიკაცია და მისი ფუნქციონალის მოთხოვნად

გარდაქმნა. ამისთვის ჩატარდა კვლევები და გამოკითხვები, როგორც მეცნიერების სფეროში მომუშავე, აგრეთვე სხვა ტიპის ტექსტურ ინფორმაციასთან მომუშავე პირებთან.

ჩატარებულ იქნა *მონაცემთა მეცნიერების* კვლევები, რის შედეგადაც მივიღეთ ჩვენი პროექტის მნიშვნელოვანი სამეცნიერო სიახლეები, მანქანური დასწავლის ინოვაციური მოდელები, სამეცნიერო ნაშრომთა *თემის* მოდელირებისა და კლასიფიკაციის გზით. ასევე განვახორციელეთ კვლევები და გამოკითხვები წიგნის თარგმნის პროცესთან დაკავშირებით, რადგან პროექტისთვის ძალიან მნიშვნელოვანია მთარგმნელთა საერთო სისტემაში კომფორტული მუშაობა;

3) *მესამე ეტაპზე* – პირველ პუნქტში შერჩეული ტექნიკური საშუალებებითა და მეორე პუნქტში მიღებული კვლევის შედეგებით, შემუშავდა სადემონსტრაციო პროტოტიპი, რომელიც ასახავს ტექსტური ინფორმაციის შექმნის ფუნქციონალს, მისი შესრულების გზებსა და შესაძლებლობებს.

➤ *ინოვაციური შედეგები*. პროექტის კვლევის მთავარი სამეცნიერო სიახლე, მანქანური დასწავლის მოდელებს უკავშირდება. როგორც უკვე აღვნიშნეთ, თემის მოდელირების ალგორითმების წყალობით, შექმნილ იქნა ინოვაციური მოდელები, რომელთა გამოყენების არაერთგზის შესაძლებლობა არსებობს. ერთმანეთს შედარდა სხვადასხვა მეთოდები თემის მოდელირებისას და გამოვიკვლიეთ საუკეთესო მათგანი. შედარებებში მონაწილეობა მიიღო ოთხმა მეთოდმა, ესენია [30, 31, 38]:

1) *დირიხლეს ლატენტური განაწილება* – სტატიის რეზიუმეებისთვის (დაახლოებით 2 მილიონი რეზიუმე);

2) *დირიხლეს ლატენტური განაწილება* – სტატიების სრული ტექსტისთვის (დაახლოებით 400 000-მდე სტატიის სრული ტექსტი);

3) არაუარყოფითი მატრიცის ფაქტორიზაცია – სტატიის რეზიუმეებისთვის (დაახლოებით ორი მილიონი რეზიუმე);

4) არაუარყოფითი მატრიცის ფაქტორიზაცია – სტატიების სრული ტექსტისთვის (დაახლოებით ოთხასი ათასამდე).

უნდა აღინიშნოს, რომ ზემოხსენებული ოთხი მეთოდის გამოყენების მიერ საბოლოო შედეგის მიღებამდე, არაერთგზის მოხდა მათი გაშვება ოპტიმიზაციის მიზნით. ოთხი მეთოდის ერთმანეთთან შედარებისთვის ჩატარდა ერთგვარი შეჯიბრი და მოხდა ორი გამარჯვებული მეთოდიდან მიღებული მონაცემების გაერთიანება. საბოლოოდ, კი მიღებული იქნა, ორი ინოვაციური მოდელი, რომლის საშუალებითაც, მოხდა არსებული სამეცნიერო ნაშრომების რეკატეგორიზაცია. მათი ცალკეული ისევე როგორც სინქრონული გამოყენება, საშუალებას მოგვცემს რომელიმე კატეგორიის სამეცნიერო მიმართულებით და ამ მიმართულებაში კონკრეტული თემით დაინტერესებულ პირებს მათთვის საინტერესო ნაშრომები შევთავაზოთ. მიღებული მეთოდების გამოყენებით მიღებული მოდელები ინოვაციურია და შედეგი წარმოადგენს სამეცნიერო სიახლეს.

აგრეთვე, აღსანიშნავია მანქანური დასწავლის მეორე მეთოდის, *ტექსტის კლასიფიკაციის* გამოყენებაც. სამეცნიერო ნაშრომების კლასიფიკაციის მოდელების შესაქმნელად, გამოყენებულ იქნა მანქანური დასწავლის ორი ყველაზე მთავარი და პოპულარული მეთოდი. აღნიშნული მოდელებია:

1. წრფივი მხარდაჭერის ვექტორული კლასიფიკატორი (The Linear Support Vector Classifier) [21];

2. მულტინომინალური ბაიესის მიამიტი ალგორითმი (Naive Bayes) [26, 27].

აღნიშული მეთოდების გამოყენების შედეგად, მიღებული იქნა მოდელი, რომლის მეშვეობითაც რამდენიმე წინადადების დახმარებით, შესაძლებელია მომენტალურად დადგინდეს ის, თუ რა სფეროს ეხება ან უბრალოდ რა თემაზეა აღნიშნული ტექსტი. ამის მიხედვით, კი შესაძლებელი იქნება, რომ მეცნიერთათვის, მათი მიმართულებისთვის საჭირო ფუნქციონალური და მომხმარებლის ინტერფეისის შეთავაზება მოხდეს.

წიგნის სამეცნიერო სიახლედ ასევე მიგვაჩნია გამოკითხვებისა და სტატიებში გამოყენებული ფუნქციონალის კვლევის შედეგებიც. კვლევის შედეგად, გაგებული იქნა ინფორმაცია იმის შესახებ, თუ რა ფუნქციონალური და ინტერფეისის საჭირო იმისთვის, რომ გამარტივდეს ტექსტურ (არასტრუქტურირებულ) ინფორმაციასთან მომუშავე პირთა მუშაობის გამოცდილება და როგორ უნდა განხორციელდეს ეს ფუნქციონალური სისტემაში. აგრეთვე გაგებულ იქნა იმ ბიზნესპროცესების შესახებ, რომელიც სამეცნიერო ნაშრომებთან მუშაობის დროს მიმდინარეობს, ამ ცონდის საფუძველზე კი ჩამოყალიბდა მათი ავტომატიზაციის მოდელები.

ამასთანავე, კვლევებიდან გამომდინარე, არსებობს ცოდნა იმასთან დაკავშირებით, კონკრეტულად თუ რა ტიპის ფუნქციონალურია გამოყენებული სხვადასხვა კატეგორიის სამეცნიერო ნაშრომში და, აქედან გამომდინარე, ინფორმაცია ამ ფუნქციონალის მნიშვნელობების შესახებ.

ამ მონაცემების საფუძველზე ჩამოყალიბებულია პროტოტიპი, რომლის მეშვეობითაც შესაძლებელია თვალსაჩინოდ მოხდეს იმის დანახვა, თუ რა იგულისხმება, სხვადასხვა ტიპის სამეცნიერო მიმართულებისათვის სპეციალურად მორგებული მომხმარებლის ინტერფეისისა და ფუნქციონალური.

რაც შეეხება ჩვენი ნაშრომის პრაქტიკულ ღირებულებას, სამეცნიერო სიახლიდან გამომდინარე, ეს უკვე ნათელია. მეტად, რომ განვაზოგადოთ, პროექტის ერთ-ერთი მთავარი შედეგი არის ის, რომ ერთის მხრივ კვლევების შედეგად მიღებულია პატენტი პროდუქტისა, სადაც ზემოთ ხსენებული ფუნქციონალი და მათი გადაწყვეტის გზები იქნება ჩართული. სისტემის პრაქტიკული გამოყენება კი ხელს შეუწყობს მსოფლიოში საგამომცემლო მარკეტინგის მოდერნიზაციას და მას უფრო გლობალურს და ხელმისაწვდომს გახდის. პროდუქტი შექმნილი იქნება ნებისმიერი მსურველისთვის, ინოვაციური სიახლეები კი მომხმარებელთა გამოცდილებას ტექსტური ინფორმაციის შექმნისა და მისი გამოყენების მხრივ საგრძნობლად გაამარტივებს, რაც მათ დროისა და რესურსების დაზოგვაში დაეხმარება.

ასევე, როგორც ზემოთ ვახსენეთ, საპროექტო ნაშრომის ერთ-ერთი მიზანია წიგნის წერა შეძლონ ისეთმა ადამიანებმა ვისაც, ჩვეულებისამებრ, ამის უნარი არ აქვთ. მათ უნდა შეეძლოთ ცოდნის, გამოცდილების, ფანტაზიებისა თუ მიღწევების სხვა ადამიანებისათვის ტექსტური ფორმით და მეთოდით გადაცემა.

ასეთი ფუნქციონალის შექმნის პროგრამული უზრუნველყოფა შესაძლებელი გახდა სისტემაში *ხმის ტექსტად გარდაქმნის* ფუნქციონალის ინტეგრაციით (SPEECH-TO-TEXT ტექნოლოგია) [39, 40]. მისი გამოყენებით შესაძლებელია ადამიანებისთვის წერის პროცესი უფრო მარტივი და ხელმისაწვდომი გახდეს.

როგორც უკვე აღვნიშნეთ, კვლევის შედეგებზე დაყრდნობით გამზადდა მცირე პროტოტიპი ტექსტური ინფორმაციის შექმნის ხელშომწყობი ფუნქციონალისათვის. აგრეთვე კვლევების შედეგად მიღებულია ინფორმაცია სფეროში არსებული ბიზნესმოთხოვნილებების შესახებ და აღწერილია ამ მოთხოვნების გადაჭრის გზები.

1.7. კვლევის მეთოდოლოგია, ინტერდისციპლინურობა და არეალი

➤ *პროექტის კვლევის მეთოდოლოგია* დაყოფილია სამ ძირითად ეტაპად, რომელთა თანამიმდევრულად და სწორად წარმართვის შემთხვევაში მისი შედეგები დამაკმაყოფილებელი იქნება. განვიხილოთ თითოეული მათგანი ცალ-ცალკე და ჩამოვყალიბოთ ძირითადი ასპექტები.

პირველ ეტაპზე ჩატარდა ანალიზი იმის შესახებ თუ რა ტექნიკური საშუალებები უნდა იქნას გამოყენებული საჭირო კვლევის განსახორციელებლად, რომ სწორად მოხდეს საპროექტო ნაშრომში მიღებულ მონაცემთა დამუშავება. ანალიზი მოიცავს პროგრამულ-აპარატურული რესურსის განსაზღვრას, რომელიც საჭიროა პროექტის მოთხოვნების შესასრულებლად და მისი ძირითადი კომპონენტების დასადგენად. ამასთანავე, განხილულ იქნა დღეისათვის ბაზარზე არსებული მიდგომები და აპლიკაციები, რომლებიც უზრუნველყოფს რეალიზაციის ახალ კონცეფციებს ტექსტური ინფორმაციის შექმნისა და მისი გამოყენებისათვის.

განსაკუთრებით გამახვილდა ყურადღება *მანქანური დასწავლის მეთოდებსა და ალგორითმებზე*, რომელთა გამოყენებაც მოხდა კვლევაში. აღწერილ იქნა მათი მუშაობის პრინციპი და გამოყენების მიზეზები, კერძოდ, თემის მოდელირებისათვის განვიხილეთ *დირიხლეს ლატენტური განაწილებისა და არაუარყოფითი მატრიცის ფაქტორიზაციის* მეთოდები. ასევე, აღწერილ იქნა *მანქანური დასწავლის კლასიფიკაციის* მეთოდები.

მოთხოვნების მიხედვით, არჩევანი გაკეთდა კლიენტისა და სერვერის მხარეს საჭირო პროგრამულ ინსტრუმენტებზე და სერვისებზე. სერვერის მხარედ იგულისხმება მონაცემთა ბაზები და სასერვერო (Back-End) აპლიკაცია. თანამედროვე მიდგომებით დღეს აპლიკაციებში ორი ტიპის მონაცემთა ბაზა გამოიყენება: SQL –

რელაციური და NoSQL – არარელაციური (იერარქიული). პირველი ტიპის ბაზებში ინახება ისეთი ინფორმაცია, როგორიცაა მომხმარებლები, მათ კავშირები, კატეგორიები, ჯგუფები და ა. შ., ხოლო მეორე ტიპის ბაზები დოკუმენტურია და ინფორმაციას JSON ფორმატით ინახავს. ეს უკანასკნელი კი კომფორტულია ისეთი ინფორმაციის შესანახად, როგორიცაა შეტყობინებები, სიახლეები და თავად წიგნებიც. სხვადასხვა ტიპის ნაშრომების არარელაციურ, NoSQL მონაცემთა ბაზებში განლაგება დადებითად მოქმედებს მათ უსაფრთხოებაზე, ძალზე სწრაფი და კომფორტულია შექმნისა და გამოყენების პროცესის თვალსაზრისით [28, 29]. კლიენტის მხარეს კი, ბრაუზერში გამწვებული ვებ-აპლიკაცია იგულისხმება.

მეორე ეტაპი, როგორც უკვე აღვნიშნეთ, ეხება პროგრამის მოთხოვნების სრულ ანალიზსა და ჩამოყალიბებას. ეს ეტაპი შეიძლება რამდენიმე ნაწილად დავყოთ. პირველი ნაწილი არის მანქანური დასწავლის მეთოდების გამოყენება, ამ მეთოდების ერთმანეთთან შედარება და საუკეთესო მეთოდის გამოყენებით მიღებული მოდელი.

პროექტზე მუშაობის პროცესში, ინტერნეტულ რესურსებში მოძიებულ იქნა სამეცნიერო ნაშრომთა რამდენიმე მილიონიანი მონაცემთა ბაზა, რომელიც გამოყენებულ იქნა *თემის მოდელირების განსაზღვრის ალგორითმებში*, როგორიცაა *დირიხლეს ლატენტური განაწილება* და *არაუარყოფითი მატრიცის ფაქტორიზაცია* [30, 31]. ერთმანეთს შედარდა ამ ორი მეთოდით მიღებული მანქანური დასწავლის მოდელები, რეზიუმეებისა და სტატიის სრული ტექსტებისათვის და გამოვლინდა მათგან საუკეთესო.

ზემოაღნიშნული, დიდი ზომის სამეცნიერო ნაშრომების მონაცემთა ბაზების საფუძველზე, მანქანური დასწავლით მოხდა *კლასიფიკაციის ალგორითმების* გამოყენება, ერთმანეთს შედარდა ყველაზე მნიშვნელოვანი ალგორითმები და არჩეულ იქნა მათგან

საუკეთესო შედეგის მქონე მეთოდი. მანქანის მიერ ამ ნაშრომთა კატეგორიების შესწავლით, შეიქმნა მოდელი, რომელიც ყოველი ახალი ნაშრომის დაწყებისთანავე, ავტომატურად მიაწოდებს მას შესაბამის მოდელს და შესთავაზებს მომხმარებელს შესაბამის ინტერფეისსა და ფუნქციონალს.

კვლევის მეთოდოლოგიაში აქტიურად ვიყენებდით სპეციალური გამოკითხვების ჩატარების მეთოდს – მეცნიერებთან, მთარგმნელებთან, მწერლებთან და პროექტის პოტენციურ მომხმარებელთა სხადასხვა კლასთან [16].

სპეციალური კითხვარის შემუშავება განხორციელდა ისე, რომ რესპოდენტთაგან სასურველი ინფორმაციის მაქსიმუმი ყოფილიყო მიღებული. რა თქმა უნდა, მოცემული კითხვარები განსხვავებულია მომხმარებელთა სხვადასხვა კლასისათვის.

გამოკთხვის პროცესი რომ დავყოთ ოთხ ნაწილად, ის შეიძლება შემდეგნაირად წარმოვადგინოთ:

- ინტერვიუს მომზადება;
- გამოსაკითხი ადამიანების შერჩევა, მათი მოძიება, დაკავშირება და გამოკითხვა;
- მიღებული პასუხების გაანალიზება გამოკითხვის თითოეული ჯგუფისთვის (მაგალითად, მწერლები, მეცნიერები და სხვ.);
- საერთო, ჯამური შედეგების გაანალიზება კვლევის ვალიდაციის, ფუნქციონალისა და პროცესების ავტომატიზაციის საჭიროებების გაანალიზებისთვის.

მეცნიერებთან და მწერლებთან ინტერვიუების შედეგად მიღებულ იქნა ინფორმაცია, თუ რა იქნებოდა მათთვის მუშაობის პროცესში ხელის შეშელობი და რა პრობლემებს აწყდებიან ისინი დღევანდელი ტექსტური რედაქტორების გამოყენებისას.

შესწავლილ იქნა მათი წერის მეთოდოლოგიების და მიდგომის დეტალები. მთარგმნელების შემთხვევაში, გარკვეულ

იქნა მათი მუშაობის პროცესის თავისებურებები, იმისათვის, რომ შემუშავებული ყოფილიყო მათთვის კომფორტული სამუშაო გარემო.

გამოკითხვების შედეგად მიღებულ იქნა ინფორმაცია ორი სახის განვითარების საჭიროებაზე. *პირველი* ეხება *ფუნქციონალს* – ანალიზის საფუძველზე განისაზღვრა თუ რა ფუნქციების შეთავაზება უნდა შეძლოს მან კონკრეტული ტიპის ნაშრომის ავტორს მუშაობის პროცესში. ხოლო *მეორე, ბიზნეს პროცესების ავტომატიზაციას* – ჩამოყალიბებულ იქნა თანაავტორებთან მუშაობის გზების ავტომატიზაციის მოდელები.

გასათვალისწინებელია ის ფაქტიც, რომ გამოყენებული იქნა აგრეთვე *კონტექსტუალური გამოკითხვის* მეთოდები. ეს გულისხმობს გამოკითხვის დროს პრიორიტეტული დაკვირვებების ისეთ პოტენციურ მომხმარებლებზე ყურადღების გამახვილებას, რომელთაც მუშაობის პროცესში ექმნებოდათ რაიმე პრობლემა. იქნება ეს მეცნიერი, მთარგმნელი თუ სხვა ტექსტურ ინფორმაციასთან მომუშავე მწერალი. კონტექსტუალური გამოკითხვის შედეგად მიღებული ინფორმაცია მეტად საიმედოა, მითუმეტეს ისეთ შემთხვევებში, როდესაც ერთიდაიმავე პრობლემის შესახებ არაერთი რესპოდენტი ამახვილებს ყურადღებას. ხშირ შემთხვევაში, ეს გვაძლევს პრობლემების გამოსწორების პერსპექტივას და საბოლოოდ, დიდ როლს თამაშობს მათ გამოსწორებაში.

გამოკითხვის დროს გარკვეულ იქნა არაერთი ინფორმაცია საკმაოდ მნიშვნელოვანი ასპექტების შესახებ, მათ შორისაა ის, თუ რატომ სარგებლობენ ადამიანები დღეისათვის მოცემულ სფეროში არსებული აპლიკაციებით, რა მიაჩნიათ მათ ძლიერ და სუსტ მხარეებად, რას გააუმჯობესებდნენ მასში და ა.შ. კარგი, საიმედო შედეგები მხოლოდ კარგად მომზადებული და ჩატარებული კვლევების საფუძველზე მიიღება. შესაბამისად, კითხვები იყო

მაქსიმალურად მარტივი, გასაგები და ლოგიკური. მათზე მიღებული ნებისმიერი პასუხი, გვაძლევდა სასარგებლო ინფორმაციას.

ამასთანავე, განხილულ იქნა სხვადასხვა კატეგორიის სამეცნიერო ნაშრომები და მოხდა იმ ფორმების, ფუნქციების თუ სხვა დეტალების გაანალიზება, რაც მხოლოდ ტექსტს არ წარმოადგენს და კონკრეტულ სამეცნიერო კატეგორიაში ყველაზე მეტადაა გამოყენებული. ეს ინფორმაცია სასარგებლოა იმის ანალიზისთვის თუ როგორ უნდა განლაგდეს მომხმარებლის ინტერფეისი სხვადასხვა ტიპის სამეცნიერო ნაშრომებთან მუშაობის დროს და ზოგადად, რა ფუნქციონალი უნდა იყოს პრიორიტეტული ამ ნაშრომზე მუშაობის დროს.

მეორე ეტაპზე მიღებული ინფორმაცია ამავე ეტაპზე, დოკუმენტურ მასალად იქნა ქცეული. გაანალიზებულ იქნა მიღებული შედეგები, რამაც ერთის მხრივ შექმნა ნაშრომისთვის საჭირო ბიზნეს მოთხოვნის დოკუმენტი. მიღებული ინფორმაციით მოხდა წერის სტილისა და მეთოდოლოგიების კლასიფიცირება და შესაბამისად, საჭირო ფუნქციონალთა შექმნის გზების პოვნა.

აღსანიშნავია, რომ მიღებული ფუნქციონალი უნდა იყოს მოქნილი, იმისათვის, რომ ნებისმიერ დროს თავად მწერალს შეეძლოს, მისთვის კომფორტული სტილის მიხედვით, ნაშრომის შექმნის პროცესში მაქსიმალურად მარტივად გამოიყენოს ახალი მეთოდი.

ამის შემდეგ მოხდა საპროექტო ნაშრომის მესამე ეტაპზე გადასვლა, რომელიც კონკრეტული ფუნქციონალების შექმნისთვის საჭირო პროგრამულ არქიტექტურას გულისხმობს. მოცემულ ეტაპზე შექმნილ იქნა მოდელი, პროდუქტის არქიტექტურული დეტალების შესახებ. ჩამოყალიბდა მოდელები, კლასები, პროგრამული კოდისა და დიზაინის ინტერფეისები და სხვ., გადაწყვეტილ

იქნა სპეციალური ფუნქციონალის შესრულების გზები და აეწყო სხვადასხვა ტიპის დიაგრამები.

სამივე ეტაპზე ჩატარებული ანალიზის დროს შეიქმნა როგორც ტექსტური დოკუმენტაცია, ასევე გამოყენებულ იქნა UML (უნიფიცირებული მოდელირების ენა) და BPMN (ბიზნეს პროცესების მოდელირება და ნოტაცია) დიაგრამებიც (იხ. § 1.8) [15].

კვლევის უმნიშვნელოვანეს, მეორე ეტაპის გავლის შედეგად, ხელთ გვაქვს ისეთი სამეცნიერო ღირებულების ცოდნა, როგორიცაა სამეცნიერო ნაშრომთა კლასტერიზაციის (კატეგორიზაციის) მოდელი, მათი კლასიფიკაციის მოდელი, კლასიფიცირებული წერის მეთოდოლოგიები მხატვრულ და სამეცნიერო ლიტერატურის შექმნის პროცესიდან.

აგრეთვე გვაქვს ინფორმაცია იმის შესახებ, თუ რომელი პროფესიის მქონე ადამიანი რა მხრივ საჭიროებს წერის პროცესში ხელშეწყობას და რა ფუნქციონალის დამატება დასჭირდება მას.

ეს ინფორმაცია კი დამხმარეა იმაში, რომ შეიქმნას სისტემა, რომელიც კონკრეტული სფეროს ავტორისათვის, მუშაობის პროცესს გახდის უფრო მარტივს და კომფორტულს.

გვაქვს ინფორმაცია მწერლების, მთარგმნელების, მკითხველების და სფეროსთან დაკავშირებული სხვა როლის მქონე პირების, პრობლემებისა და მათი აღმოფხვრის შესაძლებლობების შესახებ.

➤ *ინტერდისციპლინარობა და არეალი*. რაც შეეხება კვლევის ამ ორ ცნებას, მათი მთავარი სამეცნიერო მიმართულებები, რომელთაც ჩვენი პროექტი იყენებს, მანქანური დასწავლის მეთოდები, მართვის პროცესების ავტომატიზაცია და გამოყენებითი პროგრამული ინჟინერიაა. სწორედ მათ საფუძველზე მოხდა საპროექტო ნაშრომის გადაწყვეტის კონცეფციის ჩამოყალიბება.

პროექტის ერთ-ერთი მთავარი და მნიშვნელოვანი დარგია საინფორმაციო მეცნიერება. ეს მეცნიერება შეიძლება ეწოდოს

ნებისმიერ სწავლებას, დიდაქტიკას, რომელიც იკვლევს ამა თუ იმ ფორმის და შინაარსის ინფორმაციას, მისი მოპოვების, შენახვის, გადამუშავების, გარდაქმნის და გადაცემის საკითხებს. ამასთან აქ ძალზე საინტერესოა ინტერდისციპლინური სწავლება [5, 32].

რაც შეეხება საპროექტო ნაშრომიდან მიღებული სარგებლის (სისტემის) პოტენციურ მომხმარებლებს, ისინი შეიძლება იყვნენ ნებისმიერ დარგში მომუშავე ადამიანები, რომელთაც ტექსტური ინფორმაციის შექმნასთან აქვთ საქმე. პროექტის ფარგლებში, გამოკვლეული იქნა განსხვავებულ სამეცნიერო დარგებში მომუშავე ადამიანთა მუშაობის პროცესები. ამ ინფორმაციის საფუძველზე მოხდა ამ დარგთა კლასიფიკაცია და სხვადასხვა კლასებისთვის შემუშავდა მუშაობის განსხვავებული, მათთვის კომფორტული გარემო. შესაბამისად, სწორედ ეს ადამიანები მიიღებენ სარგებელს პროექტის მიერ მიღებული ცოდნებიდან გამომდინარე.

პროექტში, ყველაზე დიდი მნიშვნელობა მისი სამეცნიერო სიახლისა და პრაქტიკული ღირებულებებიდან გამომდინარე, ენიჭება სამეცნიერო სფეროში, ტექსტურ ინფორმაციასთან მომუშავე პირებს. პროექტის მიხედვით, მათ უნდა დაზოგონ დრო და ფინანსური რესური და შესაბამისად შესაძლებელი გახდება ამ რესურსების მიმართვა კვლევის პროცესზე, რაც თავის მხრივ ხელს შეუწყობს პროდუქტის ხარისხის ზრდას, რაც ძალზედ მნიშვნელოვანი ფაქტორია.

ამასთან ერთად, პროექტი ხელს უწყობს მთარგმნელებსაც. ეს გულისხმობს ერთგვარად, საინტერესო როლის მქონე მომხმარებელთა კატეგორიის დამატებას. თარგმანი ესაჭიროება, როგორც მხატვრულ, ასევე სამეცნიერო ლიტერატურას. იმისთვის, რომ ხარისხიანი და ღირებული ნამუშევარი არა მარტო რომელიმე ერთი ენის მქონე ადამიანებისთვის იყოს ხელმისაწვდომი, არამედ მთელი მსოფლიოსთვის, საჭიროა მისი თარგმანი, რასაც ძალიან დიდი

ძალისხმევა და შრომა სჭირდება. მთარგმნელის როლი მკითხველისთვის ისეთივე მნიშვნელოვანია როგორც მწერლისთვის.

ჩვენი პროექტი, თავის მხრივ, მოიცავდა საგამომცემლო მარკეტინგის სფეროს საინფორმაციო სისტემის არქიტექტურის განსაზღვრის ამოცანას. ნებისმიერ მომხმარებელს შეეძლება სისტემაში ნაშრომის დაწერა და გამოქვეყნება. ეს ფუნქცია კი ერთი მხრივ საშუალებას აძლევს ნაკლებად ცნობილ და დამწყებ მეცნიერებსა და მწერლებს წარმატების მიღწევაში, ხოლო მეორეს მხრივ საშუალებას აძლევს რომელიმე სფეროთი დაინტერესებულ პირებს შეძლონ სასურველი ინფორმაციის მიღება.

მოკლედ რომ ვთქვათ, ჩვენი ნასრომის მიზანია ჩამოყალიბდეს სისტემა, რომელიც გააერთიანებს სხვადასხვა სამეცნიერო დარგში მოღვაწე ადამიანებს: მწერლებს, მეცნიერებს, მთარგმნელებს, და, რა თქმა უნდა, მკითხველებს, ეს ყველაფერი კი ხელმისაწვდომი უნდა იყოს ნებისმიერი მომხმარებლისათვის.

1.8. საგამომცემლო ბიზნესპროცესების მოდელირება

BPMN სტანდარტით

საწარმოო რესურსების დაგეგმვის (ERP) სისტემის დანერგვისას საგამომცემლო კომპანიაში (იხ. § 1.4) წინასწარ ხდება ამ პროცესის მოდელირება და შემდეგ მისი ეფექტიანობის ეკონომიკური ანალიზი. მოდელირების მიზნით ხშირად გამოიყენება BPMN ინსტრუმენტი [15].

1.4 ნახაზზე, მაგალითად, ნაჩვენებია საგამომცემლო კომპანიის ვაკანსიებზე თანამშრომელთა მიღების ბიზნესპროცესის სქემა BPMN-ის გამოყენებით. ჰორიზონტალურ დონეებზე (ბილიკებზე) განთავსებულია აპლიკანტის და კომპანიის ფუნქციური დეპარტამენტების ბიზნეს-ოპერაციები (ქმედებები, რაც მათ ევალებათ) და ბიზნეს-წესები (კანონით და დებულებით განსაზღვრული წესები),

დიაგრამის აგების მიზნით გამოიყენება სხვადასხვა ინსტრუმენტული საშუალებები. ჩვენ შემთხვევაში ესაა Bizagi Process Modeler – ბიზნესპროცესების მოდელირების და დოკუმენტირების გრაფო-ანალიზური პროგრამა, იგი საშუალებას აძლევს მომხმარებელს ვიზუალურად შექმნას პროცესების დიაგრამა, მოდელები და მოახდინოს მისი დოკუმენტირება მსოფლიოში აღიარებული სტანდარტის BPMN-ის გამოყენებით [32].

1.5 ნახაზზე მოცემულია სქემაზე გამოყენებული გრაფიკული აღნიშვნების, კერძოდ, ლოგიკური ოპერატორების ფუნქციები.

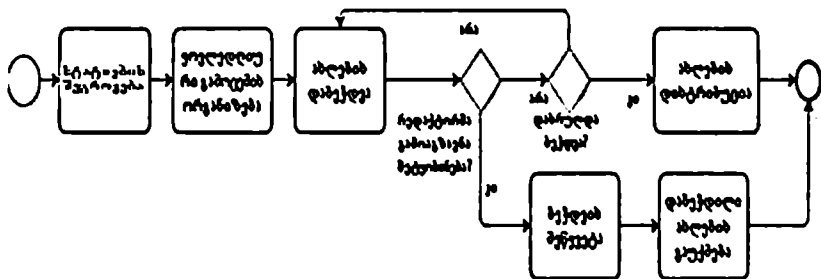
ლოგიკური ოპერატორები	
'ან' ოპერატორის გამოიყენება მონაცემთა მართვისას (Data XOR)	 ან 
მოვლენითი 'ან' ოპერატორი: ქმნის პროცესის ახალ ეგზემპლარს (Event XOR)	
'და' ოპერატორი (AND)	
'ან' ოპერატორი (OR): განმტოებისას აქტიურდება ერთი ან ყველა მტო. შერწყმისას ყველა მოქმედი შემავალი მტო იხურება	
რთული ოპერატორი. ამოღელირებს განმტოების და შერწყმის რთულ პირობებს	
მოვლენითი 'და' ოპერატორი (ქმნის პროცესის ახალ ეგზემპლარს)	

ნახ.1.5. Bizagi Process Modeler-ის ლოგიკური ოპერატორები

BPM ინსტრუმენტი და მისი სრული აღწერა იხ. [15, 33, 34].

განვიხილოთ რამდენიმე საილუსტრაციო მაგალითი საგამომცემლო მარკეტინგის სხვადასხვა ეტაპის პროცესების მოდელირებისათვის Bizagi ინსტრუმენტით [15].

მაგალითი 1: რედაქტორის მიერ გადაწყვეტილების მიღება წიგნის გამოცემის (ბეჭდვის) ბიზნესპროცესის გაუქმებაზე. 1.6 ნახაზზე მოცემულია ამ ამოცანის დიაგრამა, რომელშიც გადაწყვეტილების მიღებისთვის გამოიყენება *რომბის* ფიგურის აღნიშვნა და ციკლის ფუნქცია.



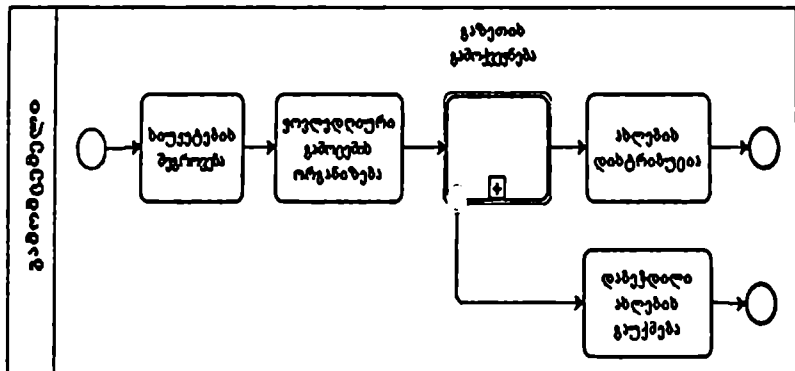
ნახ.1.6. წიგნის ბეჭდვის ან გაუქმების პროცესის BPMN მოდელი

მაგალითად, აქ გამომცემელი აგროვებს სტატიებს, ორგანიზებას უწევს დღიურ გამოცემას და შემდეგ იწყებს ასლების ბეჭდვის პროცესს. თუ რედაქტორი დარეკავს და მისცემს მითითებას, გამომცემელი შეწყვეტს პროცესს. თუ არ დარეკავს, მაშინ გამომცემელი აგრძელებს ბეჭდვის პროცესს დასრულებამდე, შემდეგ კი ავრცელებს სტატიებს.

მაგალითი 2: აქაც პროცესის შინაარსი იგივეა. გაფართოებული ქვეპროცესი არის მოცემული ორმაგი ხაზით, რაც მიუთითებს, რომ ეს არის ტრანზაქციული ქვეპროცესი. იგი შეიძლება შეწყდეს. მაგალითისთვის, თუ ასლების დაბეჭდვის

პროცესში შეტყობინება მიღებული იქნება რედაქტორისგან, მაშინ ქვეპროცესი შეწყდება.

BPMN-ით ასეთი პროცესი იმართება შედარებით უფრო მარტივად (ნახ.1.7).



ნახ.1.7. ბეჭდვის პროცესის გაუქმების შუალედური მოვლენა

ასევე უნდა აღვნიშნოთ, რომ BPMN მეთოდი საშუალებას იძლევა მივიღოთ ცხადი დიაგრამა. აქ არ გამოიყენება რომბის (გადაწყვეტილების მიღების) ფიგურა და ციკლი. ნათელია, რომ „დაბეჭდვის“ ქვეპროცესი არის ამოცანების მკაფიო კონფლექტი და ეს სიცხადე შეიძლება იყოს ძალზე მოსახერხებელი ბიზნეს-პროცესის ოპტიმიზაციისა და მართვისათვის.

თავი 2

კვლევის მეთოდები და ინსტრუმენტები

2.1. ხელოვნური ინტელექტი და მანქანური დასწავლა	55
2.2. ბუნებრივი ენის დამუშავება (NLP)	63
2.3. თემის მოდელირება	69
2.4. დირიხლეს ლატენტური განაწილება	74
2.5. არაუარყოფითი მატრიცის ფაქტორიზაცია	77
2.6. კლასიფიკაცია	82
2.7. მონაცემთა განთავსება სერვერზე	85
2.7.1. მონაცემთა ბაზები	86
2.7.2. ინსტრუმენტები რეალური დროის ფუნქციების შესასრულებლად	86
2.8. ბიბლიოთეკები და პროგრამული ენები	88
2.9. პროგრესული ვებ აპლიკაცია PWA	90

ჩვენი წიგნის ინოვაციურობა, ზოგადად, მისი ძირითადი მიზნის, ინფორმაციული საზოგადოების ფორმირების ხელშეწყობაა საგამომცემლო ინდუსტრიის და მარკეტინგის ბიზნესპროცესების სრულყოფის საფუძველზე. კონკრეტულად კი ისეთი ფრეიმვორკის შექმნაა, რომელიც საზოგადოების სხვადასხვა სფეროს წარმომადგენლებს, როგორიცაა მწერლები, მეცნიერები, მთარგმნელები და ა.შ., მიაწვდის მათი მიზნობრივი პროდუქტის (ლიტერატურული წიგნი, მონოგრაფია, სამეცნიერო სტატია, თარგმანი და სხვ.) შესაქმნელად მოხერხებულ და მარტივ ინსტრუმენტულ საშუალებას (ტექსტების, გრაფიკების, ფორმულების და სხვ. ტიპის ნაშრომის შესაქმნელ პროგრამულ რედაქტორს).

ავტორის მიერ მითითებული კონკრეტული თემატიკის და სათაურის ანალიზის საფუძველზე პროგრამულ პროდუქტს უნდა შეეძლოს ინტერნეტში მსგავსი (არსებული) ნაშრომების მოძიება და გარკვეული რეკომენდაციების ჩამოყალიბება. აგრეთვე სავარაუდო ნაშრომის კვლევის სფეროს დადგენა (მაგალითად, მათემატიკა, ფიზიკა, ბიოლოგია, კომპიუტერული მეცნიერება, საკომუნიკაციო ტექნოლოგია და ა.შ.) და სფეროს შესაბამისი ინტერფეისის შეთავაზება ავტორისადმი, ყოველგვარი ზედმეტი ფუნქციონალური დილაკებისა და პიქტოგრამებისგან.

ასეთი გლობალური, ინოვაციური ამოცანის გადაწყვეტა დაკავშირებულია პრობლემის დეტალურ და ზუსტ იდენტიფიკაციასთან, მათემატიკის ალბათობის თეორიისა და სტატისტიკის მეთოდების, ხელოვნური ინტელექტის, მანქანური დასწავლის ალგორითმების, ახალი ციფრული ტექნოლოგიების და პროგრამირების ჰიბრიდული სისტემების გამოყენებასთან.

2.1. ხელოვნური ინტელექტი და მანქანური დასწავლა

ხელოვნური ინტელექტი (*Artificial Intelligence - AI*) მიეკუთვნება კომპიუტერულ მეცნიერებათა სფეროს, რომელიც ქმნის, ამუშავებს და შეისწავლის ინტელექტუალურ მანქანებს.

ადამიანების (ბუნებრივი) ინტელექტისგან განსხვავებით იგი მანქანური (ხელოვნური) ინტელექტია, რომელიც შეიძლება იყოს პროგრამულ უზრუნველყოფაში გადატანილი. კომპიუტერები და რობოტები ასეთი პროგრამების საფუძველზე იმიტირებენ (იქცევიან) ადამიანებივით, ხშირად უკეთესადაც (მაგრამ, არა ყველაფერში). ხელოვნური ინტელექტის პროგრამებს შეუძლია მონაცემთა ანალიზის და კონტექსტუალიზაციის განხორციელება ინფორმაციის მიწოდების (გადაცემის) მიზნით ან ავტომატურად აწარმოონ მოქმედებები ადამიანის ჩარევის გარეშე. [41.]. განსაკუთრებული როლი ხელოვნურმა ინტელექტმა გამოავლინა ექსპერტული სისტემების სახით ინდუსტრიის სფეროებში, სამედიცინო დიაგნოსტიკის, სოფლის მეურნეობის, ენერგეტიკის, საფინანსო ბანკების და სხვა მრავალ დარგში [42 - 44].

მანქანური დასწავლა (*Machine Learning - ML*) ხელოვნური ინტელექტის ქვეკატეგორიაა. ის არის მონაცემთა მეცნიერების (*Data science*) ერთ-ერთი უმნიშვნელოვანესი ნაწილი, რომელიც კომპიუტერს რეალობის პროგნოზირების საშუალებას აძლევს. მისი ძირითადი მიზანი ისაა, რომ ტექნოლოგიურ მანქანას შეეძლოს მიბადოს ადამიანის ქცევებსა და აღქმას სხვადასხვა მიმართულებით. მათ შორის არსებული რეალობის გააზრებითა და გაანალიზებით.

მანქანური დასწავლა იყენებს ალგორითმებს, რათა ავტომატურად ისწავლოს (თვითსწავლება) ინფორმაცია და ამოიცნოს

მონაცემებიდან მიღებული შაბლონები, ასეთი დასწავლის გამოყენებით უფრო უკეთესი გადაწყვეტილებების მისაღებად.

მანქანური დასწავლის დისციპლინის ათვისებით და ექსპერიმენტებით, პროგრამისტ-დეველოპერები ამოწმებენ შესაძლებლობებს და საზღვრებს, თუ რამდენად ხელეწიფებათ მიზნობრივი კომპიუტერული სისტემის სრულყოფა, მათი აღქმა, შემეცნება და მოქმედება.

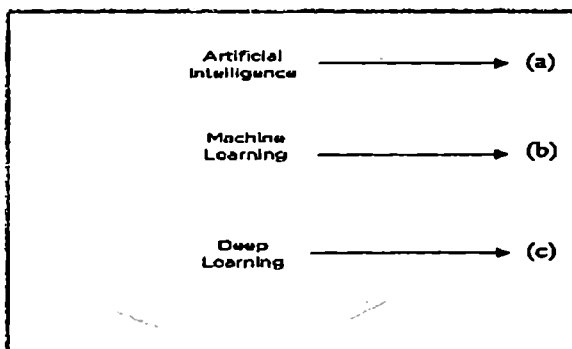
ღრმა სწავლება (Deep learning), მანქანური დასწავლის ერთ-ერთი ახალი მიმართულება, პროგრესული კვლევის მეთოდია. მისი მოდელები იყენებს დიდ *ნეირონულ ქსელებს (Neural networks)* – ქსელებს, რომლებიც ფუნქციონირებს ადამიანის ტვინის მსგავსად მონაცემთა ლოგიკური ანალიზის მიზნით, რთულ კანონზომიერებათა შესასწავლად და პროგნოზების გასაკეთებლად ადამიანის ჩარევისაგან დამოუკიდებელად.

მაშასადამე, ხელოვნური ინტელექტის სისტემების აგება, კვლევა და განვითარება მისი ქვეკატეგორიების საფუძველზე ხდება. ესენია:

- მანქანური დასწავლა;
- ღრმა სწავლება;
- ნეირონული ქსელები;
- კომპიუტერული ხედვა;
- ბუნებრივი ენის დამუშავება.

შემდგომში განვიხილავთ ამ საკითხებს უფრო დეტალურად, კერძოდ, თუ რა განსხვავებაა ხელოვნურ ინტელექტსა და მანქანურ დასწავლას შორის, როგორ გამოიყენება ისინი დღეს დიდ და პატარა ორგანიზაციებში.

a) ხელოვნური ინტელექტი (AI) არის ჭკვიანი ინტელექტუალური მანქანების შექმნის კონცეფცია (ნახ.2.1) [42].



ნახ.2.1. AI, ML და DL -ის დამოკიდებულება

b) მანქანური დასწავლა (ML) – ხელოვნური ინტელექტის ქვე-სიმრავლე, რომელიც ქმნის ხელოვნურ ინტელექტზე ორიენტირებულ აპლიკაციებს;

c) ღრმა სწავლება (DL) – მანქანური დასწავლის ქვესიმრავლეა, რომელიც იყენებს ნეირონული ქსელების დიდ მონაცემებს და რთულ ალგორითმებს მოდელის მოსაშზადებლად (მისი განსწავლის მიზნით) [45,46].

➤ მანქანური დასწავლა ეხმარება სხვადასხვა პროგრამებს ან პროგრამისტ-დეველოპერებს (უშუალო დაპროგრამების გარეშე) მოახდინონ ამა თუ იმ რეალობის პროგნოზირება. ამას კი ისინი ახორციელებს სხვადასხვა ალგორითმების და მათი გამოყენების შედეგად მიღებული მოდელების მიერ შესწავლილი ისტორიული მონაცემების საფუძველზე. მანქანურ დასწავლას იყენებენ ისეთ განსხვავებულ მიმართულებათა ამოცანებში, როგორებიცაა:

- 1) ტექსტის კლასიფიკაცია;
- 2) თემის მოდელირება;
- 3) თაღლითობის გამოვლენა;
- 4) მომხმარებელთა ინტერესების შესწავლა;

5) სპამის ფილტრაცია;

6) ბიზნეს პროცესების ავტომატიზაცია და სხვ.

მანქანური დასწავლა ძალზე მნიშვნელოვანია იმ მხრივაც, რომ იგი იძლევა ისეთ ინფორმაციას, როგორიცაა მომხმარებელთა ქცევის ტენდენციების ხედვა ბიზნესოპერაციების სტანდარტებზე. ამ მხრივ, იგი უზრუნველყოფს ისეთი გლობალური პროდუქტების განვითარებას, რომლებიც დღეს ტექნოლოგიური გიგანტების: Google, Amazon, Facebook, Netflix და სხვ., სიაშია.

მსოფლიოს უდიდესი ტექნოლოგიური კომპანიების ერთ-ერთი უმთავრესი და უმნიშვნელოვანესი კვლევებისა და მომავალი განვითარების ნაწილი სწორედ მანქანური დასწავლის გარშემოა კონცენტრირებული. 1970-იანი წლების ბოლომდე იგი ხელოვნური ინტელექტის ნაწილად მიიჩნეოდა, ხოლო შემდეგ მოხდა მისი განშტოება და საკუთარი განვითარების გზის ჩამოყალიბება.

მანქანური დასწავლა, გარკვეული თვალსაზრისით, ტვინის უჯრედების ურთიერთქმედების მოდელზეა დაფუძნებული. ასეთი მოდელი 1949 წელს შექმნა კანადელმა მეცნიერმა, ფსიქოლოგმა დონალდ ჰებმა (Donald Hebb), რომელმაც გამოსცა წიგნი „ქცევათა ორგანიზაცია“ [47]. აღნიშნულ წიგნში ავტორმა აღწერა თეორიები, თავის ტვინში არსებული ნეირონების ურთიერთქმედებისა და მათი აგზნების შესახებ.

რაც შეეხება თავად მანქანური დასწავლის სახელწოდებისა და მისი რეალური შექმნის ისტორიას, იგი კომპანია IBM-ის თანამშრომელ არტურ სემუელს (Arthur Samuel) ეკუთვნის. მან 1950-იან წლებში, შექმნა კომპიუტერული პროგრამა, ყველასთვის კარგად ნაცნობი, ერთ-ერთი უძველესი თამაშისთვის „შაში“. მაშინდელ კომპიუტერებში, მეხსიერების სიმცირის გამო, მას მოუწია ისეთი მეთოდებისა და ფუნქციების გამოგონება, რომელთა

გამოყენებითაც, შესაძლებელი ხდებოდა ნაკლები მეხსიერებით, მეტი გამოთვლების ჩატარება.

მისი დიზაინის მიხედვით, სათამაშო დაფაზე განლაგებული ქვების მიხედვით, ყოველ პერსპექტიულ ქვას თავისი ქულა ენიჭებოდა. მოცემული ინფორმაცია აძლევდა მას საშუალებას გამოეთვალა თითოეული მხარის გამარჯვების შანსები. მინიმაქსის სტრატეგიის გამოყენებით მისი პროგრამა ირჩევდა თავის შემდეგ ნაბიჯს, რაც სამომავლოდ, მანქანური დასწავლის არსებულ მინიმაქსის წესში (ალგორითმში) გადაიზარდა [48].

ა. სემუელმა შეიმუშავა არაერთი მნიშვნელოვანი ალგორითმი, რომელიც მისი შექმნილი „შაშის“ თამაშს უკეთესს ხდიდა. მისი პროგრამა თამაშის ყოველი შემდეგი პარტიისთვის უკეთესი ხდებოდა. 1959 წელს სწორედ მან პირველმა შემოიღო ტერმინი „მანქანური დასწავლა“ (თვითსწავლება), რასაც 80 წლის შემდეგ უდიდესი ადგილი უკავია ტექნოლოგიურ სამყაროში.

ალეწეროთ მანქანური დასწავლის მუშაობის პროცესი.

თვითსწავლების პროცესისთვის აუცილებელია გვექონდეს მონაცემები. ეს მონაცემები შეიძლება იყოს, ტექსტური, რიცხვითი, მედია თუ სხვ. მონაცემების გამოსაყენებლად, პირველი და ურთულესი საჭიროება, რა თქმა უნდა, მისი, მოძიება ან შეგროვებაა. ინფორმაციის მოძიების შემდეგ, ის უნდა იქნას გამოყენებული, როგორც ტრენინგის მონაცემები ან მონაცემთა ნაკრები, რომელსაც მანქანური დასწავლის მოდელი გამოიყენებს თვითსწავლებისთვის.

მას შემდეგ, რაც მანქანური დასწავლის პროგრამის შემქმნელი შეარჩევს დასწავლის მეთოდს, ის ქმნის თვითსწავლის მოდელს, რომელსაც მიაწვდის არსებულ ინფორმაციას, მოდელი კი საკუთარ თავს შეასწავლის ამ ინფორმაციის თავისებურებებს და ამ ცოდნას მომავალი პროგნოზებისა და ნიმუშების მოძებნისათვის გამოიყენებს. მომავალში კი შესაძლებელი იქნება ამ მოდელისა და მისი

სხვადასხვა პარამეტრების ცვლილება იმისთვის, რომ შესაძლებელი გახდეს უკეთესი შედეგების მიღება. აგრეთვე შესაძლებელია მოდელისათვის ინფორმაციის პერიოდული დამატება, რაც მას უფრო სრულყოფილს გახდის.

მანქანური დასწავლის დაწყებამდე, ხდება არსებული მონაცემების გაყოფა ორ ნაწილად, ესენია: *სასწავლი* და *დასატესტი მონაცემები*. როგორც წესი, სატესტო მონაცემთა ჯგუფის რაოდენობა ნაკლებია სასწავლო მონაცემთა რაოდენობაზე და დაახლოებით 20-30 პროცენტამდე მერყეობს. სატესტო მონაცემები გამოიყენება მიღებული მოდელის შესაფასებლად და როდესაც მას აჩვენებენ ახალ მონაცემებს, ის ამოწმებს თუ რამდენად ზუსტია მანქანური დასწავლის ეს მოდელი. შედეგად მივიღებთ მოდელის სისწორის პროცენტულ მაჩვენებელს, რომლის დამაკმაყოფილებლობაც სხვადასხვა შემთხვევაში შეიძლება განსხვავებული იყოს (მაგალითად, კლინიკურ შემთხვევაში მისი სიზუსტის მაღალი მაჩვენებელი უფრო მნიშვნელოვანია, ვიდრე არსებულ პროდუქტებში მომხმარებელთა ინტერესების შესაბამისი შეთავაზებების გაკეთების დროს).

მანქანური დასწავლისგან მიღებული მოდელები, სამი ძირითადი მოთხოვნის შესასრულებლად შეიძლება იქნას გამოყენებული. ეს მოთხოვნებია:

- *აღწერითი* – მანქანური დასწავლის მოცემული ფუნქცია გულისხმობს იმას, რომ იგი არსებულ ცოდნას, რაიმე მომხდარის მიმართ ახსნისათვის იყენებს;

- *პროგნოზირებადი* – ნიშნავს, რომ მოდელი მონაცემებს იმის პროგნოზისათვის გამოიყენებს, თუ რა შეიძლება მოხდეს, ან რა კლასს ან კატეგორიას შეიძლება მიეკუთვნებოდეს მისთვის მიწოდებული მონაცემი;

- *ინსტრუქციული* – მანქანური დასწავლის მოდელი გასცემს ინსტრუქციებს იმის შესახებ, თუ რა ქმედებები უნდა განხორციელდეს ამა თუ იმ შემთხვევაში.

მანქანური დასწავლის თეორიაში სამი ძირითადი მეთოდი არსებობს: *კონტროლირებადი (მართვადი) თვითსწავლება*, *არაკონტროლირებადი (არამართვადი) თვითსწავლება* და *ნახევრად-კონტროლირებადი თვითსწავლება* [93-95]:

- *კონტროლირებადი მანქანური დასწავლის* მეთოდი განისაზღვრება ისეთი მონაცემთა ნაკრებით, რომელიც რაიმე გზით არის კლასიფიცირებული და აქვს ინფორმაცია ამ კლასების შესახებ. ამ ინფორმაციის გამოყენებას ვახდენთ სხვადასხვა პროგრამული მიდგომის მეშვეობით ისე, რომ საბოლოოდ შეგვიძლია ახალი შეყვანილი მონაცემების კლასიფიკაცია ან პროგნოზი. როგორც კი არსებულ მოდელს ახალი მონაცემები დავმატებთ, ის არეგულირებს მათ იქამდე, სანამ არ მოხდება მისი სათანადო მორგება უკვე არსებულ მონაცემებზე.

კონტროლირებადი მანქანური დასწავლა ეხმარება ორგანიზაციებს გადაჭრას ისეთი საკითხები, როგორიცაა ტექსტის კლასიფიკაცია, თაღლითური ან სპამ-წერილების ფილტრაცია და ა.შ. მოცემულ მეთოდებში გამოყენებულია წრფივი რეგრესია, ლოჯისტიკური რეგრესია და ა. შ.;

- *არაკონტროლირებადი მანქანური დასწავლა* იყენებს სხვადასხვა მათემატიკურ ფუნქციებსა და პროგრამულ ალგორითმებს იმისთვის, რომ მოახდინოს ისეთი მონაცემთა კატეგორიზაცია, რომელსაც ჯერ არ აქვს რაიმე კატეგორია. მაგალითად, ჩვენი საპროექტო კვლევის შემთხვევაში, ალგორითმისათვის გადაცემულ იქნა მილიონობით სამეცნიერო სტატიის რეზიუმეს ტექსტი და მისი საშუალებით მოხდა ამ ნაშრომების ახალ კატეგორიებში განთავსება (ამ საკითხს მოგვიანებით დავუბრუნდებით).

მოკლედ რომ ვთქვათ, ზემოაღნიშნული ალგორითმები, ახდენს მონაცემთა ავტომატურ დაჯგუფებას ადამიანის ჩართულობის გარეშე. ინფორმაციაში რაიმე მსგავსებებისა თუ განსხვავებების ისეთი მოთხოვნების დასაკმაყოფილებლად, როგორცაა მომხმარებელთა კლასიფიკაცია, ანალიზი სხვადასხვა ტიპის გაყიდვებისათვის, სახეთა ამოცნობა და ა.შ.

ჩვენი კვლევის შემთხვევაში არაკონტროლირებადი მანქანური დასწავლა გამოიყენება ტექსტის თემატური კლასიფიკაციისათვის, მომხმარებელთა ინტერესების შესასწავლად სხვადასხვა თემატიკის და არა მხოლოდ წინასწარი გაწერილი კატეგორიის მიხედვით, მრავალი კატეგორიის გაერთიანებისათვის და სხვ.

– ნახევრად-კონტროლირებადი მანქანური დასწავლის მეთოდი მიდგომა კონტროლირებად და არაკონტროლირებად მანქანურ დასწავლებს შორის, რომელიც, თავის მხრივ, მოიცავს ორივე მათგანს. თვითსწავლების დროს ის იყენებს რაოდენობრივად, შედარებით მცირე კონტროლირებად (კატეგორიზებულ) მონაცემებს იმისათვის, რომ მიღებულ იქნას გარკვეული ცოდნა არსებულ მონაცემთა კლასებთან დაკავშირებით, რომელიც გამოყენებულ იქნება უფრო მოზრდილი არაკონტროლირებადი მონაცემების მახასიათებლების მოსაპოვებლად.

ნახევრადკონტროლირებად მანქანურ დასწავლას შეუძლია გადაჭრას არასაკმარისი კატეგორიზებული მონაცემების ქონა მომავალში, ამ მონაცემთა სხვადასხვა მანქანური დასწავლის მეთოდების გამოსაყენებლად.

საპროექტო კვლევაში გამოყენებული იქნა მანქანური დასწავლის ისეთი მეთოდები, როგორებიცაა ტექსტის კლასიფიკაცია და თემის მოდელირება. რაც იმას ნიშნავს, რომ გამოყენებულია მისი ორი უმნიშვნელოვანესი ქვეკატეგორია, კონტროლირებადი და არაკონტროლირებადი კვლევები. საბოლოოდ, კი ამ

მიმართულებების და მეთოდების გაერთიანებით მიღებულია, ისეთი მოდელი, სამეცნიერო სიახლე და პრაქტიკული ღირებულებები, რომელთაც შეუძლია ტექსტური ინფორმაციის შექმნის პროცესის და საგამომცემლო მარკეტინგის ხელშეწყობა.

უნდა აღინიშნოს ის ფაქტი, რომ მანქანური დასწავლის მოცემული მეთოდების დამუშავება მოხდა ისეთი მიმართულების გამოყენებით, როგორიცაა ბუნებრივი ენის დამუშავება, რაც ადამიანური ენის კომპიუტერულ ენად ტრანსფორმაციას გულისხმობს.

2.2. ბუნებრივი ენის დამუშავება (NLP)

სანამ თემის მოდელირების და კლასიფიკაციის ალგორითმებს შევიმუშავებთ, აუცილებელია არსებული სამეცნიერო ნაშრომებისა და რეზიუმეების (ტექსტების) გასუფთავება ზედმეტი ფრაზებისგან, რომლებმაც შეიძლება ხელი შეუშალოს ორივე ტიპის მანქანურ დასწავლას.

ასეთი ტექსტები შეიძლება იყოს არტიკლები, კავშირები, რიცხვები და სხვა ისეთი სიტყვები ან ფრაზები, რომლებიც კონკრეტული კვლევისთვისაა ხელისშემშლელი ფაქტორი.

ამ პრობლემების აღმოსაფხვრელად, მანქანურ დასწავლაში არსებობს მიდგომა - *ბუნებრივი ენის დამუშავება (Natural language processing – NLP)*. იგი მოიცავს *გამოთვლით ლინგვისტიკას, მანქანურ დასწავლას და ღრმა სწავლების მოდელებს* ადამიანური ენის დასამუშავებლად.

გამოთვლითი ლინგვისტიკა მეცნიერებაა ადამიანის ენის მოდელების გაგებისა და აგების შესახებ კომპიუტერების და პროგრამული ინსტრუმენტების დახმარებით.

როგორც უკვე აღვნიშნეთ, ბუნებრივი ენის დამუშავება, არის იმის შესაძლებლობა, რომ კომპიუტერმა შეძლოს ადამიანის ბუნებრივი ენის აღქმა, ისე, რომ მოხდეს გაანალიზება იმისა, თუ რა

არის ნაგულისხმები, მის მიერ ნათქვამ თუ დაწერილ ინფორმაციაში. ბუნებრივი ენის დამუშავება, მანქანური დასწავლისა და ზოგადად ხელოვნური ინტელექტის ერთ-ერთი მნიშვნელოვანი კომპონენტია.

ბუნებრივი ენის მოდელირების მიდგომები, კომპიუტერს საშუალებას აძლევს ისევე აღიქვას ადამიანის ბუნებრივი ენა, როგორც ამას სხვა ადამიანები აკეთებენ. კომპიუტერულ სისტემაში ტექსტური ინფორმაციის შეყვანის შემდეგ, რა ტიპისაც არ უნდა იყოს იგი, ნაწერი თუ წარმოთქმული, ბუნებრივი ენის მოდელირება, სპეციალური ალგორითმების მეშვეობით, ახდენს ამ ინფორმაციის დამუშავებას და მას კომპიუტერისთვის აღსაქმელ მნიშვნელობას სძენს.

ბუნებრივი ენის მოდელირების, ხელოვნური ინტელექტის ერთ-ერთ უმნიშვნელოვანეს ნაწილად წარმოჩენისთვის, წარმოვიდგინოთ, რომ როგორც ადამიანს აქვს სხვადასხვა ორგანოები იმისთვის რომ მიიღოს ინფორმაცია (მაგალითად მხედველობა (თვალები), იმისთვის რომ წაიკითხოს და სმენა (ყურები) - რომ მოისმინოს), აგრეთვე კომპიუტერებს აქვს აპარატურა, იმისთვის რომ მოახდინონ ინფორმაციის დანაბვა და გაგება. ხოლო რაც შეეხება მის დამუშავებას, ერთის მხრივ, ადამიანის ტვინი ახდენს ამ ინფორმაციის დამუშავებას და აზრის გაანალიზებას, ხოლო, მეორეს მხრივ, კომპიუტერში სპეციალური ალგორითმების საშუალებითა და ბუნებრივი ენის მოდელირების მიდგომებით ხდება ამ ინფორმაციის პროგრამულ ნაწილად კონვერტირება და შემდეგ მისი აღქმა.

ბუნებრივი ენის დამუშავებისთვის ორი ძირითადი ეტაპია: მონაცემთა წინასწარი დამუშავება და ამ მონაცემთა შესწავლის საფუძველზე, მანქანური დასწავლის მოდელის მიღება.

მონაცემთა წინასწარი დამუშავება გულისხმობს ტექსტური ინფორმაციის გაწმენდას, მისგან არასაჭირო ინფორმაციის ამოშლას და ამით მის მომზადებას მანქანური დასწავლის ალგორითმებისთვის, რათა მათ შეძლონ ამ ინფორმაციის შემდგომი ანალიზი.

წინასწარი დამუშავება ამ ყველაფერთან ერთად, მონაცემებს უცვლის ფორმატს და აყენებს სამუშაო ფორმატში. ის აგრეთვე ხაზს უსვამს ტექსტში არსებულ მახასიათებლებს, მაგალითად, სიტყვის გრამატიკულ ფორმას და ა.შ. ხოლო ამის შემდეგ, ალგორითმებს საშუალება ეძლევათ უფრო მარტივად იმუშაონ ტექსტთან.

მონაცემთა დამუშავების რამდენიმე გზა არსებობს, მათ შორის ჩვენს მიერ გავლილი გზებია:

➤ **ტოკენიზაცია (Tokenization)**, ბუნებრივი ენის დამუშავებისა და მანქანური დასწავლის სფეროში, ეხება ტექსტის თანმიმდევრობის მცირე ნაწილებად გარდაქმნის პროცესს, რომელიც ცნობილია როგორც ნიშნები (tokens). ესაა საჭირო მონაცემთა სპეციალური უნიკალური საიდენტიფიკაციო კოდებით ჩანაცვლების პროცესი.

იგი ინახავს ყველა საჭირო ინფორმაციას შენახული მონაცემების (ტექსტის) შესახებ. მისი მიდგომიდან გამომდინარე აშკარაა, რომ ის ცდილობს მინიმუმამდე დაიყვანოს არსებული მონაცემების რაოდენობა მეხსიერების მიხედვით, რაც შემდგომში მის დამუშავებას საგრძნობლად უფრო სწრაფს გახდის.

ჩვენი კვლევის შემთხვევაში საქმე გვაქვს მილიონობით სამეცნიერო ნაშრომთან, რაც ბუნებრივია დიდი პრობლემაა მონაცემთა დამუშავების ხანგრძლივობასთან დაკავშირებით.

სწორედ ტოკენიზაციაა საშუალება, რომლის გამოყენებითაც შესაძლებელი ხდება ამ დროის მინიმუმამდე შემცირება და მისი მნიშვნელოვანი დაზოგვა.

ამოსაშლელ სიტყვათა სია (Stop word) - წინასწარ შედგენილი სიაა, იმისთვის რომ ტექსტიდან მოხდეს ისეთი სიტყვების ამოღება, რომლებიც ძალიან ხშირადაა გამოყენებული, არ გააჩნია თემატური მნიშვნელობა ან უბრალოდ ხელს უშლის მანქანური დასწავლის ალგორითმებს ჩვენთვის სასურველი ინფორმაციის დამუშავებაში.

მაგალითად, როდესაც ჩვენს ხელთ არსებული სამეცნიერო სტატიების დამუშავებისას აშკარა გახდა ფაქტი, რომ ნაშრომებში დამატებული იყო ისეთი სიმბოლოები ან სიტყვები, რაც საბოლოო ჯამში ურევდა ისევე, როგორც თემის მოდელირების ალგორითმებს, ასევე კატეგორიზაციასაც კი (საპროექტო კვლევის დეტალებს მოგვიანებით განვიხილავთ).

მას შემდეგ, რაც არასაჭირო სიტყვები ამოღებულ იქნა ნაშრომიდან, დარჩა ისეთი სიტყვები, რომელთა დამუშავებაც უფრო დეტალურ და ზუსტ ინფორმაციას გვაძლევდა ამ ტექსტის კატეგორიისა თუ მისი თემის შესახებ.

ზოგადად, *ამოსაშლელ სიტყვათა სია*, ხშირად გამოიყენება სხვადასხვა საძიებო სისტემებში. მაგალითად, გუგლის საძიებო სისტემაში ძებნისას, ის არ იყენებს არტიკლებს ძებნის პროცესში თუ მას სპეციალური სიმბოლოებით არ მოვნიშნავთ.

მანქანური დასწავლის არსებულ ბიბლიოთეკებში უკვე არსებობს *ავტომატურად ინტეგრირებული მეთოდები*, სადაც ამა თუ იმ არასაჭირო გრამატიკული სიტყვების იგნორირება ხდება. თუმცა, პირველადი დამუშავების შემდეგ, საპროექტო კვლევის დროს მოხდა ისეთი შედეგების მიღება, სადაც აშკარა იყო მეტი ამოსაშლელი სიტყვის ჩამატება. ამის შემდეგ კი, კვლევათა შედეგად მიღებული მოდელების შედეგი ბევრად უკეთესი გახდა.

➤ *ლემატიზაცია და სტემინგი* – პროცესი, როდესაც წინადადებებიდან სიტყვების ამოშლა მათი ძირეული ფორმების პროცესინგისთვის ხდება.

• სტემინგი (Stemming – ფუძე) არის პროცედურა, რომლის საშუალებითაც შესაძლებელია ერთი და იგივე ფუძის მქონე ყველა სიტყვის საერთო ფორმამდე შემცირება;

• ლემატიზაცია (lemmatization) შლის ფლექსიურ დაბოლოებებს და აბრუნებს სიტყვის საბაზისო ან ლექსიკონის ფორმას.

მაგალითად, ინგლისურ ენაში არსებობს მრავალი სიტყვა, რომელთაც საერთო ფუძეს განსხვავებული მნიშვნელობა აქვს. მაგალითად „Generous“ (სულგრძელი) და „Generation“ (თაობა) (აგრეთვე Generously (სულგრძელად) და Generate (გენერირება). ამ სიტყვათა ფუძის გენერაცია, მისი მნიშვნელობების გასაგებად და ინფორმაციის საგრძნობლად შესამცირებლად საკმაოდ რთული და ჩახლართული პროცესია. სწორედ ამ პროცესს ეწოდება ლემატიზაცია და სტემინგი.

სტემინგის ყველაზე ხშირი და ეფექტური გამოყენების გზა პორტერის ალგორითმი გახლავთ [49]. 1980 წელს მარტინ პორტერსმა შეიმუშავა ალგორითმი, რომელიც იყენებს სიტყვათა შემცირების 5 ფაზას და თითოეულს გააჩნდა საკუთარი წესები სიტყვათა აგებულებისათვის. მაგალითად პირველი ფაზის წესები (ცხრილი_2.1).

პორტერის ალგორითმის პირველი ფაზის წესი ცხრ.2.1

სუფიქსი	შედეგი	მაგალითი	
SSES	SS	Plusses	Pluss
IES	I	Movies	Movi
SS	SS	Across	Across
S		Nights	Night

აქ ჩანს, რომ კონვერტაციის შემდეგ, აღნიშნული სუფიქსის მქონე სიტყვები გარდაქმნის შემდეგ შეიცვლება ისე, როგორც ეს „შედეგი“-ის სვეტშია ნაჩვენები, ხოლო კონკრეტული სიტყვებისათვის წესი ყველაზე გრძელი სუფიქსის მიხედვით იმოქმედებს.

სტემინგისაგან განსხვავებით, ლემატიზაცია გამოდის სიტყვების შემცირების ჩარჩოდან და ითვალისწინებს ენის სრულ ლექსიკას, რათა მოახდინოს მათი უკეთესი მორფოლოგიური ანალიზი. მარტივად რომ ვთქვათ, *ლემატიზაცია არის სხვადასხვა ფორმის მექანიზმი, ერთიდაიმავე სიტყვების გამოვლენა*. ლემატიზაცია ძალიან ხშირად გამოიყენება სხვადასხვა საძიებო სისტემებში. ლემატიზაცია ბევრად უფრო ინფორმატიულია ვიდრე უბრალოდ სტემინგი. თუმცა, უნდა აღინიშნოს რომ ის, შეიძლება ძალიან რთულად იქნას გამოყენებული არა ინგლისურ ენებში.

მას შემდეგ, რაც მონაცემთა წინასწარი დამუშავება შესრულდება, ხდება მათი დამუშავებისთვის სპეც-ალგორითმების დეველოპმენტი. ბუნებრივი ენის დამუშავების მრავალი ალგორითმი არსებობს, თუმცა მათ შორის ყველაზე ხშირად ორი გამოიყენება:

1. *წესებზე დაფუძნებული სისტემა* – იყენებს წინასწარ შემუშავებულ *ლინგვისტიკურ წესებს* და ძირითადად, ბუნებრივი ენის დამუშავების შექმნის დასაწყისისათვის იყო პოპულარული (თუმცა, გარკვეული თვალსაზრისით, დღესაც გამოიყენება);

2. *მანქანურ დასწავლაზე აგებული სისტემა* – იყენებს მანქანური დასწავლის ალგორითმებს, რომლებიც სწავლობს დავალებათა შესრულებას, მათთვის მიწოდებული სასწავლო მონაცემების მიხედვით. სისტემა თანამიმდევრულად სრულყოფს მიდგომას, მისთვის მეტი მონაცემების მიწოდების საფუძველზე.

ბუნებრივი ენის დამუშავების არსებული პრინციპები ემყარება სხვადასხვა ტიპის სამეცნიერო კატეგორიას და მათ განვითარებას. მასთან დაკავშირებული კვლევები 70 წელიწადზე მეტია არსებობს და სათავეს *ლინგვისტიკის სამეცნიერო სფეროდან* იღებს. მისი განვითარება განაპირობა, ისევე როგორც გამოთვლითი ლინგვისტიკის განვითარებამ (მე-20 საუკუნის 50-იანი წლებიდან), ასევე კომპიუტერული მეცნიერების თანამედროვე მიდგომებმა.

1950-იან წლებში, ალან ტიურინგმა (Alan Turing) პირველმა ჩაუყარა საფუძველი ბუნებრივი ენის დამუშავებას. მან შეიმუშავა ტესტი, იმის დასადგენად, იყო თუ არა კომპიუტერი რეალურად ინტელექტუალური. ეს ტესტი მსოფლიოში ტიურინგის ტესტითაა ცნობილი [50].

გასული საუკუნის 90-იან წლებამდე, ბუნებრივი ენის დამუშავება, ძირითადად წესებზე იყო დაფუძნებული (როგორც ზემოთ აღვნიშნეთ) და ტექსტის აღქმისათვის იყენებდა ენათმეცნიერების მიერ შემუშავებულ წესებს. ამის შემდეგ, კომპიუტერისა და პროგრამული ინჟინერიის განვითარებამ გამოიწვია ის, რომ ბუნებრივი ენის დამუშავება, ლინგვისტიკური წესებისაგან მეტად დამოუკიდებელი გახდა და ინჟინრულ მიდგომებზე გადავიდა.

ბუნებრივი ენის დამუშავება ჩვენი კვლევისა და სამეცნიერო სიახლის უმნიშვნელოვანესი ნაწილია. სწორედ მისი მეშვეობით იქნება შესაძლებელი *თემის მოდელირების კლასიფიკაციის ალგორითმების* სწორი და კორექტული გამოყენება.

2.3. თემის მოდელირება

თემის მოდელირება (Topic Modeling) არის მანქანური დასწავლის არა-კონტროლირებადი მეთოდი, რომლის მეშვეობითაც ხდება *სხვადასხვა ტიპის ტექსტური ინფორმაციისგან შემდგარი დოკუმენტების კატეგორიზაცია მასში არსებული ტექსტის, კერძოდ კი გამოყენებული სიტყვებისა და ფრაზების მიხედვით.*

თემის მოდელირება, ავტომატურად აანალიზებს ტექსტურ მონაცემებს და წინასწარ მითითებული კლასტერების რაოდენობის მიხედვით ანაწილებს ამ კლასტერებში. საბოლოო შედეგად კი ვიღებთ წინასწარ მითითებული რაოდენობის კლასტერს/კატეგორიას და მასში მოხვედრილი ტექსტურ მონაცემებს.

როგორც უკვე აღვნიშნეთ, თემის მოდელირების ალგორითმის გაშვებამდე, წინასწარ უნდა გვექონდეს განსაზღვრული მოსალოდნელი კლასტერების რაოდენობა.

რაოდენობის წინასწარი განსაზღვრა რთული პროცესია და ხშირად, მოდელირების ალგორითმის მრავალჯეროვანი გაშვებას მოითხოვს. მიღებული შედეგების მიხედვით კი უკვე ნათელი ხდება, არის თუ არა წინასწარ მითითებულ კლასტერთა რაოდენობა კვლევისათვის შესაბამისი. რადგან თემის მოდელირება მანქანური დასწავლის არაკონტროლირებად მეთოდს იყენებს, ის არ საჭიროებს სწავლებას და მათ ანალიზს ამის გარეშე ახდენს.

თუმცა, გარანტიის მიცემა იმაზე, რომ მოცემული მონაცემები აბსოლუტურად აკურატულია, შეუძლებელია. მაგრამ, ჩვენ კვლევაში, აშკარაა ფაქტი, რომ ის ძალიან კარგად მუშაობს რამდენიმე გაშვების შემდეგ. ჩვენ შევცვალეთ წინასწარ განსაზღვრული კლასტერების რაოდენობა, რამდენჯერმე დავამატეთ სიტყვები ამოსაშლელ სიტყვათა სიას და საბოლოოდ, შედეგი დამაკმაყოფილებელი გამოდგა.

ტექსტის ანალიზის დროს თემის მოდელირება ახდენს არა მხოლოდ ცალკეული სიტყვების ანალიზს, არამედ სიტყვათა ჯგუფისა და ფრაზების, რასაც მანქანური დასწავლის ენაზე „სიტყვების ჩანთა“ (Bag of words) ეწოდება. სიტყვათა ჯგუფის ანალიზის მიხედვით კი გასაგები ხდება, თუ როგორ არის დამოკიდებული სიტყვების მნიშვნელობა იმ კონტექსტზე, რომელშიც ის ბუნებრივ ენაშია გამოყენებული.

როგორც უკვე აღვნიშნეთ, ცალკეული სიტყვების ანალიზი არ არის საკმარისი იმისთვის რომ დადგინდეს ის, თუ კონკრეტულად რა იგულისხმება ამა თუ იმ კონტექსტში, ხოლო ფრაზების შესწავლით, კლასტერიზაცია მეტად ზუსტი ხდება.

ტექსტის კლასტერიზაციის მიზნით თემის მოდელირება მანქანური დასწავლის ერთადერთი საშუალება არ არის. ამისათვის კიდევ არსებობს *სემანტიკური ლატენტური ანალიზი*, *კლასტერული ანალიზი* და სხვ. მოცემული მიდგომები საკმაოდ საინტერესოა, თუმცა მიუხედავად ამისა არსებობს მიზეზები, რის გამოც ჩვენს კვლევაში გამოვიყენეთ *თემის მოდელირება* და არა ზემოთ ჩამოთვლილი სხვა მეთოდები.

თემის მოდელირება, სხვა ტიპის კლასტერული ანალიზის მარტივ ფორმებთან შედარებით, ორი მნიშვნელოვანი უპირატესობით გამოირჩევა:

1) K-means კლასტერინგში (რაც ტექსტის კლასიფიკაციის, შედარებით მარტივი და ერთ-ერთი ყველაზე პოპულარული ფორმაა), თითოეული ტექსტური დოკუმენტი (ჩვენ შემთხვევაში სამეცნიერო სტატია ან რეზიუმე), შეიძლება მიენიჭოს მხოლოდ ერთ კლასტერს ან კატეგორიას [96, 97]. თუმცა თემის მოდელირების დროს, კლასტერები სამეცნიერო ნაშრომებს (დოკუმენტებს) პროცენტულად ენიჭებათ, რაც ნიშნავს იმას, რომ თითოეულ ნაშრომს მიენიჭება ალბათობები იმისა, თუ რომელ კლასტერში შეიძლება ის მოიაზრებოდეს, მის ტექსტში არსებული ფრაზების მიხედვით. მაგალითად, რომელიმე სამეცნიერო ნაშრომი შესაძლოა 65%-ით სამედიცინო ფიზიკის თემებთან მიახლოებულ კლასტერს მიენიჭოს ხოლო 30%-ის ფიზიკის. ასეთ შემთხვევაში აშკარაა, რომ ნაშრომი სამეცნიერო ფიზიკის კლასტერს მიეკუთვნება;

2) მეორე ძირითადი უპირატესობა, რაც თემის მოდელირებას გააჩნია სხვა კლასტერული ანალიზებთან შედარებით, გახლავთ ის, რომ იგი იყენებს უფრო დახვეწილ განმეორებით ტექნიკას, იმის დასადგენად, თუ რამდენადაა კონკრეტული დოკუმენტი დაკავშირებული რომელიმე გენერირებულ თემასთან. ეს განმეორებითი ტექნიკა ბაიესის ტექნიკის სახელითაა ცნობილი [98, 99].

მის მიხედვით, დოკუმენტს თავდაპირველად შემთხვევითობის პრინციპით მიენიჭება რომელიმე კლასტერში ყოფნის ალბათობა, შემდეგ კი, რაც უფრო მეტი მონაცემის დამუშავება მოხდება, ეს ალბათობაც უფრო და უფრო ზუსტი გახდება.

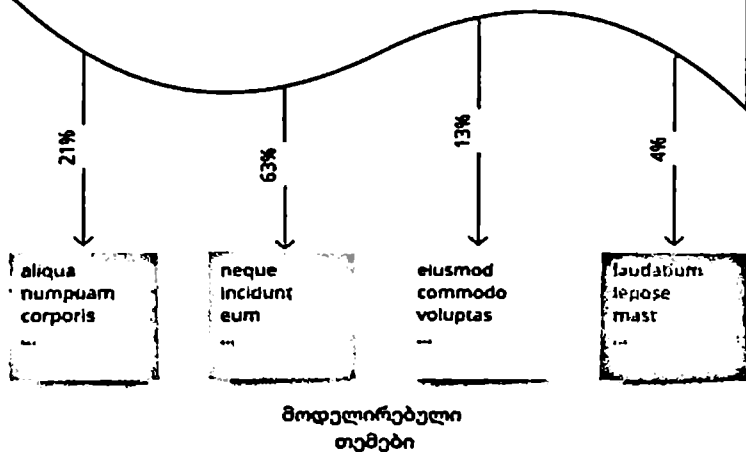
მნიშვნელოვანია აღინიშნოს ფაქტი, რომ თემის მოდელირების ალგორითმები არ ცდილობს იმას, რომ ჩაანაცვლოს ბუნებრივი ენაში არსებული ტექსტი. ამის ნაცვლად, ის, მისთვის მიცემული ტექსტური დოკუმენტების საფუძველზე, სწავლობს და შემდეგ გამოთქვამს ვარაუდებს იმის შესახებ, თუ როგორ არის გამოყენებული სიტყვები თემებში, სხვადასხვა შესწავლილი ფრაზისა და გამონათქვამის საფუძველზე.

ხშირ შემთხვევაში თემის მოდელირება, გაუგებარ და არაინფორმაციულ შედეგებს იძლევა. ეს ჩვენი კვლევის პირველ ეტაპებზეც ასე გამოიყურებოდა, თუმცა რამდენიმე ეტაპიანმა კვლევებმა და ალგორითმში ცვლილებების შეტანამ საბოლოოდ სასურველ შედეგამდე მიგვიყვანა [100].

იმისათვის, რომ უკეთ გავერკვეთ თუ როგორ მუშაობს თემის მოდელირების ალგორითმი, წარმოვადგინოთ დიაგრამა (ნახ.2.2), სადაც აღწერილია მოდელირების ძირითადი პრინციპები და მისი მიმდინარეობა.

ნახაზზე ილუსტრირებულია თუ როგორ არის მოცემული ტექსტის ნაწილი კლასიფიცირებული, სხვადასხვა მოდელირებული თემის მიხედვით. დიაგრამის ქვედა ნაწილში ხედავთ მოდელირებულ თემებს, სადაც აღნიშნულია რამდენიმე ისეთი სიტყვა, რომელიც ყველაზე ხშირადაა გამოყენებული მოცემულ თემაში. მაგალითად, პირველ თემაში ყველაზე ხშირად გამოყენებადი სიტყვებია „aliqua“, „numpuam“ და სხვ.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?



ნახ. 2.2. თემის მოდელირების ბუზაობის პრინციპი რელევანტური თემის პროცენტული შერჩევადობის მხრივ

როგორც ნახაზზე ჩანს, სხვადასხვა მოდელირებული თემები აღნიშნულია სხვადასხვა ფერით. დიაგრამის ზედა ნაწილში ხედავთ დოკუმენტს, რომელშიც კონკრეტული სიტყვები მონიშნულია მისი თემის შესაბამისი ფერებით. როგორც მაგალითშია გადმოცემული, დოკუმენტში ყველაზე ხშირად მეორე თემის შესაბამისი სიტყვები გვხვდება, შემდეგ ადგილზეა პირველი თემა და ა.შ. მოკლედ რომ ვთქვათ, იმის ალბათობა, რომ დოკუმენტი მეორე თემას ეკუთვნის 63%-ია, ხოლო დანარჩენ თემებზე ასე ნაწილდება: 1 – 21%, 3 – 13%, 4 – 4%. მიღებული შედეგიდან გამომდინარე შეგვიძლია ვივარაუდოთ, რომ დოკუმენტი მეორე თემის კლასტერს ეკუთვნის.

თემის მოდელირების ჩვენ საპროექტო ნაშრომში ინტეგრაციის შესახებ შემდეგ თავებში ვისაუბრებთ. ახლა კი თემის მოდელირების ის ორი უმნიშვნელოვანესი მეთოდი განვიხილოთ, რომლებიც ჩვენს პროექტში გამოვიყენეთ. ეს მეთოდებია დირიხლეს ლატენტური განაწილება (Latent Dirichlet allocation) და არაუარყოფითი მატრიცის ფაქტორიზაცია (Non-negative matrix factorization).

2.4. დირიხლეს ლატენტური განაწილება

როგორც წინა პარაგრაფში აღვნიშნეთ, თემის მოდელირება გვეხმარება იმაში, რომ მოხდეს სხვადასხვა ტიპის დოკუმენტების დაჯგუფება საერთო კლასტერის ქვეშ, მასში არსებული სიტყვებისა და ფრაზების მიხედვით (ნახ.2.2). აქ ვხედავთ თუ რა პრინციპით ხდება დოკუმენტისათვის კონკრეტული თემის მინიჭება, თუმცა ვერ ვხედავთ იმას, თუ როგორ შეიქმნა თავად სხვადასხვა ტიპის დოკუმენტები. ამ პრობლემის მოსაგვარებლად თემის მოდელირების ერთ-ერთი ყველაზე პოპულარული მიდგომა, დირიხლეს ლატენტური განაწილება გამოვიყენეთ.

დირიხლეს ლატენტური განაწილება (LDA) არის თემის მოდელირების, არაუარყოფითი მატრიცის ფაქტორიზაციის მეთოდთან ერთად, ყველაზე პოპულარული ტექნიკა დოკუმენტთა კრებულიდან სხვადასხვა თემების ამოსაღებად [30, 31].

ამოსაღები თემები, რომლებიც ალგორითმის გაშვების შემდეგ უნდა გენერირდეს, ჯერ კიდევ უცნობია. სწორედ ამიტომ ვიყენებთ სიტყვა ლატენტურს. სიტყვა ლატენტური, გულისხმობს რაიმეს, რაც უკვე არსებობს, მაგრამ ფარულია, ჯერ არ არის განვითარებული.

აღნიშნული განაწილება გერმანელი მათემატიკოსის პიტერ ლესტაჟ ლეჟუნ დირიხლეს (Lejeune Dirichlet) სახელს ატარებს და ამ პროცესის მთავარი იდეა გახლავთ განაწილება-განაწილებებზე, რაც იმას ნიშნავს, რომ ამ პროცესის თითოეული სვლა თავად განაწილებაა.

თეორიულად რთული წარმოსადგენია, თუმცა დირიხლეს პროცესები გახლავთ რომელიმე დოკუმენტში არსებული თემების ალბათობების განაწილება იქ, სადაც ამ განაწილების დიაპაზონი თავადვე გახლავთ ალბათობის განაწილება.

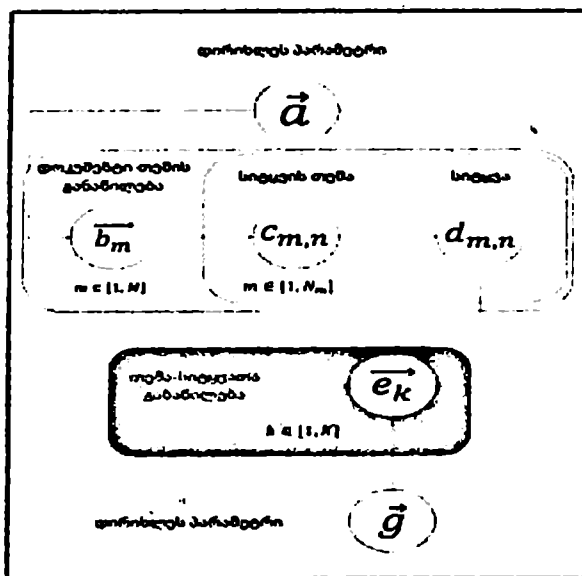
თემის მოდელირების დროს აღნიშნული პროცესი გვცემარება იმაში, რომ დავადგინოთ ის, თუ რა არის იმის ალბათობა, რომ რომელიმე სიტყვა ან სიტყვათა ჯგუფი, რომელიც კონკრეტულ დოკუმენტში გვაქვს თავმოყრილი, კიდევ გამოერდეს სხვა დოკუმენტებშიც.

ეს ალგორითმს საშუალებას აძლევს შეიმუშაოს მოდელი, რომელიც შექმნის ისეთ მონაცემებს, რომლებიც მოგვცემს საშუალებას, რომ გვქონდეს მოლოდინი იმისა, იარსებებს თუ არა ეს სიტყვები და ფრაზები სხვა მსგავს დოკუმენტებშიც.

დირიხლეს ლატენტური განაწილების წყალობით, ვახდენთ ალბათობათა გენერირებას ისეთი სიტყვებისათვის, რომლებიც

სხვადასხვა თემებს შეიძლება წარმოადგენდეს. დირიხლეს პროცესების მიხედვით, დოკუმენტები არის სხვადასხვა თემების ალბათობის შემადგენელი, ხოლო თემები კი სხვადასხვა სიტყვებისა [51].

2.3 ნახაზზე ნაჩვენებია დირიხლეს ლატენტური განაწილების რეპრეზენტაცია, სადაც მწვანე ბლოკში მოთავსებულია ყველა დასამუშავებელი დოკუმენტი (ჩვენ შემთხვევაში სამეცნიერო ნაშრომი ან რეზიუმე), რომელიც წარმოდგენილია M სიმბოლოთი. ყვითელ ფერში კი კონკრეტული დოკუმენტის სიტყვებია წარმოდგენილი, რომელთა რაოდენობაც N -ის ტოლია, d კი კონკრეტულ სიტყვას გულისხმობს.



ნახ.2.3. დირიხლეს ლატენტური განაწილების მოდელი

დირიხლეს ლატენტური გადახრის მეთოდოლოგის მიხედვით, ყოველი ეს სიტყვა რომელიმე თემასთანაა დაკავშირებული. ეს თემები c სიმბოლოთია ნაჩვენები. ამ სიტყვათა ჯამური რაოდენობის მიხედვით k_i ვიღებთ b სიმბოლოს მნიშვნელობას, რომელიც დოკუმენტის თემას წარმოადგენს.

როგორც ნახაზზე ჩანს, დირიხლეს ლატენტური განაწილების მოდელს ორი შემავალი პარამეტრი აქვს: 1 – a , რომელიც დოკუმენტისთვის თემის მინიჭებას განსაზღვრავს და 2 – g , რომელიც განსაზღვრავს თემისათვის სიტყვების განაწილებას [52]. რაც შეეხება დირიხლეს ლატენტური განაწილების დადებით და უარყოფით მხარეებს, ყველაზე მეტად აღსანიშნავია შემდეგი:

დადებითი – როგორც უკვე აღვნიშნეთ, თითოეული დოკუმენტი განიხილება, როგორც რამდენიმე სხვადასხვა თემის ნაზავი და აქვს ამ თემათა პროცენტული ალბათობა. აგრეთვე ერთ-ერთი უპირატესობაა ის, რომ მან წინასწარ არ იცის თუ რა შეიძლება იყოს პოტენციურად მიღებული თემები.

უარყოფითი – ალგორითმი დამოკიდებულია მონაცემთა სიმრავლის დალაგებაზე და, აქედან გამომდინარე, ყოველი შემდეგი კვლევა, შესაძლოა განსხვავებული გამოდგეს (თუმცა თანამედროვე მანქანური დასწავლის Python პროგრამირების ენის ბიბლიოთეკებში, არსებობს ამ პრობლემის მოგვარების სხვადასხვა გზა და სწორედ ეს გზები გამოვიყენეთ ჩვენი კვლევისთვის).

2.5. არაუარყოფითი მატრიცის ფაქტორიზაცია

არაუარყოფითი მატრიცის ფაქტორიზაცია (Non-negative matrix factorization – NMF) თემის მოდელირების, ერთ-ერთი ყველაზე პოპულარული და საუკეთესო ტექნიკაა. იგი მანქანური დასწავლის არაკონტროლირებადი ალგორითმია და ერთდროულად ახდენს როგორც განზომილებათა შემცირებას, ასევე მის

კლასტერიზაციას. არაუარყოფითი მატრიცის ფაქტორიზაცია, პირველად 1994 წელს იქნა შექმნილი ფინელი მეცნიერების პენტი პაატეროს (Pentti Paatero) და უნტო ტაპერის (Unto Tapper) მიერ [38], ხოლო პოპულარობა 1999 წლის სტატიით მოიპოვა, ამერიკელი დანიელ ლის (Daniel Lee) და ინგლისელი სებასტიან სეუნგის (Sebastian Seung) თანაავტორობით [52].

მოცემული ალგორითმის გამოყენება შეგვიძლია მოვახდინოთ TF-IDF (ტერმინის სიხშირე – შებრუნებული დოკუმენტის სიხშირე, იგივე Term frequency-inverse document frequency) -თან ერთად სხვადასხვა დოკუმენტის თემების მოდელირებისათვის.

TF-IDF სტატისტიკური საშუალებაა იმის, რომ მოხდეს გაზომვა თუ რამდენად შეესაბამება ესა თუ ის სიტყვა რომელიმე კონკრეტულ დოკუმენტს, სხვადასხვა დოკუმენტთან სიმრავლეში. აღნიშნული მიიღწევა ორი მნიშვნელობის ნამრავლით:

1) *ტერმინთა სიხშირე დოკუმენტში* – გულისხმობს იმას, თუ რამდენადაა სიტყვა გამოყენებული კონკრეტულ დოკუმენტში. ამ სიხშირის გამოთვლის რამდენიმე გზა არსებობს. ყველაზე მარტივი კი უბრალოდ დოკუმენტში სიტყვათა პირდაპირი დათვლაა. ამის შემდეგ კი შესაძლებელია ამ სიხშირეთა კორექტირება, დოკუმენტის ზომის, ან მასში ყველაზე მეტად გამოყენებული სიტყვების სიხშირის მიხედვით;

2) *დოკუმენტის ინვერსიული სიხშირე* – გულისხმობს დოკუმენტში სიტყვათა ინვერსულ სიხშირეს სხვადასხვა დოკუმენტთან ერთობლიობაში. მარტივად რომ ვთქვათ, ეს ნიშნავს იმას თუ რამდენად ხშირადაა ესა თუ ის სიტყვა გამოყენებული მთლიან დოკუმენტთან სიმრავლეში. რამდენადაც ხშირადაა რომელიმე კონკრეტული სიტყვა გამოყენებული სხვადასხვა დოკუმენტში, მით მეტად უახლოვდება მისი მნიშვნელობა 0-ს. მოცემული მეტრიკა შემდეგნაირად გამოითვლება – ვიღებთ დოკუმენტთა სრულ

რაოდენობას, ვყოფთ მას იმ დოკუმენტთა რაოდენობაზე, რომელიც შეიცავს კონკრეტულ სიტყვას და ვითვლით ლოგარითმს.

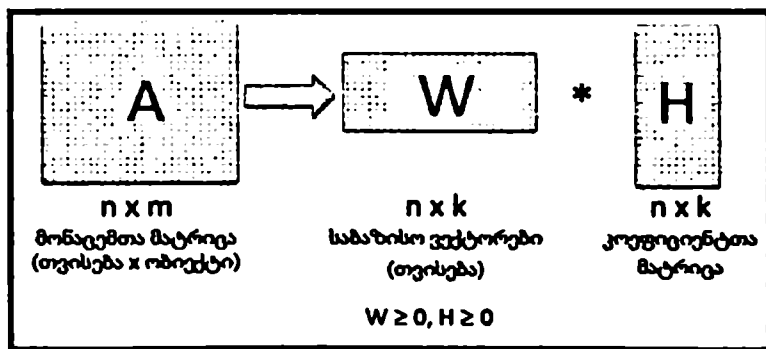
კიდევ ერთხელ აღვნიშნოთ ის, რომ რაც უფრო ხშირადაა რომელიმე სიტყვა სხვადასხვა დოკუმენტებში გამოყენებული, მით უფრო უახლოვდება ის 0-ს, ხოლო მეორეს მხრივ პირიქით, რაც უფრო იშვიათად გვხვდება ის, რიცხვის მნიშვნელობა 1-ისკენ მიიწევს. საბოლოო ჯამში, ამ ორი მონაცემის ერთმანეთზე გადამრავლებით, ვიღებთ დოკუმენტში TF-IDF ქულას და რაც უფრო მაღალი იქნება ამ ქულის მნიშვნელობა, მით უფრო მნიშვნელოვანია ეს სიტყვა დოკუმენტისათვის.

წარმოვადგინოთ TF-IDF ფორმულა:

$$tfidf(t, d, D) = tf(t, d) * idf(t, D)$$

სადაც t კონკრეტული სიტყვაა, d - რომელიმე დოკუმენტი, ხოლო D დოკუმენტთა სიმრავლე.

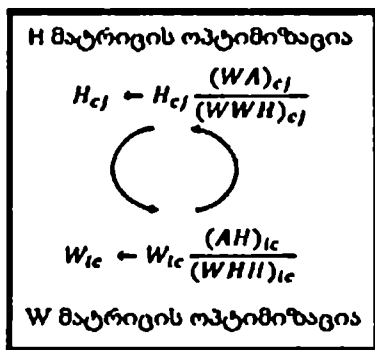
2.4 ნახაზზე გვაქვს A მატრიცა, რომლის სვეტები ობიექტებია და სტრიქონები - თვისებები, რომელიც როგორც უკვე აღვნიშნეთ TF-IDF შედეგად მიღებული მატრიცაა.



ნახ. 2.4. არაუარყოფითი მატრიცის
ფაქტორიზაცია

ჩვენ გვინდა, რომ k შემცირებული განზომილებების (საბაზისო ვექტორები) წრფივი კომბინაციით, მივუახლოვდეთ თითოეულ ობიექტს, რომელიც A მატრიცის სვეტებია. თითოეული საბაზისო ვექტორი k შეიძლება განიმარტოს როგორც კლასტერი, ხოლო ამ კლასტერის შემადგენელი ნაწილები წარმოდგენილია როგორც H მატრიცა. აქედან გამომდინარე, იმისთვის რომ ამ მიდგომამ პრაქტიკულად სწორად იმუშაოს, ალგორითმში გვაქვს ორი შემავალი პარამეტრი. ნახაზზე მოცემულია A მატრიცა (TF-IDF მატრიცა) და ბაზისური ვექტორების რაოდენობა k , რაც ჩვენ შემთხვევაში, თემის მოდელირებისას შერჩეულ თემათა რაოდენობას გულისხმობს. ასე, რომ ისევე, როგორც დირიხლეს ლატენტურ განაწილებაში, საჭიროა, რომ წინასწარ განვსაზღვროთ, იმ თემათა რაოდენობა, რა რაოდენობაზეც გვინდა, რომ ჩვენი სამეცნიერო ნაშრომები და რეზიუმეები კლასტერიზდეს.

საპროექტო კვლევაში (ისევე როგორც არაუარყოფით მატრიცის ფაქტორიზაციაში მიღებული) თემათა რაოდენობის არჩევის შემდეგ, 2.5 ნახაზზე ასახულ W და H მატრიცების, მნიშვნელობების მინიჭება შემთხვევითობის პრინციპით მოხდა,



ნახ. 2.5. H და D მატრიცათა ოპტიმიზაცია

შემდეგ k ჩატარდა ამ მნიშვნელობათა ოპტიმიზაცია. ასე რომ გვაქვს ოპტიმიზაციის მაქსიმიზაციის მოლოდინი, რათა არსებითად დავხვეწოთ W და H მატრიცები.

მატრიცათა დახვეწა მდგომარეობს იმაში, რომ უნდა გაგრძელდეს კოეფიციენტთა განახლება მანამდე, სანამ W და H მატრიცების აპროქსიმაცია აზრს არ შეიძენს. ამის საერთო მიდგომა კი მდგომარეობს იმაში, რომ მოხდეს H და შემდეგ W მატრიცების განმეორებითი განახლება, სანამ ისინი შესათვისნი არ გახდება.

განხილული არაუარყოფითი მატრიცის ფაქტორიზაციის მეთოდის დადებით და უარყოფით მხარეები ასეთია:

დადებითი – აღნიშნული მიდგომა წრფივი ალგებრის მოდელია, რომელიც მაღალგანზომილებიან ვექტორებს, დაბალგანზომილებიან წარმოდგენაში აყალიბებს და იყენებს იმ ფაქტს, რომ ვექტორები არაუარყოფითია;

უარყოფითი – მიდგომა შედარებით უკეთ მოქმედებს მცირე ზომის დოკუმენტებზე, თუმცა ასევე არ მუშაობს ცუდად დიდი ზომის დოკუმენტებზეც.

რაც შეეხება დირიხლეს ლატენტური განაწილების და არაუარყოფით მატრიცის ფაქტორიზაციის შედარებას, ეს ალგორითმები განსხვავებულ მიდგმას იყენებს (დირიხლეს ლატენტური განაწილება – *სავარაუდო მოდელია*, ხოლო არაუარყოფით მატრიცის ფაქტორიზაცია – მულტიპარამეტრული ანალიზის ტექნიკა), თუმცა მიზანი ერთია.

მეტი სიზუსტისათვის კვლევაში ორივე მეთოდი გვაქვს გამოყენებული, ხოლო კვლევის დროს ამ ორი მეთოდის გამოყენების შედეგად მიღებული მოდელი ინოვაციური სამეცნიერო სიახლეა. მათი მუშაობის მცირე განსხვავებები შემდეგნაირია:

დირიხლეს ლატენტური განაწილება:

- 1) კარგად მუშაობს დიდ დოკუმენტებთან, როგორებიცაა, ჩვენ კვლევაში გამოყენებული სრული სამეცნიერო ნაშრომთა ტექსტები;
- 2) მოდელი ვითარდება ყოველი ახალი დოკუმენტის დამატების შემდეგ;

3) ვარაუდობს, რომ თითოეული დოკუმენტი რამდენიმე თემას მიეკუთვნება;

4) ერთიდაიმავე მონაცემებისათვის შესაძლოა ყოველთვის განსხვავებული შედეგი მიიღოს;

არაუარყოფით მატრიცის ფაქტორიზაცია:

1) უკეთ მუშაობს მოკლე ტექსტებთან, ისეთთან როგორიცაა ჩვენ შემთხვევაში სამეცნიერო სტატიების რეზიუმეები;

2) ჩვენი კვლევების დროს გამოჩნდა, რომ ის საკმაოდ სწრაფია, ვიდრე დირიხლეს ლატენტური განაწილება;

3) დოკუმენტი აგრეთვე მიეკუთვნება რამდენიმე თემას;

4) ვარაუდის ნაცვლად, ალგებრულად ითვლის, თუ რამდენად შეესაბამება თითოეული დოკუმენტი რომელიმე თემას.

2.6. კლასიფიკაცია

მანქანური დასწავლის კიდევ ერთი მნიშვნელოვანი მეთოდი, რომელსაც კვლევისათვის ვიყენებთ, *ტექსტის კლასიფიკაციაა*. კლასიფიკაცია მანქანური დასწავლის ერთ-ერთი უმთავრესი და ყველაზე პოპულარული ტექნიკაა.

შესაბამისად, ტექსტის კლასიფიკაციაც ფუნდამენტური მიდგომაა ბუნებრივი ენის დამუშავების სფეროში და გამოიყენება ისეთი დავალებების შესასრულებლად, როგორიცაა ჩვენ შემთხვევაში, ახალი ნაშრომისთვის *კატეგორიის ავტომატური მინიჭება*, აგრეთვე *სენტიმენტური ანალიზისათვის*, *სპამის აღმოსაჩენად*, *თემის ლაბელინგისთვის* და სხვ.

მოცემული მეთოდი კონტროლირებადია და ტრენინგისათვის საჭიროებს მარკირებულ მონაცემებს (მონაცემებს უკვე არსებული კატეგორიით).

კლასიფიკაციის მოდელს ჩვენ ნაშრომში მნიშვნელოვანი როლი უკავია. როგორც უკვე აღვნიშნეთ, როდესაც მომხმარებელი

დაიწყებს ნაშრომზე მუშაობას, სისტემამ მისი რეზიუმეს ტექსტის მიხედვით, პირდაპირ უნდა გამოიცნოს ის, თუ კონკრეტულად რა სფეროს ნაშრომზე მუშაობს ავტორი და შესთავაზოს შესაბამისი ინსტრუმენტები და ინტერფეისი მუშაობის პროცესის გასამართივებად.

კონკრეტულად, ტექნოლოგიის გამოყენების მხრივ, სამი უმთავრესი მიზეზია იმისა თუ რატომ ვიყენებთ ტექსტის კლასიფიკაციას.

1) *მასშტაბურობა* – ხელით კლასიფიკაცია საკმაოდ რთულია და დიდ დროს მოითხოვს. მანქანური დასწავლის კლასიფიკაციის მეთოდის გამოყენებით კი შევისწავლეთ მილიონობით სამეცნიერო ნაშრომი და მიღებული მოდელის მეშვეობით ბევრად უფრო სწავად მოვახდინეთ ახლი ნაშრომის კლასიფიკაცია;

2) *ანალიზი რეალურ დროში* – როგორც უკვე აღვნიშნეთ, მომხმარებელი, რეალურ დროში მუშაობის დასაწყისშივე იღებს სასურველ შედეგს. უნდა აღინიშნოს, რომ სისტემაში, ყოველი ახალი დამატებული დოკუმენტი მას კიდევ უფრო ჰკვიანს ხდის;

3) *უტყუარი შედეგები* – მანქანური დასწავლის შედეგად მომხმარებლისთვის ავტომატურად შეთავაზებული ინტერფეისი ბევრად უფრო ზუსტია, ვიდრე თავად მომხმარებლის მიერ შერჩეული კატეგორია, რადგან გამორიცხულია ადამიანური შეცდომის ფაქტორი. მანქანურ დასწავლაში ტექსტის კლასიფიკაციის სხვადასხვა მეთოდი არსებობს. ჩვენ კვლევაში რამდენიმე ყველაზე პოპულარული მეთოდი გამოვიყენეთ და შედეგები ერთმანეთს შევადარეთ, რომ საბოლოოდ სასურველი მეთოდი აგვეჩრია [53].

კერძოდ, კვლევაში გამოყენებული პოპულარული ალგორით-მებია: ბაიესის მიაშიტ ალგორითმთა ოჯახი [26, 27] და დამხმარე ვექტორული მანქანები (Support vector machines - SVM) [53, 54].

სტატისტიკურ ალგორითმთა მიამიტი ბაიესის ოჯახი, ტექსტის კლასიფიკაციის სფეროში, ერთ-ერთი ყველაზე ხშირად გამოყენებული ალგორითმია, რადგან მისი ერთ-ერთი შემადგენელი ალგორითმის, მულტინომინალური ბაიესის მიამიტი ალგორითმის მეშვეობით, სასურველი შედეგის მიღება, შედარებით ნაკლები რაოდენობის მონაცემებითაცაა შესაძლებელი.

ბაიესის მიამიტი ალგორითმები დაფუძნებულია ბაიესის თეორემაზე, რომელიც გვცხმარება გამოვთვალოთ ორი მოვლენის ალბათობა, თითოეული მათგანის ერთმანეთისგან დამოუკიდებლად დადგომის საფუძველზე.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

ფორმულიდან გამომდინარე, შეგვიძლია ვიპოვნოთ A-ს დადგომის ალბათობა იმის გათვალისწინებით რომ B მოხდა. შესაბამისად, B - რეალურად მომხდარი ხოლო A ნავარაუდევია მნიშვნელობაა, რომელთაც ერთმანეთზე დამოკიდებულება არ აქვს.

დამხმარე ვექტორული მანქანები, არის ერთ-ერთი ყველაზე ძლიერი ალგორითმი ტექსტის კლასიფიკაციისთვის. ბაიესის მსგავსად, არც მოცემულ ალგორითმს სჭირდება სასწავლო მონაცემთა ძალზე დიდი რაოდენობა, იმისთვის რომ შესწავლა მოახდინოს. თუმცა, დამხმარე ვექტორული მანქანების ალგორითმი შედარებით მეტ გამოთვლით რესურსს მოითხოვს, თუმცა, ამ ალგორითმის მიღებული შედეგები კიდევ უფრო ზუსტია.

ალგორითმი მონაცემთა დამუშავების დროს ავლებს წრფეს და მონაცემებს 2 ნაწილად ჰყოფს. პირველი მონაცემები – მიეკუთვნება კონკრეტულ ჯგუფს და მეორე მონაცემები – რომლებიც მას არ მიეკუთვნება. რაც უფრო შორსაა წრფე თითოეული მათგანისგან, მით უფრო ოპტიმალურია იგი.

2.7. მონაცემთა განთავსება სერვერზე

სერვერული მონაცემების განთავსება მნიშვნელოვანი პროცესია ნებისმიერ ვებ აპლიკაციისთვის. მოცემულ პარაგრაფში ვისაუბრებთ იმაზე, თუ სად უნდა განთავსდეს, პროგრამული კოდი, სისტემის სერვერული მხარე და მონაცემთა ბაზები. აგრეთვე ვისაუბრებთ, რა ტექნოლოგიური საშუალებები უნდა იქნას გამოყენებული სისტემის სერვერულ მხარეს.

როდესაც სისტემა განკუთვნილია გლობალური მოხმარებისთვის და არ არის ორიენტირებული რომელიმე გეოგრაფიულ ლოკაციაზე, საჭიროა ის განთავსებულ იყოს სერვერებზე გლობალური მასშტაბით, რადგან სერვერზე წვდომის სისწრაფე პირდაპირ პროპორციულია, მოთხოვნის გამშვების მასთან გეოგრაფიულად ახლოს მყოფთან. თუ სერვერი ტერიტორიულად მდებარეობს ჩრდილოეთ ამერიკაში და მასზე წვდომა ავსტრალიიდან ხდება, ინფორმაციის გაცვლის სისწრაფე, მანძილის შესაბამისად, დაბალია. აქედან გამომდინარე, შეგვიძლია დავასკვნათ, რომ თუ გვჭირდება სისტემის ისეთ სერვერზე განთავსება, სადაც ის სწრაფად იქნება ხელმისაწვდომი ნებისმიერი ტერიტორიისთვის, საჭიროა სერვერების გლობალური განთავსებაც.

თანამედროვე მიღწევები ამის საშუალებას *ღრუბლოვანი სერვისების* საშუალებით იძლევა (მაგალითად. AWS) [56]. შესაბამისად, დასაწყისისთვის საჭიროა გამოვიყენოთ ჩვენთვის ყველაზე კომფორტული ღრუბლოვანი სერვისი. ჩვენ შემთხვევაში ამ სერვისით შესაძლებელი უნდა იყოს თავისუფლად მოხდეს მონაცემთა ბაზებისა და სერვერის აპლიკაციის უსაფრთხოდ განთავსება, უნდა იყოს ხელშემწყობი ავტომატური ტესტირებისა და CI/CD მენეჯმენტში.

დღეისათვის, მსოფლიოში ერთ-ერთი ყველაზე გავრცელებული და სანდო ღრუბლოვანი სერვისია AWS (Amazon Web

Services), იგი გვთავაზობს კვლევებისთვის მნიშვნელოვან და საინტერესო ღრუბლოვანი სერვისებს, როგორებიცაა AWS S3 (Simple Storage Service), EC2 (Elastic Compute Cloud) და სხვ. [57-59].

AWS ზურგს უმაგრებს ტექნოლოგიური გიგანტი ამაზონი, რაც მას კონკურენტებთან (მაიკროსოფტი, გუგლი და სხვ.) შედარებით კიდევ უფრო სანდოს ხდის, როგორც სტაბილურობის, ასევე უახლესი ტექნოლოგიების გამოყენების თვალსაზრისით [56].

2.7.1 მონაცემთა ბაზები

საპროექტო სისტემის სრულყოფილი მუშაობისთვის საჭიროა სამი სახის მონაცემთა ბაზა:

1) *რელაციური მონაცემთა ბაზა* [61, 62], რომელიც განთავსებულ იქნება სერვერზე, ამ შემთხვევაში არჩევანი გაკეთდა MySQL პაკეტის ბაზაზე, რომელიც გამოირჩევა მისი პრაქტიკული შესაძლებლობებით და სიმარტივით [63, 64];

2) *რეალური დროის მონაცემთა ბაზა*, რომლის მუშაობის დროს მის მონაცემთა ცვლილებებისთანავე, კლიენტის აპლიკაციის მხარესაც ავტომატურად განახლდება. მათ სისტემაში გამოყენებას შეეძლებოდა Google Firebase პროდუქტის საშუალებით;

3) *NoSQL მონაცემთა ბაზა*, რომელიც განთავსებულია სმარტ მოწყობილობებში და პროგრამის ინტერნეტთან წვდომის გარეშეც მუშაობს ბრაუზერის ლოკალურ მეხსიერებაში [27, 33].

2.7.2 ინსტრუმენტები რეალური დროის ფუნქციების

შესასრულებლად

როგორც წინა პარაგრაფში აღვნიშნეთ, მონაცემთა ბაზების სერვერზე შესანახი ფაილებისა და პროგრამული კოდისთვის, ერთი შეხედვით, ოპტიმალური ღრუბლოვანი სერვისია AWS, რომელიც იძლევა საშუალებას სწრაფად და ტრაფიკისა და მეხსიერების

შეუზღუდავად იქნას მიღებული სასერვერო კომპიუტერის მომსახურება. ჰოსტინგი საშუალებას იძლევა ვმართოთ და ავტომატურ რეჟიმში გავუშვათ მონაცემთა ბაზების სარეზერვო ასლების შექმნა, ვუზრუნველყოთ სისტემის უსაფრთხოება და სხვ.

აღსანიშნავია, რომ სერვისს აქვს საკმაოდ კომფორტული მომხმარებელთა სტატისტიკის, ელ-ფოსტის მართვის, SSH და სხვა მრავალი ფუნქციონალი, რომელიც პროგრამის მუშაობისას იქნა გამოყენებული.

Google კომპანიის Firebase პროდუქტი არის მრავალფუნქციურობით გამორჩეული სისტემა. იგი აერთიანებს ისეთ ფუნქციებს, როგორიცაა რეალური დროის მონაცემთა ბაზები, ღრუბლოვანი მესიჯების სისტემა, აუტენტიფიკაცია და ავტორიზაცია, ჰოსტინგი, დინამიკური ბმულები, ნოტიფიკაციები და სხვა მრავალი [64].

პროდუქტს აქვს მრავალი ისეთი ფუნქცია, რომელიც მონაცემთა განთავსების პროცესს ამარტივებს, თუმცა, ფასიანია და საკმაოდ ძვირიცაა, სხვა ღრუბლოვანი ტექნოლოგიებისგან განსხვავებით. შესაბამისად, შესაძლებელია მისი რესურსების რაიმე კონკრეტული მცირე დავალებებისთვის გამოყენება. მაგალითად, როდესაც საჭიროა რეალური დროის მონაცემთა ბაზები, ან სისტემის სიახლეთა გამოწერა შეტყობინებების მისაღებად, კომენტარების დასაწერად და ა.შ. უნდა მოხდეს Firebase-ის შესაძლებლობის გამოყენება, მაგრამ პერიოდულად მონაცემები უნდა იქნას გადატანილი უფრო სანდო რელაციურ მონაცემთა ბაზაში, აპლიკაციის სერვერზე და ამით განთავისუფლდეს Firebase სისტემაში დაკავებული მეხსიერება. ოპტიმიზაციის მიზნით ძალზე კომფორტულია ანალიტიკის ფუნქციონალის გამოყენება გამოკვლევებისა და ანალიზისთვის, რაც დაგეხმარება სწორი მიმართულებების არჩევაში პროგრამის განვითარების დროს.

Firestore ინსტრუმენტი, როგორც უკვე აღვნიშნეთ, ნაშრომში გამოიყენება რეალური დროის პროცედურების დროს. იგი მნიშვნელოვან როლს თამაშობს კომენტარების დაწერის და პროდუქტის განხილვისას, მომხმარებლებს შორის შეტყობინებების გაგზავნისა და მიღების დროს, გამოყენებული იქნება ნოტიფიკაციები და ღრუბლოვანი შეტყობინებები.

Firestore-ს ყველა ფუნქცია სწრაფი და ადვილად გამოსაყენებელია. მაგალითად, Google გარანტიას იძლევა მისი პროდუქტიდან ღრუბლოვანი მესიჯინგის გამოყენების დროს, რომელიც შესაძლებელია განხორციელდეს მოწყობილობაზე ნოტიფიკაციების გაგზავნის დროს, ინტერნეტთან დაკავშირებულ მომხმარებელთა 95%-ს ის მეოთხედი წამის განმავლობაში მიუვა. ასეთივე, მაღალი სისწრაფით მუშაობს რეალური დროის მონაცემთა ბაზაც.

სასარგებლო და ძალზე მნიშვნელოვანი ფუნქციაა სისტემის უსაფრთხოების შესაძლებლობა, რაც მონაცემთა ბაზისთვის მომხმარებელთა როლების და უფლებების არსებობას გულისხმობს.

2.8. ბიბლიოთეკები და პროგრამული ენები

პროექტისთვის საჭიროა პროგრამირების არაერთი ენისა და ბიბლიოთეკის გამოყენება. მათ სამ ნაწილად ვყოფთ, ესენია:

- 1) სამეცნიერო კვლევა მანქანური დასწავლის ალგორითმების გაშვებისთვის;
- 2) აპლიკაციის სერვერული ნაწილი და
- 3) კლიენტის აპლიკაცია.

მანქანური დასწავლის ალგორითმების გასაშვებად, ზოგადად პროგრამირების Python ენა და მასში არსებული ბიბლიოთეკები იქნა გამოყენებული. რაც შეეხება დამატებით ბიბლიოთეკებს, ამ შემთხვევაში მოკვლეული და შემდეგ გამოყენებულ იქნა პროექტის შესასრულებლად საჭირო, მსოფლიოში საუკეთესო ღია კოდის მქონე ბიბლიოთეკები.

პირველ რიგში აღსანიშნავია „sklearn“ ბიბლიოთეკა. მანქანური დასწავლისთვის განკუთვნილი მისი მრავალი ფუნქციონალის საშუალებით, სამეცნიერო კვლევის პროცესი ბევრად უფრო სწრაფი იყო, ვიდრე მის გარეშე. ბიბლიოთეკა გამოყენებულ იქნა, როგორც მონაცემთა წინასწარი მომზადებისთვის, აგრეთვე მანქანური დასწავლის ალგორითმების გასაშვებად და მისი ტესტირებისთვის.

მიღებული მოდელების შესანახად გამოყენებულ იქნა ბიბლიოთეკა „pickle“, ხოლო შედეგების ვიზუალიზაციისთვის კი „pyLDAvis“ და „matplotlib“-ი [65-67].

სერვერის მხარეს, პირველ რიგში საჭიროა პროგრამული კოდები, რომლებიც მოახდენს მონაცემთა ბაზებთან კავშირს, უზრუნველყოფს უსაფრთხოებას, მიაწვდის კლიენტს მოთხოვნილ მონაცემებს და ა.შ. ასეთი პროგრამული ენებიდან აღსანიშნავია Python, JavaScript და C# ენები [35, 68-71]. ჩვენ შემთხვევაში, გამოყენებული იქნა JavaScript-ი, node.js ტექნოლოგიით.

რაც შეეხება კლიენტის მხარეს გამოყენებულ ბიბლიოთეკას, ამ შემთხვევაში უპირატესობა კომპანია google-ის თანამშრომლის მიერ შექმნილ Angular-ს მიენიჭა, რაც იქიდანაა გამომდინარე, რომ მისი ფუნქციონალი სხვა ბიბლიოთეკებისგან განსხვავებით უფრო სრულყოფილია და ნაკლებად მოითხოვს დამატებითი ბიბლიოთეკების ჩართულობას [72].

ასევე, ერთ-ერთი მიზეზია ის, რომ მისი პროგრამირების ძირითადი ენა TypeScript-ია. აღნიშნული ენა კი გაუმჯობესებული JavaScript-ია მონაცემთა ტიპების დამატებითა და მეტად ობიექტზე ორიენტირებული მიდგომით.

აღსანიშნავია, რომ საჭიროა გამოყენებულ იქნას ზემოთ აღნიშნული პროგრამული ენებისა და ბიბლიოთეკების უახლესი ვერსიები [73-75].

2.9. პროგრესული ვებ აპლიკაცია (PWA)

როგორც უკვე შევნიშნეთ, პროექტში შექმნილი ბიზნეს მოთხოვნის მიხედვით, მიღებული სამეცნიერო სიახლე და პრაქტიკული ღირებულების საბოლოოდ გამოყენება ვებ აპლიკაციის საშუალებით ხდება. ამის შესახებ შეიძლება გაჩნდეს კითხვა, რატომ ვებ აპლიკაცია და არა სხვა, ლოკალური პროგრამა, როგორც Microsoft Word ან LaTeX-ია?

ამის მთავარი მიზეზი პროგრესული ვებ აპლიკაციების (PWA - Progressive Web Apps) უპირატესობაა. როგორც web.dev პლატფორმაზე ვკითხულობთ: „ვები უნიკალური სისტემაა, რომელიც საშუალებას გვაძლევს ერთი პროგრამული კოდის საშუალებით, შევქმნათ აპლიკაცია ყოველი პლატფორმისთვის და შევქმნათ მომხმარებელზე ორიენტირებული უსაფრთხოების მოდელი. ამასთან ერთად, ვები არ კონტროლდება რომელიმე კომპანიის მიერ და გვაძლევს აბსოლუტურ თავისუფლებას“ [76].

ამასთან ერთად, უნდა აღვნიშნოთ, რომ პროგრესული ვებ აპლიკაციები მუშაობს ისევე, როგორც ლოკალური პროგრამები იმ დამატებითი პრიორიტეტით, რომ მომხმარებელს არ უწევს მისი საკუთარი ხელით განახლება, ან მისი შესყიდვა ყოველი ახალი ვერსიის გამოსვლისას.

აღნიშნული აპლიკაციები თავისუფალია ლოკალური ოპერაციული სისტემების პრობლემისაგან და, აქედან გამომდინარე, ბევრად ნაკლებ ხარვეზს შეიცავს.

თავი 3

მანქანური დასწავლა: კვლევა და შედეგები

3.1. მონაცემთა მოძიება	92
3.2. ნაშრომთა თემის მოდელირება	93
3.2.1. მონაცემთა წინასწარი დამუშავება	94
3.2.2. LDA და NMF მეთოდების გაშვება და მოდელების მიღება	96
3.3. სამეცნიერო სტატიების რეზიუმეთა კლასიფიკაცია	104
3.3.1. მონაცემთა წინასწარი დამუშავება	105
3.3.2. კლასიფიკაციის მოდელის მიღება	107

3.1 მონაცემთა მოძიება

მონოგრაფიის სამეცნიერო ღირებულებისა და სიახლის უმნიშვნელოვანიესი ნაწილი მანქანური დასწავლის მეთოდებისა და ალგორითმების გამოყენებით *ინოვაციური მოდელების* შექმნაა.

ჩვენ საპროექტო ნაშრომში, მანქანური დასწავლის რომელი მეთოდების და ალგორითმების გამოყენებას ვაპირებთ და რატომ, ამაზე შემდეგ პარაგრაფში დეტალურად შევჩერდებით.

ამჯერად კი წარმოვადგენთ მანქანური დასწავლის მეთოდების გამოყენებას და მათ შედეგებს. ერთმანეთს შევადარებთ სხვადასხვა მეთოდებს და ავირჩევთ საუკეთესოს იმისთვის, რომ შევქმნათ *ინოვაციური მოდელები*, რაც საბოლოოდ ჩვენი ნაშრომის უმთავრესი სამეცნიერო სიახლე და პრაქტიკული ღირებულებაა.

იმისთვის, რომ ნებისმიერი მანქანური დასწავლის ალგორითმი გამოვიყენოთ, პირველ რიგში აუცილებელია გვქონდეს სასწავლო მონაცემთა ბაზა (Dataset). რაც უფრო დიდია მონაცემთა ბაზა, მით უფრო უკეთესია შედეგიც.

მონაცემთა შესასწავლად საჭირო ბაზები ინტერნეტის მეშვეობით იქნა მოძიებული. კერძოდ, მოკვლეულ იქნა Arxiv.org-ის მონაცემთა ბაზა, რომელიც ორ მილიონამდე სამეცნიერო ნაშრომს მოიცავს [77, 78].

იგი შედგება მრავალი კატეგორიის ტექნიკური საგნებისაგან, ისეთებისგან როგორიცაა, მათემატიკა, ფიზიკა, კომპიუტერული მეცნიერებები, სტატისტიკა, ფინანსები, კვანტური და გამოთვლითი ბიოლოგია და სხვ. ამასთანავე, დამატებით მოძიებულ იქნა Pubmed-ის მონაცემთა ბაზა [79], რომელიც 130000-მდე სამეცნიერო სტატიისგან შედგება. აღნიშნული სამეცნიერო სტატიები, ძირითადად ბიოსამედიცინო და სიცოცხლის მეცნიერებების კატეგორიებს მიეკუთვნება.

მიღებული მონაცემების დამუშავების პირველ ეტაპზე, ისინი ერთ ბაზაში იქნა გაერთიანებული, შემდეგ, ზედმეტი ინფორმაციისგან (ავტორები, გამოყენებული ლიტერატურა და სხვ.) გაწმენდილი და ორ ნაწილად დაყოფილი. ეს ორი ნაწილი კი მომავალი კვლევებისათვის უმნიშვნელოვანესი გახდა.

პირველი ნაწილი გახლავთ სამეცნიერო სტატიების რეზიუმეები. მათი შესწავლა გამოირჩევა იმით, რომ ის ბევრად უფრო მცირეა ვიდრე მთლიანი სტატიის ტექსტი და მასში კარგად არის გადმოცემული საკვლევი ობიექტის თემა.

მეორე ნაწილს რაც შეეხება, ეს სამეცნიერო სტატიების ძირითადი ტექსტია. მონაცემთა ნაკრების ზომიდან გამომდინარე, საჭირო გახდა სრული ტექსტის მქონე ბაზების შემცირება. მაშინ როდესაც სწავლებები ჩავატარეთ დაახლოებით ორი მილიონი სამეცნიერო ნაშრომის რეზიუმეზე, სრული ტექსტისთვის გაშვებული ალგორითმები სწავლობდა დაახლოებით 350000 ნაშრომის სრულ ტექსტს.

3.2. ნაშრომთა თემის მოდელირება

როგორც უკვე აღვნიშნეთ, მანქანური დასწავლის თემის მოდელის ალგორითმების საშუალებით შევქმენით ინოვაციური მოდელი, რომლის მეშვეობითაც მოვახდინეთ სამეცნიერო ნაშრომთა ხელახალი კატეგორიზაცია, მასში გამოყენებული თემების მიხედვით. აღნიშნული მოდელის ერთ-ერთი ყველაზე მნიშვნელოვანი პრაქტიკული ღირებულება მისი საგამომცემლო სფეროში გამოყენებაა. როდესაც სისტემაში დაფიქსირდება მომხმარებლის ინტერესი რომელიმე სამეცნიერო სტატიის (ან წიგნის) მიმართ, სისტემაში დამახსოვრებული იქნება ამ კონკრეტული ნაშრომის თემის მოდელირებით გენერირებული რეზიუმესა და ტექსტის მოდელების შესაბამისი იდენტიფიკატორები. მათი დახმარებით კი

შესაძლებელი გახდება მომხმარებელს მისთვის საინტერესო ნაშრომები შევთავაზოთ. ამ ყველაფრით კი მნიშვნელოვნად გავაუმჯობესებთ სამეცნიერო საგამომცემლო მარკეტინგს.

როგორც წინა პარაგრაფებიდანაა ცნობილი, ნაშრომში გამოყენებულია თემის მოდელირების ორი საუკეთესო მეთოდი [30,31]. დირიხლეს ლატენტური განაწილების (LDA) და არაუარყოფითი მატრიცის ფაქტორიზაციის (NNMF) ალგორითმების მეშვეობით, შესწავლილი იქნა, როგორც სამეცნიერო ნაშრომთა რეზიუმეები, ასევე მათი სრული ტექსტები. საბოლოო შედეგები ერთმანეთს შევადარეთ და გამოვლენილი იქნა საუკეთესო მოდელები, ნაშრომის ორივე ნაწილისათვის. საბოლოოდ, შერჩეული ორი მოდელის სინქრონული მუშაობით კი მიღებული იქნა ამ სამეცნიერო სიახლის ინოვაციური და პრაქტიკული ღირებულება.

3.2.1 მონაცემთა წინასწარი დამუშავება

სანამ თემის მოდელირების ალგორითმი იქნება გაშვებული, საჭიროა მონაცემთა წინასწარი დამუშავება (Pre-processing). გარკვეული შეზღუდვების გამო, LDA და NMF, მონაცემთა წინასწარი დამუშავება განსხვავებული მეთოდებით ჩატარდა. იმის გამო, რომ LDA დამოკიდებულია სიტყვათა რაოდენობის ალბათობათა განაწილებაზე, არაა შესაძლებელი TF-IDF-ის გამოყენება, განსხვავებით NMF-ისგან, რომელიც კოფიციენტების საშუალებით მუშაობს.

შესაბამისად, LDA-ში გამოყენებულ იქნა დათვლის ვექტორიზატორის მეთოდი (CountVectorizer) [80]. ამ მეთოდის გამოყენებით, მთლიანი დოკუმენტთა სიმრავლისთვის, თითოეულ უნიკალურ სიტყვაზე დათვლილ იქნა თუ რამდენჯერაა ის გამოყენებული თითოეულ დოკუმენტში. შედეგად კი მიღებული იქნა დოკუმენტის ტერმინის მატრიცა (Document term matrix – DTM) (ნახ. 3.1).

	Formula	Star	Molecule	Learning	Spin	Temperature
Doc1	1	0	1	1	0	0
Doc2	0	1	0	0	1	0
Doc3	0	0	0	1	1	1

ნახ. 3.1. დოკუმენტის ტერმინთა მატრიცა

შესაბამისად, მონაცემთა წინასწარი დამუშავების ბოლო ეტაპად LDA-სთვის „CountVectorizer“-ის, ხოლო NNMF-ისთვის TF-IDF-ის (TF-IDF-ზე დეტალური ინფორმაციისთვის იხილეთ პარაგრაფი 2.5) გამოყენება მოხდა (ნახ. 3.2, 3.3).

ნახ. 3.2. NNMF მეთოდში მონაცემთა წინასწარი დამუშავებისთვის TF-IDF-ის გამოყენება

ნახ. 3.3. LDA მეთოდში მონაცემთა წინასწარი დამუშავებისთვის CountVectorizer-ის გამოყენება

როგორც ამ ნახაზებიდან ჩანს, საბოლოო “fit_transform” მეთოდის გაშვებამდე, რომელიც გარდაქმნის დოკუმენტებს სიტყვათა ვექტორებად, ორივე ტიპის ვექტორიზერის გენერირების დროს გამოყენებულია მასში შემავალი სამი პარამეტრი. ესენია:

1) "max_df" – პარამეტრის გამოყენებით, მონაცემთა სიმრავლეებში ფილტრავს (აიგნორირებს) ისეთ სიტყვებს, რომლებიც მითითებული რიცხვის პროცენტულ რაოდენობაშია გამოყენებული. ჩვენი კვლევის შემთხვევაში, ეს რიცხვი 0.5-ია. შესაბამისად, ყველა ის სიტყვა, რომელიც სამეცნიერო ნაშრომთა 50%-შია გამოყენებული, თემის მოდელირების დროს იგნორირებული იქნება;

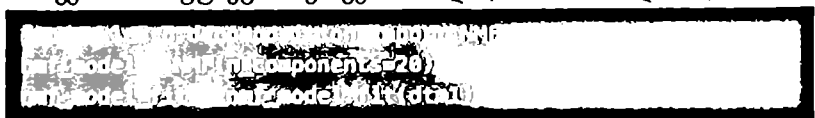
2) "min_df" – წინა პარამეტრის საპირისპიროდ, მოცემული პარამეტრის მნიშვნელობით ფილტრავს სიტყვებს, რომლებიც მინიმუმ 2 სტატიაშია გამოყენებული;

3) "stop_words" – მოცემული პარამეტრი, მეთოდისთვის პარაგრაფ 2.2-ში აღწერილი, ამოსაშლელ სიტყვათა სიის გადასაცემად არის გამოყენებული.

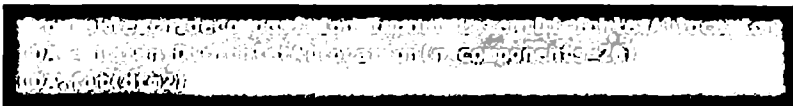
აღსანიშნავია, რომ ზემოაღწერილი პარამეტრები, თემის მოდელირების მეთოდების არაერთგზის გაშვების შემდეგ იქნა ჩამოყალიბებული, ეს საკითხი დეტალურად იქნება განხილული შემდეგ პარაგრაფში.

3.2.2. LDA და NMF მეთოდების გაშვება და მოდელების მიღება

მას შემდეგ, რაც ტექსტის წინასწარი მომზადება დასრულდა და უკვე მზად იყო დოკუმენტში სიტყვათა სიმრავლის მატრიცები, მოხდა თემის მოდელირების ალგორითმების გაშვება. ამისთვის, ჯერ სასურველი NMF და LDA კლასები გამოცხადდა და როგორც მანქანური დასწავლის უმეტესი მოდელების შემთხვევაში, მორგების "fit" ფუნქციის გაშვება მოხდა (ნახაზი 3.4. და 3.5).



ნახ. 3.4. NMF მორგების მეთოდების გაშვება



ნახ. 3.5. LDA მორგების მეთოდების გაშვება

როგორც ნახაზებიდან ჩანს, თემის მოდელირების მეთოდის კლასის გამოცხადებისას, ორივე შემთხვევაში გამოყენებულია შემავალი პარამეტრი „n_components“, რაც თემის მოდელირების მიერ თემების რიცხვს შეადგენს.

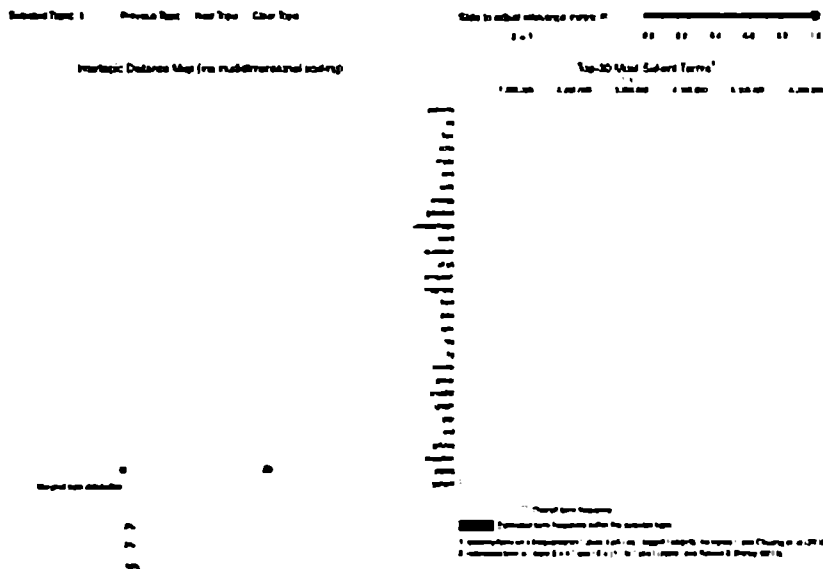
როგორც უკვე აღვნიშნეთ, პოტენციური თემების ან კატეგორიების რაოდენობა ალგორითმის გაშვებამდე ჩვენ მიერ უნდა იყოს განსაზღვრული. ამ შემთხვევაში, ჩვენ მიერ გაშვებული მეთოდი მილიონობით სამეცნიერო ნაშრომს დაყოფს ოც კატეგორიად და თითოეულს ამ კატეგორიებში მოათავსებს.

როდესაც მოდელირების ალგორითმების გაშვების პირველი შედეგები იქნა მიღებული, მოლოდინი დიდი იყო, თუმცა საბოლოოდ, შედეგი არასასურველი გამოდგა. განსაკუთრებით აღსანიშნავია მდგომარეობა სრული ტექსტის თემის მოდელირების შესახებ. აღმოჩნდა, რომ „max_df“, „min_df“, „stop_words“ და „n_components“-ის შემავალი პარამეტრები არასაკმარისი იყო. ტექსტი მაინც შეიცავდა მრავალ არასასურველ სიტყვას და საბოლოო შედეგიც ძალიან არასასურველი გამოდიოდა (ნახ. 3.6 და 3.7). ამას მოჰყვა მეორე, მესამე გაშვება, მაგრამ სასურველი შედეგის დადგომას მეტი მცდელობა დასჭირდა.

LDA კლასტერირება (ლატენტური დირიხლეს განაწილება) მანქანური დასწავლის უკონტროლო ალგორითმია. კლასტერიზაციის ალგორითმის (K-means) მსგავსად, LDA აჯგუფებს სიტყვებს და დოკუმენტებს კლასტერების წინასწარ განსაზღვრულ რაოდენობაში (მაგალითად. თემებში), შემდგომი გამოყენებისთვის.



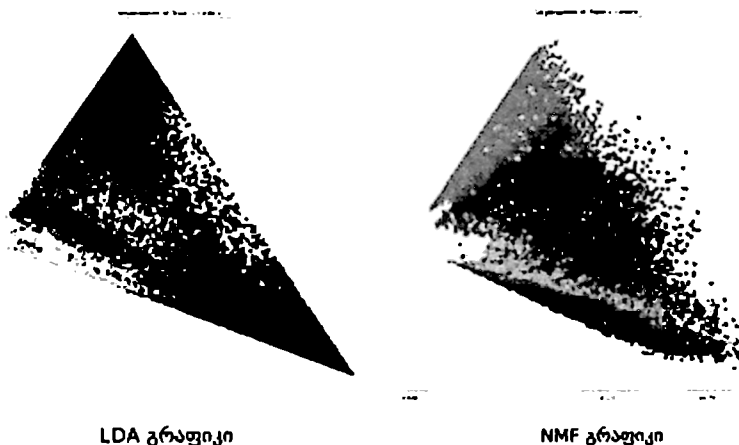
ნახ. 3.6. LDA კლასტერული სეგრეგაციის გრაფიკი რეზიუმეებისთვის მეორე გაშვებაზე (ა) და სრული ტექსტებისთვის მესამე გაშვებაზე (ბ)



ნახ. 3.7. LDA კლასტერები არასწორი მონაცემებით (pyLDAvis გრაფიკული დიაგრამა)

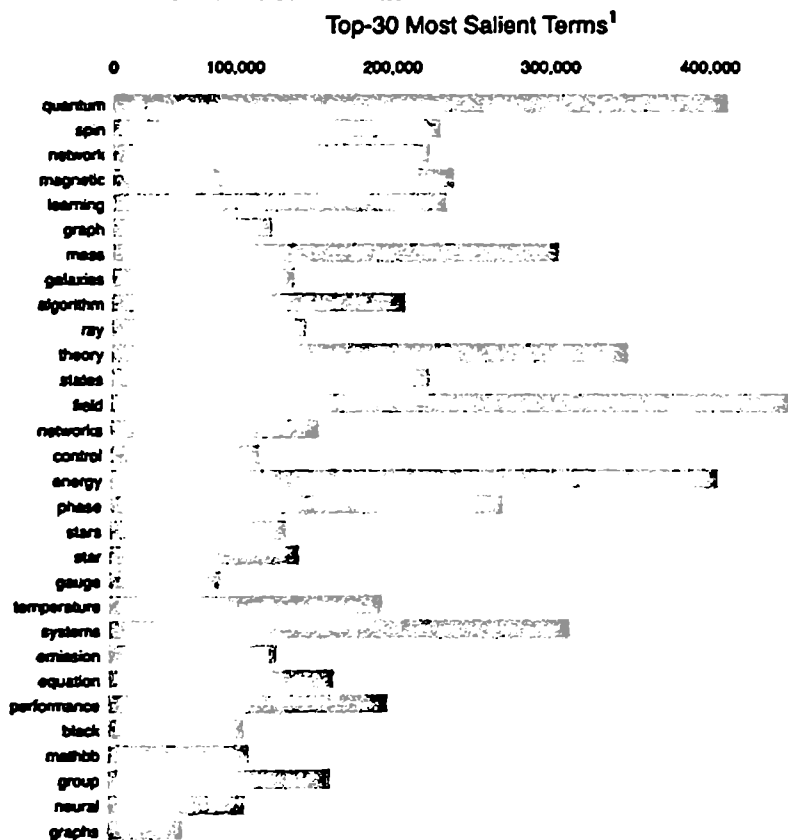
როგორც 3.6 ნახაზზე ჩანს, სასურველი შედეგის მიღწევამდე კლასტერული სეგრეგაციის გრაფიკი არ იძლევა კარგ შედეგს. აგრეთვე, ნახაზი 3.7-ის მიხედვით, ჩანს მრავალი არასასურველი სიტყვა. განსაკუთრებით: „xcite“, „fig“, „eq“, „xmath0“ და სხვ.

რამდენიმე ეტაპიანმა მონაცემთა დამუშავებამ, განსაკუთრებით „stop_words“ პარამეტრის ოპტიმიზაციამ, „max_df“ და „n_components“ გაუმჯობესებასთან ერთად, მრავალი საათისა და ასეულობით გეგაბაიტი მეხსიერების დახარჯვის შემდეგ, სასურველ შედეგები გამოიღო. საბოლოოდ თემის მოდელირების გაშვებულმა ალგორითმებმა უკეთესი შედეგი დააბრუნეს ვიდრე მათმა წინამორბედმა მცდელობებმა. შედეგად გაუმჯობესდა კლასტერული სეგრეგაციის გრაფიკიც [81] (ნახ. 3.8) და სხვადასხვა თემებში ყველაზე ხშირად გამოყენებული სიტყვების მნიშვნელობებიც მეტად საინტერესო გახდა.



ნახ. 3.8. კლასტერული სეგრეგაციის გრაფიკი სწორი მონაცემებით

ნახაზზე ჩანს კლასტერული სეგრეგაცია იგი ახალ კატეგორიებში ბევრად უკეთ გამოიყურება. განსაკუთრებით, აღსანიშნავია NNMF მეთოდის შედეგად მიღებული სეგრეგაცია. წარმოვადგინოთ 30 გამორჩეული სიტყვა განაწილების მოდელიდან (ნახ. 3.9).



ნახ. 3.9. LDA მეთოდით მიღებული 30 ყველაზე მნიშვნელოვანი სიტყვა

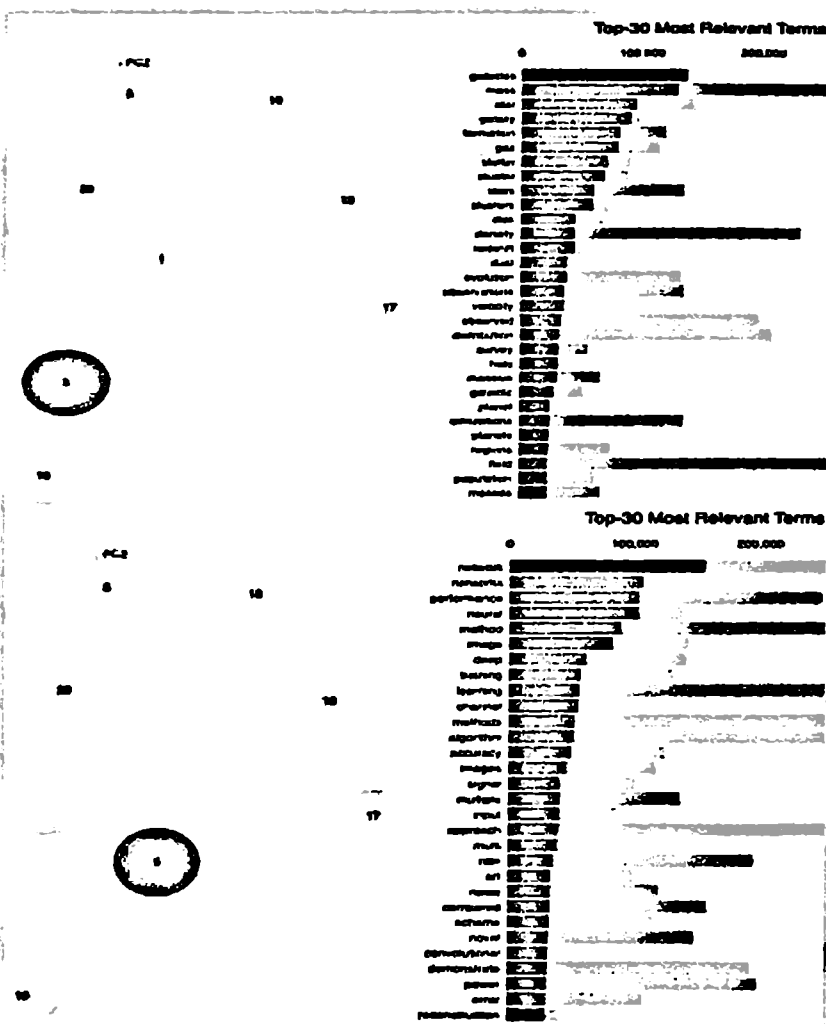
როგორც ნახაზზე ჩანს, გამორჩეულ სიტყვათა სიაში მოხვდა სიტყვები: „კვანტური“, „სპინი“, „მასა“, „გალაქტიკა“, „ვარსკვლავი“, „სისტემები“, „ენერგია“ და სხვ. მიღებული შედეგიდან ვხედავთ, რომ ამ სიტყვათა უმეტესობა გამოსადეგია დოკუმენტის სავარაუდო კატეგორიის წარმოსადგენად.

ვნახოთ რამდენიმე მაგალითი თუ როგორ გამოიყურება კონკრეტულ თემაში არსებული სიტყვების მნიშვნელობები და რამდენადაა იგი საკმარისი დოკუმენტის კატეგორიის გამოვ-ნობისათვის (ნახ. 3.10).

როგორც ამ ნახაზზე ჩანს, 30 ყველაზე მნიშვნელოვანი სიტყვა განაწილებიდან, უკვე მეტყველებს სხვადასხვა სამეცნიერო ნაშრომის კატეგორიაზე.

გვხვდება ისეთი სიტყვები, როგორიცაა *კვანტური* (quantum), *ენერგია* (Energy), *მასა* (mass) და სხვ. ამასთან ერთად, აქ გამოსახულია მნიშვნელოვან სიტყვათა განაწილება ორი კონკრეტული კლასტერისათვის. პირველ მათგანში ყველაზე ხშირად გამოყენებადი სიტყვებია: *გალაქტიკები* (galaxies), *მასა* (mass), *ვარსკვლავი* (star), *გალაქტიკა* (galaxy), *ფორმირება* (formation), *გაზი* (gas) და სხვ. აღნიშნული სიტყვები, მეტყველებს იმაზე, რომ საქმე გვაქვს ასტრონომიის მიმართულებასთან.

მეორეს მხრივ, კი შემდეგ თემაში ვხვდებით სიტყვებს: *ქსელი* (network), *ნეირონული* (neural), *პერფორმანსი* (performance), *ღრმა* (deep), *ტრენინგი* (training), *სწავლება* (learning) და სხვ. სიტყვებიდან გამომდინარე აშკარაა, რომ კლასტერში ნაგულისხმევი თემა ყველაზე ახლოს მანქანურ დასწავლასთანაა.



ნახ. 3.10. ყველაზე საგულისხმო სიტყვები მეორე და მესამე კლასტერებისთვის

იმის შესაბამისად, რომ მონაცემები დაყოფილია ორ ნაწილად (ნაშრომის რეზიუმე და სრული ტექსტი) და ვიყენებთ დამუშავების ორ განსხვავებულ მეთოდს, ალგორითმების გაშვების შემდეგ, მიღებულია ოთხი მოდელი. ეს მოდელებია:

- 1) LDA - რეზიუმეებისთვის;
- 2) NNMF - რეზიუმეებისთვის;
- 3) LDA - სრული ტექსტისთვის;
- 4) NNMF - სრული ტექსტისთვის.

უნდა აღინიშნოს, რომ კვლევის დრო NNMF მოდელის მიღება, 5-10-ჯერ უფრო სწრაფად ხდებოდა ვიდრე LDA მოდელის, თუმცა LDA-ს მონაცემები, არანაკლებად საინტერესოა. საჭირო იყო გამოვლენილი ყოფილიყო საუკეთესო მოდელები რეზიუმესა და სრული ტექსტისათვის. ამისთვის, მოდელიდან ამოღებულ იქნა ოთხივე მოდელში მიღებული, 20-20 თემის, 50-50 ყველაზე მნიშვნელოვანი სიტყვა (ჯამში 80 თემა და 4000 სიტყვა).

ამოღებული მონაცემები, თითოეული მეთოდისათვის სათითაოდ იქნა განხილული და შედეგად მიღებულ იქნა ინფორმაცია კონკრეტული კლასტერის სასარგებლოდ გამოყენების მხრივ პროცენტულად რომელ მეთოდს აქვს უკეთესი შედეგი. შედეგები იხილეთ 3.1 ცხრილის სახით.

თემის მოდელირების შედეგები, პოტენციური

გამოყენების მხრივ

ცხრ.3.1

Abstract NMF	Abstract LDA
≈0,9	≈0,95
Full text NMF	Full text LDA
≈0,6	≈0,67

როგორც ცხრილში ჩანს, LDA მეთოდის გამოყენებამ ორივე, რეზიუმეებისა და სრული ტექსტების გამოყენების დროს, მცირედით აჯობა NNMF მეთოდს. მიღებული თემების 50-50-ი ყველაზე რელევანტური სიტყვების გაანალიზების დროს აღმოჩნდა, რომ LDA მეთოდის მიერ კლასტერიზებული სიტყვები, NNMF-თან შედარებით მცირედით უფრო გამოსადეგარია და მეტად მიუთითებს რომელიმე კატეგორიასა თუ კვლევის მეთოდზე.

3.3. სამეცნიერო სტატიების რეზიუმეთა კლასიფიკაცია

მანქანური დასწავლის შემდეგი მეთოდი, რომელიც საპროექტო კვლევის პროცესშია გამოყენებული, ტექსტის კლასიფიკაციაა. აღნიშნული მეთოდის გამოყენებით მიღებულ იქნა ნაშრომისთვის, ძალზე მნიშვნელოვანი, კიდევ ერთი ინოვაციური სამეცნიერო სიახლე. ტექსტის კლასიფიკაციის მეთოდების მუშაობის დეტალებისთვის შეიძლება ვიხილოთ პარაგრაფი 2.6. ახლა კი მის პრაქტიკულ გამოყენებას შევხებით.

კლასიფიკაციის მეთოდის გამოსაყენებლად არსებულ სამეცნიერო ნაშრომთა მონაცემთა ბაზის მხოლოდ ერთი ნაწილი – რეზიუმეები იქნა გაანალიზებული. მიზეზი მისი პრაქტიკულობაა.

როგორც უკვე აღვნიშნეთ, კვლევის მიზანია, რომ როდესაც მომხმარებელი ნაშრომზე მუშაობას დაიწყებს, კლასიფიკაციის მოდელის მეშვეობით, სისტემა რამდენიმე წინადადების (ან რეზიუმეს) შეყვანის შემდეგ, მომენტალურად მიხვდეს იმას, თუ რა თემაზეა საუბარი და მომხმარებელს ამ თემისთვის საჭირო ფუნქციონალსა და ინტერფეისს შესთავაზებს. კვლევა, თუ რა ფუნქციონალია საჭირო ამა თუ იმ თემისათვის, განხილულია მეტავეში.

3.3.1 მონაცემთა წინასწარი დამუშავება

ისევე როგორც თემის მოდელირების შემთხვევაში, კლასიფიკაციის ალგორითმების გაშვებამდე, საჭიროა არსებულ მონაცემთა წინასწარი დამუშავება, თუმცა მისგან განსხვავებით, ტექსტის კლასიფიკაცია კონტროლირებადი მანქანური დასწავლის მეთოდია.

შესაბამისად, მას ორი შემავალი პარამეტრი აქვს. ეს პარამეტრებია თავად შესასწავლი ტექსტი და მისი კატეგორია. ჩვენ შემთხვევაში, დასახული მიზნებიდან გამომდინარე გამოყენებული იქნა არა ჩვენ მიერ, არამედ მონაცემთა მწარმოებლის მიერ მინიჭებული კატეგორიები.

მრავალი კატეგორიის მქონე სტატიების შემთხვევაში არჩეულ იქნა პირველი კატეგორია. ამ მიზეზით არსებული ხარვეზები, გადაჭრილ იქნა მონაცემთა ბაზის სიდიდის მეშვეობით. ორ მილიონამდე სტატიის არსებობამ, შეძლო არასწორად განაწილებული კატეგორიების დაბალანსება და საბოლოო შედეგი საკმაოდ დამაკმაყოფილებელი გამოდგა.

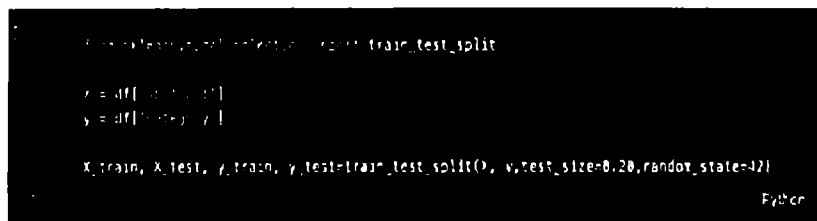
შესწავლილი კატეგორიების ჯამური რაოდენობა იყო 178, რაც ძალზე დიდი რიცხვია და ნიშნავს იმას, რომ სტატიები ზედმეტად დაკონკრეტებული კატეგორიებით უნდა ყოფილიყო შესწავლილი.

კატეგორიათა დიდი რაოდენობა, კლასიფიკაციის მეთოდის გამოყენებისას. მისი წარმატებულად მუშაობის ეჭვს ბადებდა. თუმცა უნდა აღინიშნოს ისიც, რომ ამ კატეგორიებს ჰქონდა მშობელი კატეგორია.

მაგალითად, შესწავლილ იქნა ისეთი სამეცნიერო ნაშრომთა კატეგორიები, როგორებიცაა: „კომპიუტერული მეცნიერება და თამაშთა თეორია“, „ადამიანისა და კომპიუტერის ურთიერთქმედება“, „ინფორმაციის თეორია“, „ლოგიკა კომპიუტერულ

მეცნიერებებში" და მრავალი სხვა ქვეკატეგორია, რომელიც საბოლოოდ კომპიუტერულ მეცნიერებებს აერთიანებს. აქედან გამომდინარე, იმ დოკუმენტთა უმრავლესობა, რომლის გამოცნობის დროსაც კლასიფიკაციის მოდელი შეიცდება, მაინც იქნება ახლოს ძირითად სფეროსთან.

მას შემდეგ, რაც მონაცემთა ბაზამ პირველად გაიარა ელემენტარული დამუშავება, ის გასუფთავდა ზედმეტი მონაცემებისგან და მოხდა მისი მაქსიმალური შემცირება. ამის შემდეგ დაიწყო მისი მომზადება კლასიფიკაციის ალგორითმების გასაშვებად. ამისთვის, პირველ ეტაპზე მონაცემები ორ ნაწილად იქნა დაყოფილი (ნახ. 3.11).



ნახ. 3.11. კლასიფიკაციის შესასწავლ მონაცემთა ორ ნაწილად დაყოფა

როგორც ნახაზზე ჩანს, წინასწარ მომზადებულია მანქანური დასწავლისათვის საჭირო ორი, რეზიუმეებისა და მათი კატეგორიების x და y მონაცემთა სიმრავლე. გავიხსენოთ 2.6 პარაგრაფში ახსნილი ინფორმაცია იმის შესახებ, რომ ტექსტის კლასიფიკაციის მოდელების გასაშვებად მონაცემები ორ ნაწილად უნდა დაიყოს. პირველი ნაწილი მოდელის შესასწავლად გამოიყენება, მეორე კი მისი ტესტირებისთვის. „sklearn“ ბიბლიოთეკის model_selection მოდულიდან იმპორტირებული train_test_split მეთოდის გამოყენებით, კი ამ მეთოდებისგან მიიღებულია ოთხი X_{train} , X_{test} , y_{train} , y_{test} სიმრავლე.

X_{train} და y_{train} სიმრავლეები, შესაბამისად, მონაცემთა შესასწავლად საჭირო რეზიუმეებსა და კატეგორიებს მოიცავს, ხოლო X_{test} და y_{test} – მათი ტესტირებისთვის საჭირო რეზიუმეებსა და კატეგორიებს.

მე-15 ნახაზზე ჩანს რომ `train_test_split` მეთოდს კიდეც ორი შემავალი პარამეტრი აქვს. `test_size` შედეგების დასატესტი სრულ მონაცემთა ნაწილის პროცენტს გულისხმობს. ამ პარამეტრის მნიშვნელობა 0.20-ია, რაც იმას ნიშნავს, რომ ჩვენ შემთხვევაში 1948654 სტატიის რეზიუმეს დამუშავებისას, შესასწავლი მონაცემების რაოდენობა 1558923-ია, ხოლო სატესტო მონაცემებისა 389731.

კიდეც ერთი პარამეტრი, რომელსაც მეთოდის გამოყენებისას ვხედავთ, `random_state`-ია. აღნიშნული პარამეტრი გამოყენებულია იმისთვის, რომ პარამეტრების ოპტიმიზაციის პროცესში, რამდენიმე განსხვავებული მანქანური დასწავლის მეთოდების გაშვების დროს სასწავლო და სატესტი მონაცემების გადანაწილება განსხვავებულად არ მომხდარიყო, რასაც შემდეგ, შესაძლოა რეზულტატის შედარებების დროს შეცდომაში შეყვანა გამოეწვია.

3.3.2 კლასიფიკაციის მოდელის მიღება

სამეცნიერო ნაშრომთა კლასიფიკაციის ოპტიმალური მეთოდის მისაღებად, დღეისათვის არსებული სწავლების ორი საუკეთესო მეთოდი იქნა გამოყენებული. ისევე როგორც თემის მოდელირებისას, კლასიფიკაციის შემთხვევაშიც, დოკუმენტთა ვექტორიზაცია მონაცემთა წინასწარი დამუშავების ერთ-ერთი მნიშვნელოვანი კომპონენტია. ზემოაღნიშნული თითოეული მეთოდისთვის ვექტორიზაციის სხვადასხვა მეთოდი იქნა გამოყენებული. საბოლოოდ, მოდელების მიღების შემდეგ კი „predict“ მეთოდის გამოყენებით, მოდელში წინასწარ გადაღებული სატესტო მონაცემები იქნა გაშვებული. მიღებული ინფორმაცია კი განსაზღვრავს

მოდელის სიზუსტის მაჩვენებელს, რაც მის საბოლოო ხარისხზე მიუთითებს. ორივე მეთოდის გამოყენების შემდეგ ჩვენ მიერ მიღებული სიზუსტის მაჩვენებელი ქულები, ერთმანეთს შედარადა და შედეგად საბოლოო „გამარჯვებული“ გამოვლინდა.

იმ მიზეზით, რომ TF-IDF ვექტორიზერი, ავტომატურად ვექტორიზაციის დროს სიტყვას 0-დან 1-მდე ანიჭებს მნიშვნელობას, ამ სიტყვის გამოყენებადობის გათვალისწინებით ყველა დოკუმენტში, მეთოდები გაშვებულ იქნა მონაცემთა წინასწარი დამუშავების სამი სხვადასხვა გზის მიხედვით. ესენია:

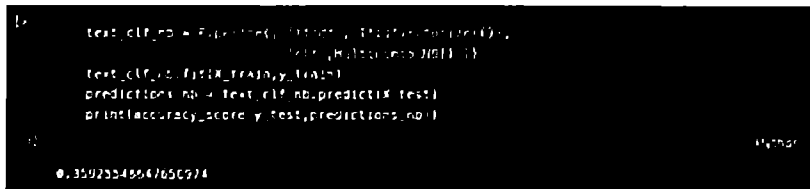
1) მონაცემთა ვექტორიზაცია TF-IDF მეთოდის გამოყენებით დამატებითი პარამეტრების გარეშე;

2) მონაცემთა ვექტორიზაცია TF-IDF მეთოდის გამოყენებით დამატებითი პარამეტრებით – $\text{max_df}=0.5$, $\text{min_df}=2$;

3) მონაცემთა ვექტორიზაცია CountVectorizer მეთოდის გამოყენებით დამატებითი პარამეტრებით – $\text{max_df}=0.5$, $\text{min_df}=2$.

პირველი მეთოდი, რომელიც სტატიების რეზიუმეთა კლასიფიკაციისათვის იქნა გაშვებული, ბაიესის მიაბიტი (Naive Bayes) ალგორითმია [30]. პირველ ეტაპზე ვექტორიზერად TF-IDF მეთოდი იქნა გამოყენებული, მასში დამატებითი ფილტრების არსებობის გარეშე. შემდეგ გაშვებულ იქნა სწავლების „fit“ მეთოდი და მიღებულ მოდელზე „prediction“ მეთოდის გამოყენებით, სატესტო მონაცემები დამუშავდა.

ამ ეტაპზე შედეგი არადამაკმაყოფილებელი აღმოჩნდა. მონაცემთა ტესტირების შემდეგ, სიზუსტის მაჩვენებელი ქულა დაახლოებით 0.36-ის ტოლი გამოვიდა, რაც იმას ნიშნავდა, რომ დასატესტი 400000-მდე სამეცნიერო სტატიის გამოცნობის მაჩვენებელი მხოლოდ 36% გახლდათ. მეთოდის გაშვების კოდის ფრაგმენტი მოცემულია 3.12 ნახაზზე.



ნახ. 3.12. ბაიესის მიაშიტი ალგორითმით პირველი კლასიფიკაციის მეთოდის გაშვება

ამის შემდეგ, ჩვენს მიერ მოხდა TF-IDF ვექტორიზერის გამოყენება შემავალ პარამეტრებთან ერთად და საბოლოოდ სიზუსტის მაჩვენებელი ქულა 0.41 გამოვიდა, რაც წინასთან შედარებით გაუმჯობესებული ვარიანტია, თუმცა მაინც არ არის დამაკმაყოფილებელი. ამის შემდეგ უკვე ოპტიმიზებული შემავალი პარამეტრებით, იმავე კლასიფიკაციის მეთოდის გაშვებისას კიდევ ერთხელ იქნა გამოყენებული, თუმცა, ამჯერად CountVecorizer ვექტორიზერით.

შედეგი ბევრად უკეთესი გამოვიდა ვიდრე TF-IDF-ის შემთხვევაში. სიზუსტის მაჩვენებელი ქულა ამჯერად 0.54 იყო, რაც ძალზე კარგი მაჩვენებელია 200-მდე კატეგორიის შემთხვევაში, რომელთაგან ბევრი ერთმანეთთან კლასტერულად ძალიან ახლოსაა. იმის დადგენა, რამდენად კარგია ამა თუ იმ სიზუსტის მაჩვენებელი ტექსტის კლასიფიკაციაში, ყოველი მონაცემთა სიმრავლისათვის განსხვავებული მიდგომით ხორციელდება.

მაგალითად, ონკოლოგიაში, სასიცოცხლოდ მნიშვნელოვანია მოდელს უმაღლესი ხარისხის მაჩვენებელი ჰქონდეს, თუმცა მანქანური დასწავლის უმეტესობაში, მაშინ როდესაც შესასწავლი მხოლოდ ორი კატეგორია ან კლასია, კარგ შედეგად ითვლება 70-80%-იანი მაჩვენებლის მიღება.

რაც შეეხება მეორე მეთოდს, ეს წრფივი მხარდაჭერის ვექტორული კლასიფიკატორია – SVC (Support Vector Classifier) [81].

ზემოაღნიშნული ვექტორიზაციის სამივე მეთოდის გამოყენების შედეგები შემდეგია:

პარამეტრებიანი CountVectorizer-ის გამოყენების შემდეგ მიღებული მოდელის სიზუსტის მაჩვენებელი 0.51-ია, რაც უკვე კარგი მაჩვენებელია, თუმცა ჩამოუვარდება ბაიესის მიაშიტი ალგორითმის მაჩვენებელს. შემდეგ, აღნიშნული მეთოდი, პარამეტრებიანი TF-IDF ვექტორიზერის მიერ გენერირებული მატრიცით დამუშავდა. ამ შემთხვევაში კი შედეგად სიზუსტის მაჩვენებლის ქულა ამ ეტაპისთვის საუკეთესო 0.591 აღმოჩნდა, რაც მაღალი მაჩვენებელია 200-მდე ერთმანეთთან დაახლოებული კატეგორიების შემთხვევაში. საბოლოოდ კი გამოვიყენეთ TF-IDF მეთოდი შემავალი პარამეტრების გარეშე. შედეგად კი ამ შემთხვევაში საუკეთესო 0.593 სიზუსტის მაჩვენებელი მივიღეთ. მეთოდების შედარების შედეგები ასახულია მე-3 ცხრილში.

ტექსტის კლასიფიკაციის მეთოდების შედარება
სხვადასხვა ვექტორიზაციის გამოყენებით

ცხრ.3.2

მეთოდი	ბაიესის მიაშიტი ალგორითმი	SVC
CountVectorizer პარამეტრებით	0.54	0.51
TF-IDF	0.36	0.593
TF-IDF პარამეტრებით	0.41	0.591

როგორც ცხრილში ჩანს, ბაიესის მიაშიტი ალგორითმი უკეთეს შედეგს აჩვენებს CountVectorizer-ის გამოყენების დროს, მაგრამ არადამაკმაყოფილებელია TF-IDF-ის შემთხვევაში. ხოლო SVC ორივე TF-IDF მეთოდის გამოყენების დროს დამაკმაყოფილებელ შედეგს გვაძლევს. შედარების შემდეგ საბოლოოდ გამოვლინდა გამარჯვებულიც (მწვანე ფერით). ეს SVC მეთოდითა და პარამეტრების გარეშე გაშვებული TF-IDF ვექტორიზერით მივიღეთ ინოვაციური მოდელი, რომელიც სამეცნიერო ნაშრომთა 200-მდე კატეგორიაში, 60%-იანი სიზუსტის მაჩვენებელს აფიქსირებს.

თავი 4

ტექსტური ინფორმაციის შექმნისათვის საჭირო ფუნქციონალის და ინტერფეისის კვლევა

4.1 გამოკითხვა და მისი ანალიზის შედეგები	112
4.1.1 ფუნქციონალი	112
4.1.2 ბიზნესპროცესების სრულყოფა	129
4.2 სხვადასხვა ტიპის სამეცნიერო ნაშრომებში არსებული ფუნქციონალის კვლევა და შედეგები	134

4.1 გამოკითხვა და მისი ანალიზის შედეგები

4.1.1 ფუნქციონალი

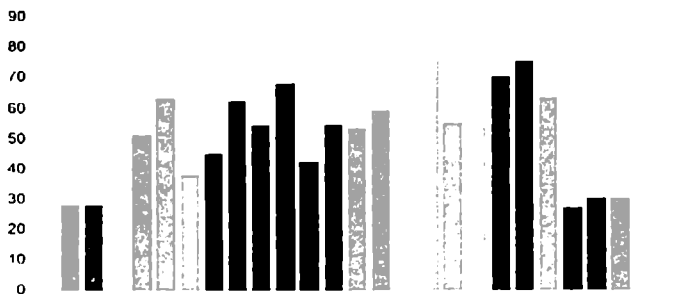
ჩვენი მონოგრაფიის კვლევის ერთ-ერთი მნიშვნელოვანი ნაწილია ადამიანური ფსიქოლოგიის, სტილის, იდეებისა და მათ მიერ დანახული პრობლემების გათვალისწინება, ამისათვის გამოკითხვების ჩატარება შესაბამისი სფეროს მცოდნე ექსპერტებთან – საპროექტო ნაშრომის მნიშვნელოვანი კვლევა.

ჩვენ მიერ ჩატარებული იქნა გამოკითხვა სხვადასხვა სამეცნიერო სფეროში მომუშავე ქართველ და უცხოელ მეცნიერებთან. გამოკითხვები ატარებდა როგორც კონტექსტუალურ ხასიათს, აგრეთვე ორიენტირებული იყო ახალი იდეების მოძიებასა და სხვა სტატისტიკური მონაცემების გაანალიზებაზე.

გამოკითხვა გახლდათ კონფიდენციალური (ერთადერთი პირადი შეკითხვა რესპოდენტთან მისი ასაკი იყო). ადამიანთა დამოკიდებულების გაზრდა ასაკის მატებასთან ერთად, რომელიმე კონკრეტულ ინსტრუმენტზე საინტერესოდ იქნა მიჩნეული და შესაბამისად გამოკითხული იქნა მაქსიმალურად ყველა ასაკის მქონე ადამიანი (ნახ. 4.1).

როგორც ნახაზიდან ჩანს, გამოკითხულთა (რესპოდენტების) ასაკი საკმაოდ მრავალფეროვანია. ის დაახლოებით 20 დან 80 წლამდე მერყეობს და საშუალო ასაკი 50 წელია. კვლევამ აჩვენა, რომ **ასაკის მატებასთან ერთად, ტექნოლოგიური პროდუქტების განვითარების იდეების გამოვლენა არა თუ არ მცირდება, შეიძლება ითქვას ტექნიკური სფეროს წარმომადგენელი მეცნიერების შემთხვევაში, ახალგაზრდა არატექნიკური სფეროს წარმომადგენლებთან შედარებით იზრდება** კიდევ.

გამოკითხულთა ასაკი



ნახ. 4.1. გამოკითხულთა ასაკის მაჩვენებელი

თუმცა, უნდა აღინიშნოს, რომ *ასაკის მატებასთან ერთად, არა ტექნიკურ სფეროებში, იზრდება მიჩვევადობა*, რომელიმე კონკრეტული ტექსტის რედაქტორის მიმართ. განსაკუთრებით უნდა აღინიშნოს მიჩვევადობა ყველაზე გამოცდილ და პოპულარულ *Microsoft Office რედაქტორზე*. ზოგადად, Microsoft Office-ზე მიჩვევადობის მაღალი მაჩვენებელი აღმოჩნდა არა მხოლოდ უფროსი ასაკის მეცნიერების მხრიდან, არამედ სხვა არატექნიკურ სფეროებთან მომუშავე ახალგაზრდა მეცნიერების მხრიდანაც.

მოცემული ინფორმაციის გამოყენებით მიღებულ იქნა გადაწყვეტილება იმასთან დაკავშირებით, თუ როგორი უნდა იყოს პროექტის ტექსტური რედაქტორის სამომხმარებლო ინტერფეისი, რომ არ მოხდეს მომხმარებლის მხრიდან მისი მიჩვევადობის მიხედვით, სისტემის ინტერფეისის რთული აღქმა.

მომხმარებელთა გამოკითხვების მთავარი მიზანი იყო ორი ძირითადი მიმართულების გაუმჯობესება. პირველია მუშაობის ბიზნესპროცესის გაუმჯობესება. ბიზნეს პროცესში ამ შემთხვევაში იგულისხმება ის, თუ ნაშრომის შექმნის დროს, წერის პროცესის

გარდა, რა სხვა პროცესები არსებობს, რომელთა ავტომატიზაციაც დღეისათვის არსებულ სისტემებში არ გვხვდება და რისი გაუმჯობესების შემთხვევაშიც, გვექნება საინტერესო პრაქტიკული ღირებულების ავტომატიზებული სისტემები.

მოგესენებათ, *სამეცნიერო ნაშრომების შექმნა, მხოლოდ წერის პროცესს არ მოიცავს*. ასეთი პროცესები, შეიძლება იყოს, ნაშრომის შექმნის დროს *თანაავტორებთან ურთიერთობა*, როგორც მათი განხილვით ავტომატიზება, აგრეთვე *განხილვების ისტორიის სტანდარტიზაცია*. მნიშვნელოვანი ფაქტორია სამეცნიერო ნაშრომ-
თა შენახვის სტანდარტიზაცია, მათი *გამოქვეყნების მარტივი გზები*. ჟურნალის რედაქციაში გაგზავნის შემდეგ, უარყოფითი პასუხის დადგომის შემთხვევაში (რეცენზირების ხარვეზების არსებობისას) გაკეთებული ცვლილებების ისტორიის შენახვა და სხვა მნიშვნელოვანი ბიზნეს პროცესები.

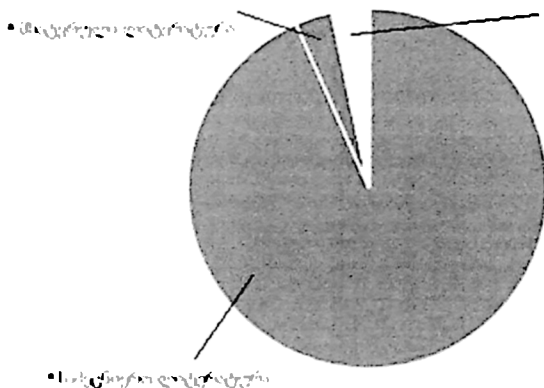
მეორე მნიშვნელოვანი გამოწვევა, რომლის გაუმჯობესების მიზნად დასახვის გამოც მოხდა აღნიშნული გამოკითხვის ჩატარება, *ტექსტურ ინფორმაციაზე მომუშავე რედაქტორის ფუნქციონალის ნაკლოვანებებისა და მისი არასრულყოფილების გამოვლენაა*.

მარტივად რომ ვთქვათ, გამოკითხვების მეშვეობით, გაანალიზებულ იქნა იმ ფუნქციონალის ნაკლებობა, რომელიც სამეცნიერო ნაშრომზე მუშაობის დროს, დღევანდელ ტექსტურ რედაქტორებს აქვს. სპეციალური შეკითხვების მეშვეობით, რესპოდენტები აფიქსირებდნენ მოსაზრებებს იმ ფუნქციონალის შესახებ, რომლის დანაკლისსაც განიცდიან ამ რედაქტორში ყოველდღიური მუშაობისას.

გამოკითხვაში მონაწილეობას იღებდნენ ორი ძირითადი სფეროს წარმომადგენლები, სამეცნიერო ნაშრომებთან მომუშავე ადამიანები და მთარგმნელები. გამოკითხულთა უმრავლესობა სამეცნიერო სფეროს წარმომადგენელი გახლდათ, რადგან ჩვენი

ნაშრომის ძირითადი სამეცნიერო ღირებულება და სიახლე, სწორედ სამეცნიერო ნაშრომზე მომუშავე პირების მუშაობის გამარტივება და ხელშეწყობაა, მათთვის სპეციალურად შექმნილი ფუნქციონალთა ნაკრებით. თუმცა, არ უნდა ვუგულებელვყოთ მთარგმნელებისა და სხვა მწერლების მნიშვნელობა.

4.2 ნახაზზე მოცემულია დიაგრამა, სადაც წარმოდგენილია სამივე კატეგორიაში მომუშავე პირთა რაოდენობრივი შედარება.



ნახ. 4.2. გამოკითხულ ადამიანთა პროფილების განაწილება

რაც შეეხება გამოკითხულთაგან სამეცნიერო ნაშრომების სფეროში მომუშავე პირთა მიმართულებებს, ამ მხრივაც შევეცადეთ, რომ წარმოდგენილი გვექნოდა სამეცნიერო მიმართულებათა მრავალფეროვნება, რათა არ ყოფილიყო რომელიმე კონკრეტული მიმართულების მიმართ გადატანილი ზედმეტი ყურადღება.

უნდა აღინიშნოს ის ფაქტი, რომ *ფუნქციონალის მხრივ*, სხვადასხვა სამეცნიერო მიმართულებით, არსებობს განსხვავებული მოთხოვნები, მაგალითად, ქიმიის და ბიოლოგიის მიმართულებაში

საინტერესოა მოლევუღური სტრუქტურის ამგები ფუნქციონალი, ხოლო დედამიწის შემსწავლელ მეცნიერებებში კი რუკების გამოყენება.

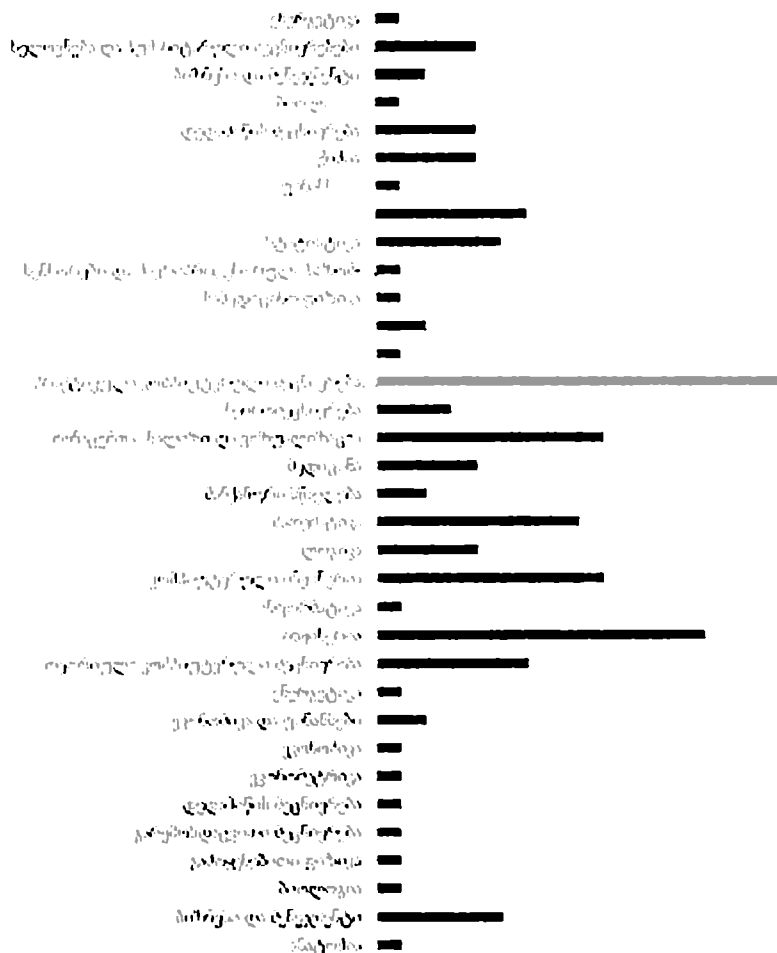
ასევე აღსანიშნავია ფაქტი, რომ როგორც თემის მოდელირების შედეგებში ვთქვით, მანქანური დასწავლის ალგორითმების მეშვეობით, შექმნილ იქნა სამეცნიერო ნაშრომთა ახალი კლასტერი, თუმცა ამან ხელი არ შეუშალა გამოკითხვაში მითითებულ არსებულ სამეცნიერო ნაშრომებს, რადგან ბიზნეს პროცესების სრულყოფა, არსებული ტექსტური რედაქტორების ფუნქციონალის ნაკლოვანებებზე და ახალი ფუნქციონალის დამატება კვლევისთვის ძალიან მნიშვნელოვანია.

როგორც 4.3 ნახაზზე ჩანს რესპოდენტთა სამეცნიერო მიმართულებით გამოკითხულ ადამიანთა მრავალფეროვნება აღინიშნება. თუმცა, უმეტესობა ტექნიკური მიმართულებით მუშაობს. ეს მიმართულებებია:

- კომპიუტერული მეცნიერება;
- კომპიუტერული ინჟინერია;
- პროგრამული ინჟინერია;
- მონაცემთა ანალიზი და ვიზუალიზაცია;
- ხელოვნური ინტელექტი;
- ფიზიკა;
- მათემატიკა;
- ბიზნესი;
- მენეჯმენტი და სხვ.

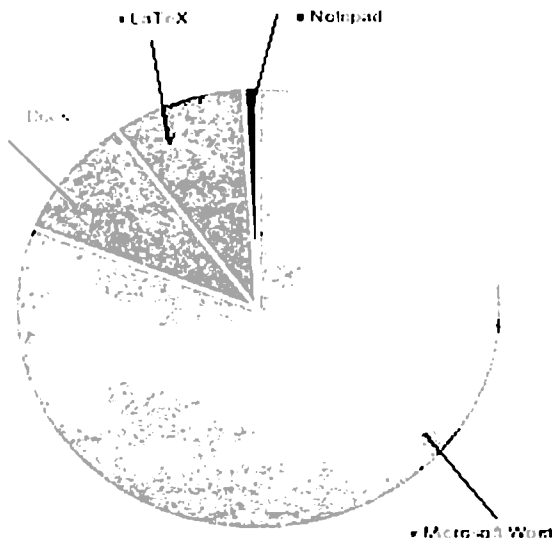
რაც შეეხება კონკრეტულ შეკითხვებს, რომლებიც განკუთვნილი იყო უმნიშვნელოვანესი ინფორმაციის მისაღებად, იმისა თუ რა ბიზნესპროცესების და ფუნქციონალის გაუმჯობესებაა საჭირო არსებული ტექსტური რედაქტორებისთვის სხვადასხვა მიმართულების სამეცნიერო და სხვა ტიპის ნაშრომებთან მუშაობის დროს, მათი მნიშვნელობები მოგვიანებით გვექნება წარმოდგენილი.

რესპოდენტთა საზოგადოებრივი მოძრაობები



ნახ. 4.3. რესპოდენტთა სამეცნიერო მიმართულებები

პირველი შეკითხვა, რა თქმა უნდა, გახლავთ იმასთან დაკავშირებით, თუ რა ტიპის ტექსტურ რედაქტორს იყენებდა რესპოდენტი. კითხვარში შეთავაზებული პასუხები გახლდათ Microsoft Office Word, Google Docs, LaTeX და Notepad. ტექსტური რედაქტორების პროცენტული გადანაწილება იხილეთ 4.4 ნახაზზე.



ნახ. 4.4. რესპოდენტთა მიერ გამოყენებული რედაქტორების თანაფარდობა

როგორც ნახაზიდან ჩანს გამოკითხულთა უმეტესობა, იყენებს Ms Word ტექსტურ რედაქტორს. მეორე და მესამე ადგილს დაახლოებით თანაბარი მნიშვნელობებით LaTeX და Google Docs იკავებს. როგორც ხედავთ, რაც არ უნდა გასაკვირი იყოს აგრეთვე გვაქვს Notepad გამოყენების შემთხვევები, რასაც თავისი მიზეზები

გააჩნია. რესპოდენტებს ჰქონდათ შესაძლებლობა სხვა აპლიკაციის შემთხვევაში, თავად დაემატებინათ რაიმე სხვა ტექსტური რედაქტორი, თუმცა ასეთი შემთხვევა არ დაფიქსირებულა, თუ არ ჩავთვლით არასწორ ინფორმაციას, რომელიც შევითხვისათვის არსებულ ბარიერს, რომ ეს უნდა იყოს ტექსტური რედაქტორი სცდება. ასეთი შემთხვევები დაფიქსირდა რამდენჯერმე, მაგალითად, რამდენიმე რესპოდენტმა დააფიქსირა, რომ იყენებს Microsoft Powerpoint-ს ძირითად ტექსტურ რედაქტორად. შესაბამისად მოხდა მსგავსი მონაცემების ამოშლა საბოლოო რეზულტატებიდან.

თითოეულ არჩეულ აპლიკაციას, თავისი მიზეზები გააჩნია და გვადლევს მიმართულებებს, განსხვავებული გზებით. მოდით, სათითაოდ აღვწეროთ რისი მანიშებელია ამა თუ იმ აპლიკაციის არჩევა.

Microsoft Word – როგორც 4.4 ნახაზზე ჩანს, მომხმარებელთა უმეტესობა სწორედ კომპანია Microsoft-ის აღნიშნულ საშუალებას იყენებს მუშაობის პროცესში. აღნიშნულ ფაქტს რამდენიმე მიზეზი აქვს. პირველი ის რომ იგი იმავე კომპანიის შექმნილია, რომელმაც მსოფლიოში ყველაზე გავრცელებული ოპერაციული სისტემა შექმნა, აგრეთვე მოცემული აპლიკაცია ერთ-ერთი ყველაზე სტაჟიანია და შესაბამისად მასზე მიჩვევადობის მაჩვენებელიც ძალიან დიდი გახლავთ. უნდა აღინიშნოს ის ფაქტი, რომ მისი მაღალი სტაჟიანობა არ აძლევს მას მოქნილობას განვითარებისთვის, სწორედ არსებულ ფუნქციონალსა და დიზაინზე არსებული მაღალი მიჩვევადობის გამო. თუ გადავხედავთ ამ პროდუქტის განვითარების ისტორიას წლების განმავლობაში, დავინახავთ რომ შესამჩნევი განვითარება, მხოლოდ სხვადასხვა ღილაკების დიზაინის ცვლილებაშია და არა აპლიკაციის ფუნქციონალის მნიშვნელოვან განვითარებაში [82].

ფაქტი, რომ Microsoft Word ყველაზე ხშირად გამოყენებული ტექსტური რედაქტორია, გვადლევს ინფორმაციას იმის შესახებ, რომ მომხმარებელთა უმეტესობა, ინტუიტიურად სწორედ ამ აპლიკაციის სამომხმარებლო ინტერფეისს უფრო მარტივად მიიღებს, შესაბამისად ჩვენი პროექტის სამომხმარებლო ინტერფეისი მიმსგავსებული უნდა იყოს Microsoft Word-ის ინტერფეისს და ძირითადი და უმნიშვნელოვანესი ფუნქციონალი, რომელსაც ყველა მწერალი იყენებს მუშაობის პროცესში, განლაგებული უნდა იყოს იმავე ადგილას, სადაც ის აღნიშნულ აპლიკაციაშია.

Google Docs – უნდა აღინიშნოს, რომ მოცემული აპლიკაცია, Microsoft Word-თან შედარებით ახალგაზრდა პროგრამაა, თუმცა მომხმარებელთა სიმრავლით მაინც გამოირჩევა. ამ ფაქტის, რამდენიმე მიზეზი არსებობს. პირველი ის, რომ იგი ტექნოლოგიური გიგანტის Google-ის შექმნილია და საიმედოობით ღირსეულად უწევს კონკურენციას პირველ ადგილოსანს. თუმცა, უნდა აღინიშნოს ისიც, რომ მას არ აქვს საკუთარი ფორმატის დოკუმენტები და შეიძლება ითქვას Microsoft Word-ის დოკუმენტებთან სამუშაოდაა შექმნილი. თუმცა, ამის მიუხედავად მას მაინც გააჩნია რამდენიმე უპირატესობა, მის პირველ კონკურენტთან შედარებით. პირველ უპირატესობად ჩაითვლება ის, რომ იგი უფასოა. ეს კი მეორე უპირატესობას წარმოქმნის, მომხმარებელი არ იქნება იძულებული ყოველი ახალი ვერსიის გამოსვლის შემდეგ ახლიდან შეიძინოს რედაქტორი. აგრეთვე ერთ-ერთი მნიშვნელოვანი უპირატესობა გახლავთ ის, რომ რედაქტორი Web-ზეა დაფუძნებული, რაც მრავალ უპირატესობას მატებს კომპიუტერის აპლიკაციებთან შედარებით. მას ყველა ის უპირატესობა აქვს, რაც პროგრესული ვებ აპლიკაციების პარაგრაფშია გადმოცემული. მისი მნიშვნელოვანი უპირატესობებია გაზიარების სიმარტივე და ნაშრომების შენახვის ბიზნესპროცესი. სისტემაში ნაშრომები

ავტომატურად ინახება იმ უპირატესობით, რომ მომხმარებელმა იცის, ისინი სპეციალურ ადგილასა საიმედოდ არის შენახული და მისი დაკარგვის ან წაშლის შანსი არ არსებობს. თუმცა უნდა აღვნიშნოთ ისიც, რომ Google Docs-ს ახასიათებს მის კონკურენტი, MsWord-თან შედარებით, ფუნქციების სიმწირე. აგრეთვე მისი მომხმარებლის ინტერფეისი შედარებით რთული აღსაქმელია, მეტად რთულია არსებული, ისედაც მწირი ფუნქციონალის პოვნა.

შეიძლება აღინიშნოს, რომ Google Docs-ის უპირატესობები კარგადაა გამოყენებული ჩვენ საპროექტო ნაშრომში.

LaTeX - სისტემა, სამეცნიერო ნაშრომებთან მუშაობის ერთ-ერთი ყველაზე მრავალფუნქციური პროგრამაა. უნდა აღინიშნოს, რომ LaTeX აგრეთვე უფასოა, თუმცა არ არის პროგრესული Web აპლიკაცია და მონაცემთა შენახვისა და ორგანიზების საკითხებში არაა მხარდამჭერი. დღეისათვის LaTeX-ს ძირითადად სამეცნიერო ნაშრომების შესაქმნელად იყენებენ, რადგან მას აქვს კონკურენტებთან შედარებით კარგი ფორმულების შეყვანის ფორმატი, თუმცა უნდა ითქვას ისიც, რომ მის უკან არ დგას რომელიმე ტექნოლოგიური გიგანტი. აპლიკაცია ღია კოდითაა და ეს მისი განვითარების უმთავრესი გზაა, ეს კი მის სტაბილურობაზე არც თუ ისე კარგად მეტყველებს.

LaTeX-ის გამოყენების მთავარ მიზეზად შეგვიძლია ჩავთვალოთ ის, რომ ფასის უპირატესობასთან ერთად, მას გააჩნია უპირატესობა ფუნქციონალშიც. თუმცა, იგი არც ისეთი მდიდარია, რომ ყველა ტიპის სამეცნიერო ნაშრომზე იყოს მორგებული. ეს ფაქტი, იმასთან ერთად, რომ LaTeX-ს მრავალი მომხმარებელი ჰყავს, კიდევ ერთხელ უსვამს ხაზს ჩვენი კვლევის აქტუალობასა და მის სამეცნიერო ღირებულებას.

ჩვენი საპროექტო ნაშრომის ერთ-ერთი უმნიშვნელოვანესი მიზანი ემსახურება იმას, რასაც ნაწილობრივ LaTeX-ის არსებობა,

თუმცა ჩვენი მიზანია სამეცნიერო კვლევების საფუძველზე შეიქმნას ისეთი მიდგომა, რომელიც ფუნქციონალის მხრივ უფრო სრულყოფილი იქნება ვიდრე ხსენებულის. მიიღებს ყველა იმ უპირატესობას რაც პროგრესულ ვებ აპლიკაციას გააჩნია და ხელშემწყობი იქნება არა მხოლოდ წერის, არამედ ნაშრომზე მუშაობის სხვა ბიზნეს პროცესებისთვისაც.

რაც შეეხება Notepad-ის გამოყენებას, უნდა აღვნიშნო, რომ ყველა, შემთხვევაში, როდესაც რესპოდენტმა ნაშრომის შექმნის აპლიკაციად Notepad-ი აირჩია, მასთან ერთად აირჩია რაიმე სხვა რედაქტორიც. მაგალითად, ერთად იქნა არჩეული Microsoft Word და Notepad, ან LaTeX Docs და Notepad, რაც მეტყველებს იმაზე, რომ აღნიშნული აპლიკაციების მულტიპლატფორმულობა და დოკუმენტების სერვერზე შენახვის შესაძლებლობა არც ისე კომფორტულია მომხმარებლისათვის, და ისინი რიგ შემთხვევაში ელემენტარულ ტექსტის რედაქტორებს იყენებენ მარტივი დავალებებისათვის.

გამოკითხვაში მონაწილე რესპოდენტებმა, ფავორიტი ტექსტური რედაქტორის არჩევის შემდეგ გასცეს პასუხი შეკითხვას: „გთხოვთ აღწეროთ რატომ იყენებთ შერჩეულ აპლიკაციას“. მიღებულმა შედეგებმა მოლოდინები სწორად გაამართლა და ისინი ასე გამოიყურება: MsWord-ის მომხმარებელთა უმეტესობამ, მისი გამოყენების მთავარ მიზეზად მასთან მიჩვევა დაასახელა. მისი გამოყენების მიზეზები, პასუხების მიხედვით ყოველთვის ზოგად ხასიათს ატარებდა და არ ყოფილა შემთხვევა, როდესაც რომელიმე კონკრეტულ ფუნქციასა თუ ბიზნესპროცესზე ყოფილიყო ხაზის გასმა. MsWord გამოყენების მიზეზებში ვხვდებით ისეთ პასუხებს როგორებიცაა: „კომფორტულია“, „მოსახერხებელი და მრავალ-ფუნქციურია“, „მოსახერხებელია ინტერფეისის გამო“, „ყველაზე მოხერხებულია და დიდი ხნის წინ დავიწყე“, „დიდი ხნის წინ დავიწყე და გამოცვლის მიზეზი ჯერ არ მაქვს“, „დიდი ხანია მასთან

ვმუშაობ, ალბათ მივეჩვიე“, „სიმარტივე, მრავალფუნქციურობა, შეგუებადობა წლების განმავლობაში“ და სხვ. იყო ასეთი პასუხებიც კი: „კარგია“, „მოსახერხებელია“ მიღებული შედეგებიდან, ორივე ქართულ და ინგლისურ ვერსიებში პასუხების 99% აღნიშნავდა სამ ფაქტს:

1) *აპლიკაცია მოსახერხებელია* – რაც შეიძლება ნიშნავდეს იმას, რომ მომხმარებელი მიჩვეულია მოცემულ აპლიკაციას და ახლა ის მისთვის მოსახერხებელია, მას საინტერესო მომხმარებლის ინტერფეისი აქვს, მომხმარებელს უკვე აქვს MsWord ნაყიდი და კომპიუტერში დაყენებული. არსებული აპლიკაციის ფუნქციონალი მომხმარებლისთვის დამაკმაყოფილებელია (თუმცა ეს ხშირად არ ნიშნავს იმას, რომ ეს რეალურადაც ასეა);

2) *აპლიკაციას მიჩვეული ვარ* – რესპოდენტები აფიქსირებდნენ ფაქტს, რომ ისინი მიჩვეულები იყვნენ MsWord-თან მუშაობას და ამით ერთგვარ ლოიალურობას გამოხატავდნენ. ჩვენი პროექტისთვის მნიშვნელოვანია თუ რამდენადაა ეს მომხმარებლები სიახლის მიმართ განწყობილნი. საკითხი საკმაოდ რთულია, თუმცა შეკითხვებიდან მიღებული პასუხებით და აგრეთვე იმ ფაქტიდან გამომდინარე, რომ Google Docs-ს და LaTeX-ს მრავალრიცხოვანი მომხმარებელი ყავს, შეგვიძლია ვივარაუდოთ, რომ ინფორმაციულ საზოგადოებაში, სიახლის მიმართ მიმღებლობა საკმაოდ მაღალია;

3) *აპლიკაცია მარტივი გამოსაყენებელია* – რაც აგრეთვე მეტყველებს MsWord-ის მიმართ მიჩვევადობაზე. აქედან გამომდინარე, როგორც უკვე აღვნიშნეთ, ჩვენი საპროექტო ნაშრომის სამომხმარებლო ინტერფეისი, მიმსგავსებულია MsWord-ის ინტერფეისს, რაც მის მიმართ მიჩვევადობის ბარიერის გადალახვისთვის მნიშვნელოვანი ნაბიჯია.

ამასთან ერთად იყო ისეთი პასუხებიც, რომლებიც ხაზს უსვამდა აპლიკაციის დადებით მხარეებს, მათ შორის ერთ-ერთი

ასეთი პასუხი იყო: „ამ აპლიკაციაში კომენტარებისა და ჩასწორებების ნახვა შედარებით მარტივია სხვა აპლიკაციებისგან განსხვავებით“. ყოველი ასეთი პასუხი აუცილებლად გასათვალისწინებელია და ყურადღება იქნა მინიჭებული საპროექტო ნაშრომის აპლიკაციის პროექტირებისას.

მორიგი შეკითხვა, რომელსაც რესპოდენტებმა გასცეს პასუხი, არია შემდეგი – „ყველაზე მეტად რა მოგწონთ შერჩეულ აპლიკაციაში?“. მოცემული შეკითხვა, ხშირ შემთხვევაში ხაზს უსმევს იმის მიზეზს, თუ რატომ არ იყენებენ მომხმარებლები ისეთ სტანდარტულ აპლიკაციებს, როგორიცაა მაგალითად MsWord-ი. რაც შეეხება, სტანდარტულ აპლიკაციებს, ამ შემთხვევაში პასუხები ძალიან ჰგავდა წინა შეკითხვის პასუხებს. გამოკითხულთა უმეტესობა ხაზს უსვამდა მათი ფავორიტი აპლიკაციის სტაბილურობას, მოქნილობასა და სიმარტივეს.

LaTeX გამოყენების შემთხვევაში, პასუხები განსხვავებულია და ნათელს ჰფენდა როგორც ამ პროგრამის მნიშვნელოვან ფუნქციონალს, აგრეთვე არსებული სხვა პოპულარული რედაქტორების ნაკლოვანებებსაც. მაგალითად, წარმოვადგინოთ რამდენიმე პასუხი აღნიშნული შეკითხვიდან და განვიხილოთ ისინი:

„სტრუქტურის თავისუფლების მაღალი დონე და მათემატიკურ ფორმულათა წერის სიმარტივე“ აღნიშნული კომენტარი, ერთ-ერთი გამოხატულებაა იმისა, თუ რა ნაკლოვანება აქვს დღევანდელ, ყველაზე პოპულარულ ტექსტის რედაქტორებს. ესენია სტრუქტურის თავისუფლების არც თუ ისეთი მაღალი დონე. ჩვენ ნაშრომში, პროექტი ვებზეა დაფუძნებული, რაც მას სტრუქტურის სიმდიდრეს მატებს. აგრეთვე, აღსანიშნავია რედაქტორში ფორმულების წერის არც თუ ისეთი მარტივი გზა. უნდა აღინიშნოს, რომ ყველაზე პოპულარული რედაქტორების უახლეს ვერსიებში, დამატებულია LaTeX ფორმულების კოდის აღქმის საშუალება,

თუმცა ამ ფუნქციის გამოყენება ისეთი მარტივი არაა როგორც თავად LaTeX-ში და არც დიდი მნიშვნელობა ენიჭება. მისი დამატების მთავარ მიზეზად შეგვიძლია დავასახელოთ ის, რომ ლატექსის მომხმარებლებს არ ჰქონოდათ შეზღუდულობა ამ პროგრამებში მუშაობის პროცესში. აგრეთვე LaTeX-თან დაკავშირებული საინტერესო პასუხებია, რომ ის მრავალმხრივია, რაშიც არ იგულისხმება იგივე მრავალმხრივობა რაც Microsoft Word-ში. აგრეთვე, ერთ-ერთი კომენტარი ეხება იმას, რომ შესაძლებელია ფუნქციებისა და პარამეტრების პროგრამირების გზით დაწერა, რაც წერის პროცესს უფრო სწრაფს და მოხერხებულს ხდის. პროექტის არქიტექტურაში, რა თქმა უნდა გათვალისწინებულია აღნიშნული ფუნქციონალიც.

შემდეგი შეკითხვა, რომელსაც რესპოდენტებმა გასცეს პასუხი იყო „ყველაზე მეტად რა არ მოგწონთ შერჩეულ აპლიკაციაში?“.

შეკითხვაზე მიღებული პასუხების გაანალიზების მიზანი, წერის პროცესის გაუმჯობესება, არსებული პრობლემების აღმოჩენა და მათი გამოსწორებაა. აღნიშნულ შეკითხვაზე პასუხი საკმაოდ რთული გასაცემია და შესაბამისად ამ პასუხების რაოდენობა შედარებით ნაკლებია, თუმცა უნდა აღინიშნოს ისიც, რომ ეს პასუხები, ფუნქციონალის სრულყოფის მხრივ, შედარებით უფრო მნიშვნელოვანია, ვიდრე წინა შეკითხვაზე მიღებული პასუხები.

აღნიშნულ შეკითხვაზე, გაფილტვრის შემდეგ, არც ისე ბევრი სასარგებლო პასუხი დარჩა, თუმცა თითოელი მათგანი ძალიან ძვირფასი და დამხმარეა პროექტის არქიტექტურული მოდელის შექმნისათვის. გამოკითხვის შედეგებში, იშვიათად, მაგრამ გვხვდება პასუხები, სადაც რესპოდენტები აღნიშნავენ, რომ ყველაზე პოპულარული რედაქტორები არასრულყოფილია და მას რთული სამომხმარებლო ინტერფეისი აქვს. მაგრამ, ამჯერად

კონცენტრირდეთ ისეთ პასუხებზე, რომლებიც ხელშემწყობი იქნება ახალი ფუნქციონალის საჭიროებების განსაზღვრისთვის.

ასეთი პასუხებიდან ერთ-ერთში მითითებულია მისი არასრულფასოვანი გრაფიკული სერვისი, რაც აღსანიშნავია, რომ ჩვენი პროექტის ერთ-ერთი მნიშვნელოვანი კომპონენტია. აგრეთვე გვხვდება კომენტარები, სადაც აღნიშნულია ის სირთულე, რაც ამ პროდუქტის სხვადასხვა ვერსიებში მუშაობისას გვხვდება.

ერთ-ერთი ასეთი ფუნქციონალური გვერდის ნუმერაციასთანაა დაკავშირებული. როგორც უკვე აღვნიშნეთ, ვებ პლატფორმის ერთ-ერთი უპირატესობის მიხედვით, ჩვენ დაპროექტებულ აპლიკაციას არ ექნება პრობლემები, სხვადასხვა ვერსიის გამოსვლასთან დაკავშირებით. სტანდარტულ რედაქტორებთან დაკავშირებით, აგრეთვე გვხვდება კომენტარი, რომ დიდ დოკუმენტებთან მუშაობა უფრო რთული ხდება და სისწრაფეც საგრძნობლად იკლებს. ასეთ პროცესებთან მუშაობის ასაჩქარებლად, სისტემაში შესაძლებელია Angular ტექნოლოგიის ვირტუალური სეროლინგის ფუნქციის გამოყენება. ამ დროს ეკრანზე სრული გვერდების მხოლოდ ის ნაწილი იქნება რენდერირებული, რომლებსაც მომხმარებელი ფიზიკურად ხედავს.

ერთ-ერთი აღნიშნული ნაკლოვანება აგრეთვე გახლავთ შემდეგი: „სიმბოლოების პოვნა საკმაოდ რთულია“. მოცემულ პარაგრაფში დეტალურად განვიხილავთ სიმბოლოების მნიშვნელობას, სხვადასხვა ტიპის სამეცნიერო ნაშრომთან მუშაობის პროცესში.

კიდევ ერთი მეცნიერის პასუხი ზემოაღნიშნულ შეკითხვაზე შემდეგია: „ვისურვებდი, რომ ეს უფრო კომფორტული იყოს ჩვენს უნარებში მუშაობისთვის.“ - აღნიშნული პასუხი სრულად აღწერს რამდენად მნიშვნელოვანია, სხვადასხვა ტიპის სამეცნიერო ნაშრომებთან მუშაობის დროს შესაბამისი ფუნქციონალისა და ინტერფეისის არსებობა.

რაც შეეხება პასუხებს LaTeX-თან დაკავშირებით, ერთ-ერთი პასუხი გახლდათ ის, რომ მისი ფუნქციონალის მოძებნისთვის აღნიშნული მეცნიერი დიდ დროს ხარჯავდა Google-ში. აღნიშნული პასუხი კიდევ ერთხელ უსვამს ხაზს, თუ რამდენად მნიშვნელოვანი შეიძლება იყოს, სხვადასხვა ტიპის სამეცნიერო ნაშრომებთან მუშაობის დროს სპეციალური ინტერფეისის შეთავაზება, სადაც ამ ტიპის ნაშრომებზე ყველაზე ხშირად გამოყენებული ფუნქციონალი შედარებით მარტივი საპოვნელი იქნება, ვიდრე ინტერფეისზე, რომელიც ზოგადი ტექსტ-ინფორმაციის აგებისთვისაა შექმნილი.

შეკითხვათა სერიაში ერთ-ერთი ყველაზე მნიშვნელოვანი იყო „რას გააუმჯობესებდით (რის დანაკლისს განიცდით) შერჩეულ ინსტრუმენტში?“. შესაბამისად, მიღებული პასუხებიც ყველაზე მეტადაა ინფორმატიული და საკმაოდ საინტერესო. როგორც აღმნოვნდა, რესპოდენტები ხშირად საკმაოდ კრეატიულები იყვნენ. მრავალი პასუხიდან გამოვყოფთ რამდენიმეს და მათ მნიშვნელობას გაუვსვამთ ხაზს.

ერთ-ერთი იდეაა „ტექსტის შეტანა ხმით“, რაც ჩვენ პროექტში გათვალისწინებულია. აგრეთვე არაერთი კომენტარია იმასთან დაკავშირებით, რომ „*კარგი იქნებოდა თუ რედაქტორი რესპოდენტის სამუშაო კატეგორიისთვის მეტად მორგებული იქნებოდა*“, რაც ჩვენი საპროექტო კვლევის მნიშვნელოვანი მიზანია. აგრეთვე, არაერთი კომენტარი იყო იმასთან დაკავშირებით, რომ არსებულ ტექსტურ „*რედაქტორებს არ შეუძლია ფორმულებთან და გრაფიკებთან სრულყოფილი მუშაობა*“. სურვილები იქნა გამოთქმული *მოლეკულური სტრუქტურის აგებასთან, ელექტრული სქემების ხაზვასთან* და სხვა ისეთ ფუნქციონალთან დაკავშირებით, რომელიც არ გააჩნია დღევანდელ ტექსტურ რედაქტორებს. ისინი გათვალისწინებულია ჩვენ საპროექტო კვლევაში. ასევე პასუხებში

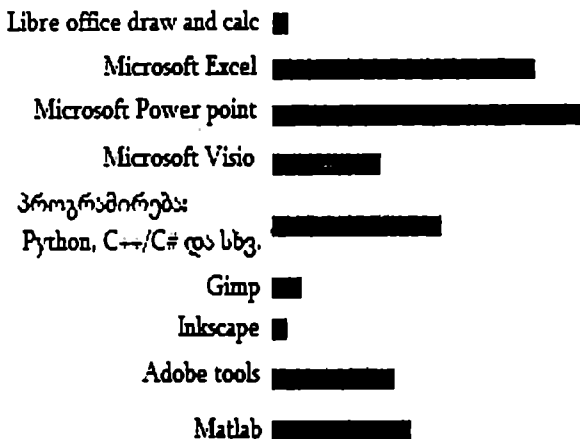
მოიაზრებოდა მრავალი იდეა, რომელთა ნაწილიც ჩვენი პროექტის არქიტექტურულ მოდელში უკვე გათვალისწინებულია.

საბოლოოდ, იმის გასაგებად, თუ კიდეც რა ფუნქციონალის ნაკლებობას განიცდიდნენ რესპოდენტები ტექსტურ რედაქტორებთან მუშაობის პროცესში, დავსვით ორი შეკითხვა:

1) „რა დამატებით ინსტრუმენტებს იყენებთ ნაშრომის შექმნის პროცესში?“ და

2) „კონკრეტულად რაში იყენებთ შერჩეულ დამხმარე ინსტრუმენტებს“.

გამოკითხულთა პასუხები მოცემულია 4.5 ნახაზზე.



ნახ. 4.5. რესპოდენტთა მიერ გამოყენებული დამხმარე ინსტრუმენტები

როგორც ჩანს, გამოკითხულ რესპოდენტთათვის, ტექსტურ ინფორმაციაზე მუშაობისას ყველაზე ხშირად გამოყენებული

პროგრამები Microsoft Power Point და Microsoft Excel-ია, აგრეთვე მაღალი მაჩვენებელია პროგრამირების, Matlab-ისა და Adobe Tools-ის გამოყენების მიმართაც. მოცემული მონაცემები კიდევ ერთხელ ადასტურებს შემდეგ ფაქტებს:

სისტემაში აუცილებელია არსებობდეს *გრაფიკების აგების ფუნქციონალი*, რომელიც დღეისათვის არსებულ რედაქტორთა უმეტესობას არ გააჩნია. მათ ინსტრუმენტებში იპოვიან ღილაკებს ახალი გრაფიკების აგებისთვის, თუმცა მასზე დაჭერის შემდეგ იხსნება Microsoft Excel ან Google sheet აპლიკაციები და მომხმარებელი ვალდებულია დიაგრამები სხვა პროგრამებში ააწყოს;

- Microsoft Powerpoint-ის და Adobe Tools-ის პაკეტების გამოყენების მაღალი პროცენტი პირდაპირ გულისხმობს იმას, რომ ტექსტური რედაქტორები განიცდის დიაგრამების აგების ფუნქციონალის ნაკლებობას;

- მთარგმნელთა გამოკითხვის საფუძველზე იქნა მიღებული ინფორმაცია სასურველი რეკომენდაციით იმის შესახებ, რომ *კარგი იქნება თუ მომხმარებლის ინტერფეისი ორ პანელად გაიყოფა, პირველ ნაწილზე მოთავსდება სათარგმნი ტექსტი, ხოლო მეორეზე იწარმოებს თარგმნის პროცესი*. ამასთან ერთად, თარგმნის დროს სათარგმნი ტექსტის ის აზრად, რომელზეც მუშაობა მიმდინარეობს, სხვა ფერით უნდა გამოირჩეოდეს ტექსტის დანარჩენი ნაწილისაგან. ამასთან ერთად კი, წერის პროცესში სათარგმნი ტექსტი ავტომატურად უნდა ჩამოდიოდეს წერის პროცესის მიხედვით და ა.შ.

4.1.2. ბიზნესპროცესების სრულყოფა

ზემოაღნიშნული შეკითხვები, პასუხს სცემს ისეთი პრობლემების მოგვარების საჭიროებას, როგორიცაა: დღევანდელ ტექსტურ რედაქტორებში მუშაობის პროცესში ფუნქციონალის, არასაკმარისობა და იძლევა ინფორმაციას იმის შესახებ, თუ რა ფუნქციონალი უნდა იყოს გათვალისწინებული ჩვენი პროექტის არქიტექტურაში.

ფუნქციონალის გარდა არის მეორე მხარეც, რისი განვითარებაც ჩვენი ნაშრომის ერთ-ერთი მნიშვნელოვანი საკითხია. ეს მხარე ბიზნესპროცესების გამარტივებაა. ამ შემთხვევაში ბიზნესპროცესებში იგულისხმება მონაცემთა შენახვის და მისი გამოქვეყნების პროცესი. მნიშვნელოვანია ასევე თანაავტორობთან ურთიერთობის გამარტივება და ამ პროცესის ავტომატიზაცია.

რესპოდენტებთან თანამშრომლობის დროს კონკრეტული შეკითხვები სწორედ ამ პრობლემების გამოსაწვევად და გასარჩევად იქნა დასმული. მიღებული შედეგებიდან გამომდინარე, გამოიკვეთა, რომ არ არსებობს რაიმე მიღებული ან საერთო გზა დოკუმენტთა შესანახად და მისი ვერსიების საწარმოებლად.

Ms Word-ის და LaTeX-ის მომხმარებლები, თავად აგვარებენ მონაცემთა შენახვასა და მისი ვერსიების კონტროლს. მართალია უახლესი MsWord-ის ვერსიაში ამავე კომპანიის ღრუბელი One Drive-ია ინტეგრირებული, თუმცა ამ ეტაპზე ის დიდი პოპულარობით ჯერ ვერ სარგებლობს. შესაბამისად, მომხმარებლებს უწევთ იზრუნონ მათი დოკუმენტების შენახვაზე და კრიტიკულ სიტუაციებში ეს პრობლემადაც კი იქცევა ხოლმე.

ხშირად ხდება მყარი დისკების დაზიანება, მონაცემთა დაკარგვა, მათი უსაფრთხოების არასათანადო უზრუნველყოფა და ა.შ. ამ პრობლემათა უმეტესობა აღმოფხვრილია Google Docs-ის შემთხვევაში, რადგან აპლიკაცია თავად ინახავს მომხმარებელთა

ფაილებს და უზრუნველყოფს მათ უსაფრთხოებას. თუმცა, უნდა აღინიშნოს, რომ Google-ის ეს პროდუქტი არ უზრუნველყოფს დოკუმენტთა ვერსიების კონტროლს.

რაც შეეხება თანაავტორებთან ურთიერთობას, გამოკითხვის შედეგად აღმოჩნდა, რომ არსებობს თანაავტორებს შორის ნაშრომის განხილვის სამი ძირითადი გზა (ნახ. 4.6).

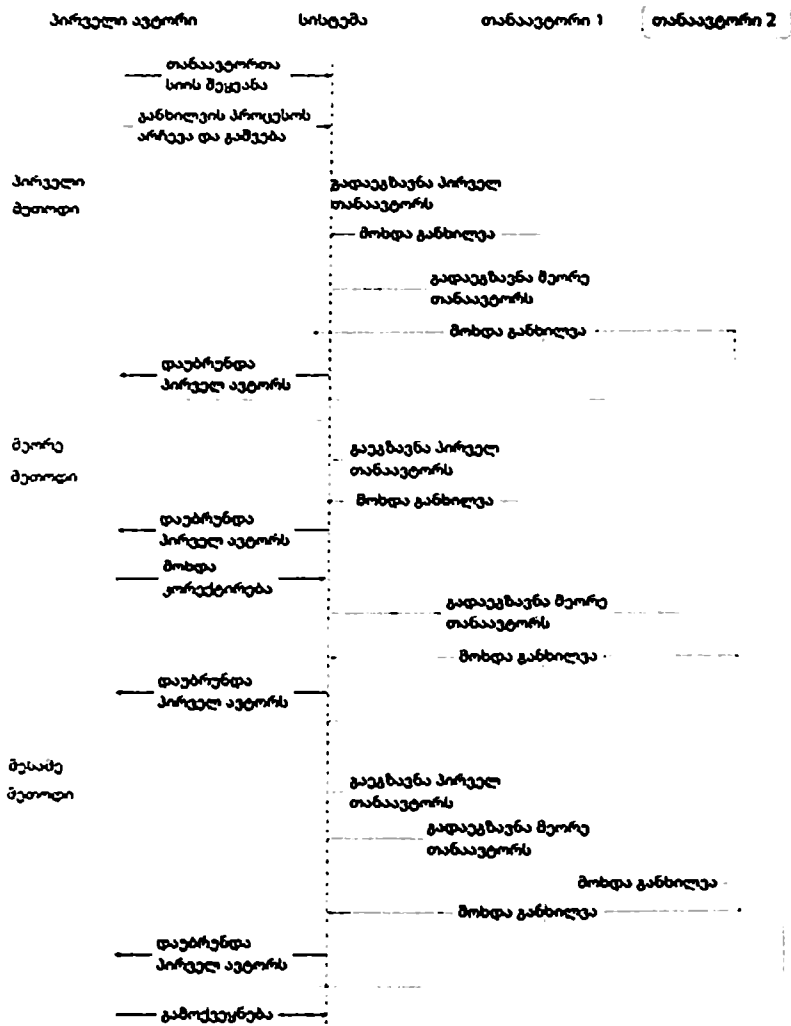
როგორც ნახაზზეა გამოსახული, არსებობს თანაავტორებთან ურთიერთობის სამი ძირითადი მეთოდი. ეს მეთოდები სტანდარტულია, თუმცა არ არის გადაწყვეტილი მათი პროცესების ავტომატიზებული სისტემის შექმნა.

ასე რომ, მნიშვნელოვნად მიგვაჩნია სამეცნიერო ნაშრომების შექმნის პროცესების ხელშეწყობი ავტომატიზებული სისტემის აგების კონცეფცია, რომელსაც ექნება ფუნქციონალი, თანაავტორებს შორის საქმიანი ურთიერთობის სრულყოფის მიზნით.

განვიხილოთ 4.6 ნახაზზე გამოსახული თანაავტორებთან ურთიერთობის სამივე გზა. იმ შემთხვევაში, თუ ნაშრომს მხოლოდ ერთი თანაავტორი ჰყავს, პროცესი მარტივია, მაგრამ თუ თანაავტორთა რაოდენობა რამდენიმეა ამ შემთხვევაში პროცესი უფრო საინტერესო ხდება.

წარმოვიდგინოთ, რომ ჩვენ სტატიას ორი თანაავტორი ჰყავს და გავუშვათ პროცესი ჩვენი სისტემის მეშვეობით. როდესაც გავაზიარებთ ნაშრომის პირველ ვერსიას და უკვე დროა ის თანაავტორებმა განიხილონ, პირველ რიგში, სისტემაში უნდა შევიყვანოთ ელ-ფოსტის მისამართები. ამის შემდეგ ავირჩევთ ნაშრომის განხილვის სამი შესაძლო პროცესიდან ერთ-ერთს და გავუშვებთ ნაშრომს განსახილველად.

პირველი მეთოდის არჩევის შემთხვევაში, ნაშრომი ჯერ გადაეგზავნება რიგით პირველ თანაავტორს, მისი განხილვის, კომენტარებისა და შენიშვნების ჩამატების შემდეგ კი გადაეგზავნება მეორე თანაავტორს.



ნახ. 4.6. თანაავტორებთან კავშირის ავტომატიზაციის გზები

მხოლოდ მას შემდეგ, რაც მეორე ავტორიც დაასრულებს განხილვას, ჩვენ დაგვიბრუნდება მათ მიერ შესწორებული ვერსია, როგორც პირველ ავტორს.

რაც შეეხება მეორე მეთოდს, ამ შემთხვევაშიც ნაშრომი ჯერ რიგით პირველ თანაავტორს გადაეგზავნება და როდესაც ის განხილვას დაასრულებს, განსხვავებით პირველი მეთოდისაგან, ეს განხილვა დაგვიბრუნდება კვლავ ჩვენ, როგორც პირველ ავტორებს. მას შემდეგ, როდესაც უკვე ჩვენ შევიტანთ სასურველ ცვლილებებს პირველი განხილვის შემდეგ, სისტემაში მივუთითებთ, რომ ნაშრომი მზადაა შემდეგი განხილვისათვის. შესაბამისად ამის შემდეგ ნაშრომი გადაეგზავნება მეორე თანაავტორს და ა. შ.

რაც შეეხება მესამე მეთოდს, იგი ძირითადად ისეთ შემთხვევებში გამოიყენება, როდესაც ნაშრომის განხილვისას დროის დეფიციტთან გვაქვს საქმე. მოცემულ მეთოდში, როდესაც ნაშრომს განსახილველად გავგზავნით, ის პარალელურად გაეგზავნება ორივე თანაავტორს.

სისტემაში შეიქმნება განხილვის ორი ასლი და ორივე თანაავტორი დამოუკიდებლად განიხილავს ნაშრომს და დაუბრუნებს რედაქციას (პირველ ავტორს). ასეთ დროს, როგორც წესი ცვლილებების გაკეთება შედარებით უფრო რთულია ვიდრე პირველი ორი მეთოდის გამოყენებისას, მაგრამ დროის მოგების მხრივ ეს მეთოდი საგრძნობლად უფრო სწრაფია. ასევე უნდა აღინიშნოს ფაქტი, რომ აღნიშნული პროცესები შეიძლება რამდენიმე ეტაპად დაიყოს და ერთი ნაშრომის განხილვის პროცესში სამივე მეთოდი იქნას გამოყენებული.

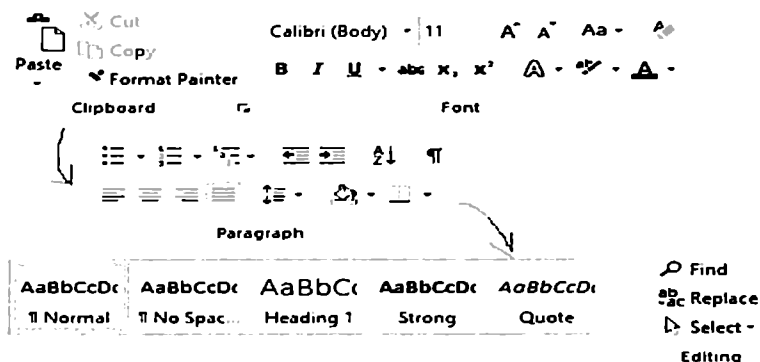
4.2. სხვადასხვა ტიპის სამეცნიერო ნაშრომში არსებული ფორმების კვლევა და შედეგები

ჩვენი პროექტის ერთ-ერთი უმნიშვნელოვანესი მიზანია ავტორთა ხელის შეწყობა წერის პროცესში, სხვადასხვა ტიპის სამეცნიერო ნაშრომებზე მომუშაობის დროს. სამუშაო ფუნქციონალისა და მომხმარებლის ინტერფეისის შესაბამისობა საკვლევ თემასთან, მეცნიერისათვის და მწერლისათვის ნამდვილად მხარდამჭერი იქნება მუშაობის პროცესში.

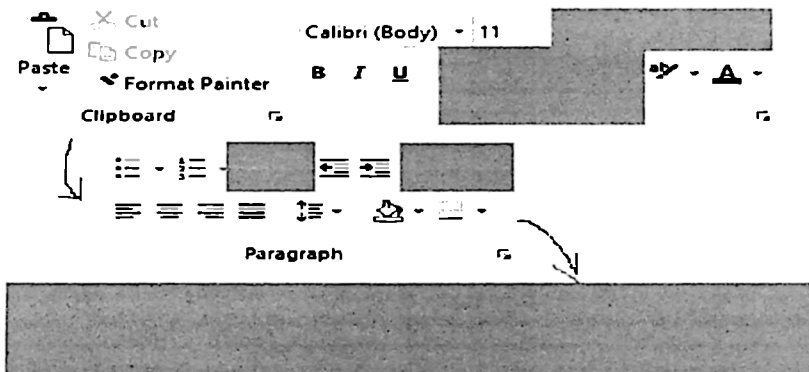
ისეთი ტიპის ფუნქციონალი, რომელიც სხვადასხვა კატეგორიის მწერლისთვის საერთოა, ყველა დოკუმენტთან მომუშავე რედაქტორს აქვს. ეს ფუნქციები საჭიროა ყველა ტიპის ტექსტის შესაქმნელად და რამდენიმე მოიცავს მნიშვნელოვან ფუნქციას, რომლებიცაა ფონტის ფერი, ზომა და ა. შ., მისი ადგილმდებარეობა და მისი განლაგება დოკუმენტში.

4.7 ნახაზზე ჩანს ერთ-ერთი ყველაზე პოპულარული ტექსტური რედაქტორის ინსტრუმენტთა მთავარი გვერდი. 4.8 ნახაზზე კი ჩანს ამ ინსტრუმენტთა ის ნაწილი, რომელიც ყველა მწერლისთვისაა სასარგებლო.

თუ ნახაზების სხვაობას დავაკვირდებით, დავინახავთ, რომ მომხმარებლებს პირველივე გვერდზე აქვთ მრავალი უსარგებლო ფუნქცია. მათ უწევთ სამუშაო პროექტის თემისათვის უფრო მნიშვნელოვანი ფუნქციების შედარებით რთული ძებნა.



**ნახ. 4.7. Microsoft Office Word-ის ინსტრუმენტების
პირველ ჩანართზე მოცემული ფუნქციონალი
(ინსტრუმენტის მთლიანი გეგმის საში ნაწილი)**



ნახ. 4.8. Ms Office Word-ის ინსტრუმენტების პირველ ჩანართზე მოცემული ფუნქციონალი, რომელზეც მონიშნულია ისეთი ფუნქციები რომელიც არ არის განკუთვნილი ყველა ტიპის მწერლისთვის

ხშირია აგრეთვე საჭირო ფუნქციონალის არარსებობის შემთხვევებიც. მაგალითად, თუ ნაშრომზე მუშაობს ნეირომეცნიერი და მას უნდა, რომ მუშაობის პროცესში დამატებით ინსტრუმენტების გარეშე შეძლოს სამედიცინო ფაილების გახსნა, მას ამ ფუნქციის შესრულება საერთოდ არ შეუძლია, ის აუცილებლად უნდა მუშობდეს მოწყობილობაზე, სადაც ამ ფაილების გახსნისთვის საჭირო პროგრამული უზრუნველყოფაა დაინსტალირებული, გახსნას ფაილი სხვა პროგრამაში, გადაიღოს სურათი და შემდეგ ჩასვას ტექსტურ რედაქტორში. აღნიშნული პრობლემის გადასაჭრელად, ჩატარებული იქნა კვლევა სხვადასხვა კატეგორიის სამეცნიერო ნაშრომების განხილვის გზით.

თითოეული ნაშრომისთვის ჩაინიშნა მასში გამოყენებული განსხვავებული ფორმების სია. ყურადღება იქნა გამახვილებული ისეთი ფორმების მიმართ, რომელთა გამოყენებისთვისაც რაიმე სპეციალური ფუნქციების არსებობაა საჭირო. ასეთ ფორმებად აღიქმება შემდეგი:

- 1) დიაგრამები;
- 2) გრაფიკები (Charts);
- 3) ცხრილები;
- 4) ფორმულები;
- 5) სამედიცინო სურათები (Nifty, X-ray);
- 6) მოლეკულური სტრუქტურები;
- 7) პროგრამული კოდები;
- 7) რუკები.

ამასთან ერთად ყურადღება იქნა გამახვილებული კონკრეტულად – რა ტიპის გრაფიკები და დიაგრამებია გამოყენებული ამ ნაშრომებში (მაგალითად, კლასების დიაგრამა, BPMN დიაგრამა, სხვადასხვა ტიპის გრაფიკები და სხვ.). აგრეთვე – ნაშრომებში გამოყენებულ სიმბოლოებზე, რადგან გამოკითხვის შედეგად,

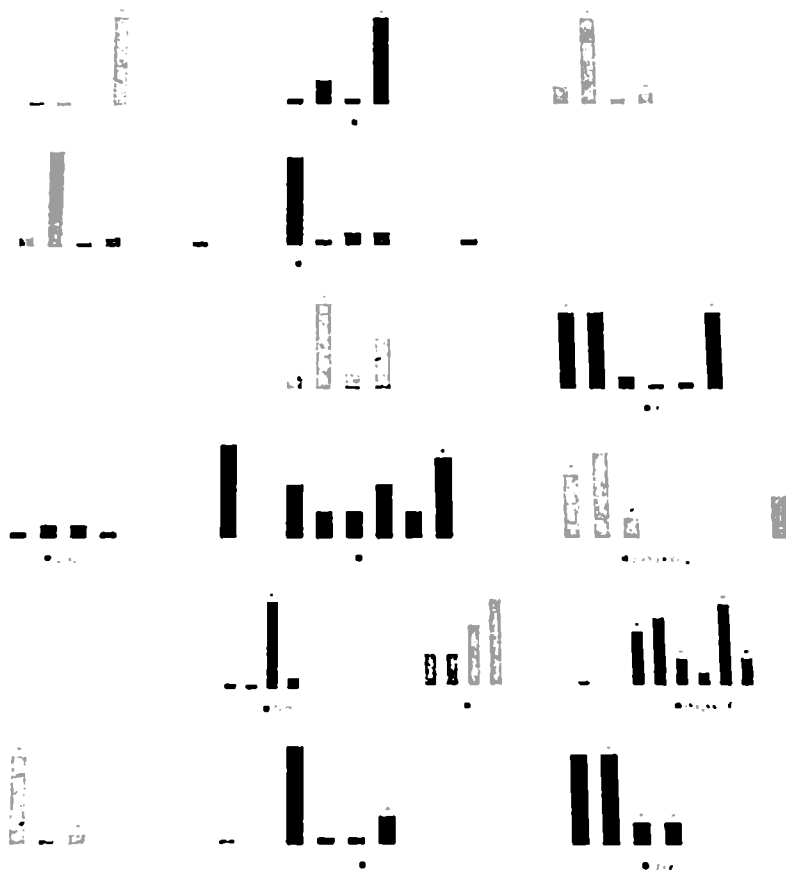
მივიღეთ არაერთი კომენტარი იმასთან დაკავშირებით, რომ ნაშრომზე მუშაობისას, არსებულ ტექსტურ რედაქტორებში, საკმაოდ რთულია სასურველი სიმბოლოების მოძებნა. სწორედ ამიტომ, ისეთ სამეცნიერო მიმართულებებში, სადაც ხშირადაა გამოყენებული სიმბოლოები, ჩანიშნული იქნა ისინი. მაგალითად, მათემატიკურ სიმბოლოებში ჭარბობს კალკულაციის, ტოლობის და სხვა მსგავსი სიმბოლოები, ხოლო ლოგიკაში ხშირია - თანაკვეთა, გაერთიანება და ა.შ.

ამგერად დავუბრუნდეთ ფორმებს და წარმოვადგინოთ იმის პროცენტული მაჩვენებელი, თუ რამდენადაა თითოეული ფორმა გამოყენებული განსხვავებულ სამეცნიერო თემებში (ნახ. 4.9, 4.10).

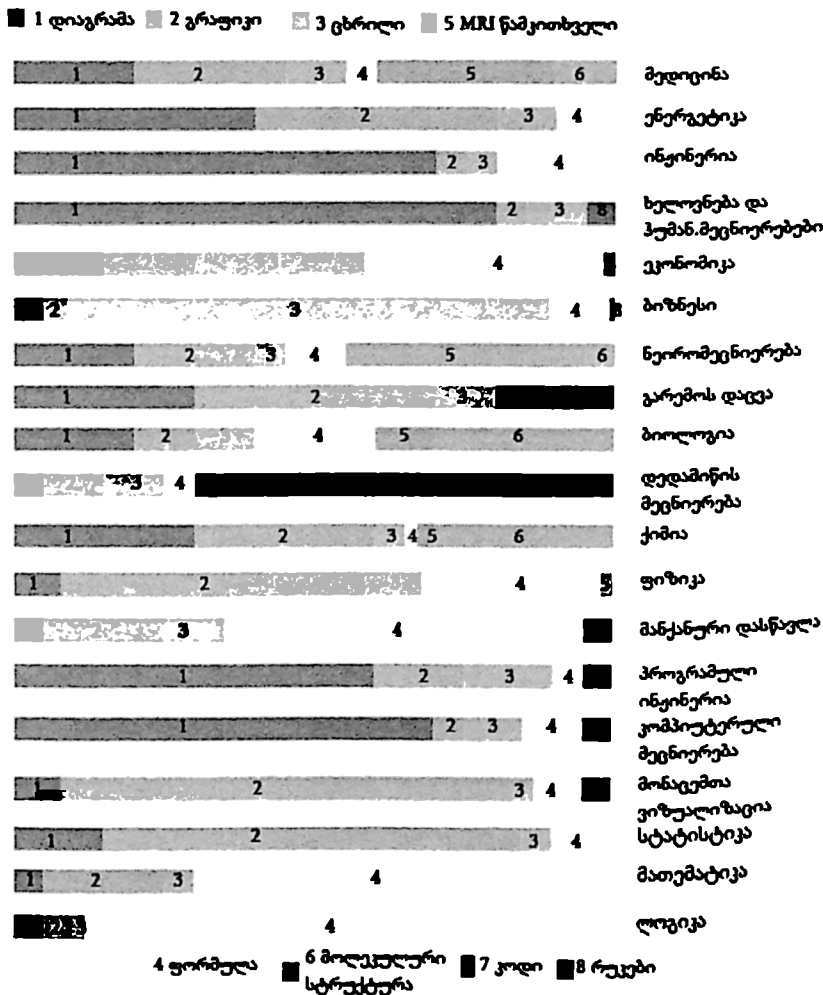
როგორც ამ ნახაზებიდან ჩანს, სხვადასხვა კატეგორიის სამეცნიერო ნაშრომებში გამოყენებული ფუნქციონალი საკმაოდ მრავალფეროვანია. თუ განვიხილავთ თითოეული მიმართულების გამოყენებულ ფორმებს, აღმოვაჩენთ, რომ სხვადასხვა ტიპის ნაშრომში სხვადასხვა ფორმებია პრიორიტეტული [83].

მაგალითად, ლოგიკასა და მათემატიკაში, ფორმების უმრავლესობა ფორმულებია, როდესაც სტატისტიკასა და მონაცემთა მეცნიერებაში გრაფიკები ჭარბობს, ხოლო თეორიულ კომპიუტერულ მეცნიერებაში, ხელოვნებასა და ჰუმანიტარულ მეცნიერებებში დიაგრამები. ასევე აღსანიშნავია, რომ მონაცემების მიხედვით, მნიშვნელოვანია მოლვეკუალური სტრუქტურების ამგები რუკები, პროგრამული კოდის ფორმატირება და სამედიცინო ფაილების წამკითხველი ფუნქციონალი.

მიღებული მონაცემების მიხედვით გამოკვლეული იქნა ფუნქციონალის დამატების საშუალებები და გზები და ერთმანეთს შედარდა მომხმარებლის ინტერფეისები, თითოეული კატეგორიის სამეცნიერო ნაშრომის შემქმნელთათვის. არსებული პროტოტიპი შეგიძლიათ იხილოთ შემდეგ პარაგრაფებში.

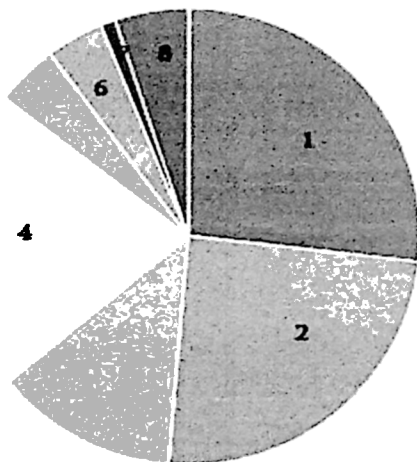
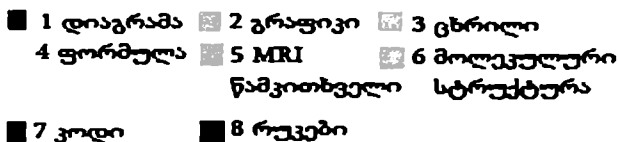


ნახ. 4.9. გამოყენებული ფორმების მაჩვენებელი თითოეული სამეცნიერო მიმართულებისთვის



ნახ. 4.10. ერთიანი გამოყენებული ფორმების მაჩვენებელი თითოეული სამეცნიერო მიმართულებისთვის

რაც შეეხება, კონკრეტული ფუნქციის მნიშვნელობას, ზოგადად სამეცნიერო ნაშრომის შექმნის პროცესში, პროტოტიპის ჩამოყალიბებამდე აუცილებელია იმის ცოდნა, თუ რამდენადაა საჭირო რომელიმე კონკრეტული ფუნქციონალის დამატება და მასში რესურსების დახარჯვა. ამისთვის წარმოვადგინოთ დიაგრამა, სადაც ჩანს ჯამურად რა რაოდენობის ფორმები გამოყენებული ყველა ტიპის სამეცნიერო ნაშრომებში (ნახ. 4.11).



ნახ. 4.11. სამეცნიერო ნაშრომებში ჯამურად გამოყენებული ფორმები

როგორც ნახაზიდან ჩანს, გვაქვს სამი უმნიშვნელოვანესი ფორმების ასაგები ფუნქცია. სამეცნიერო ნაშრომებში, ყველაზე

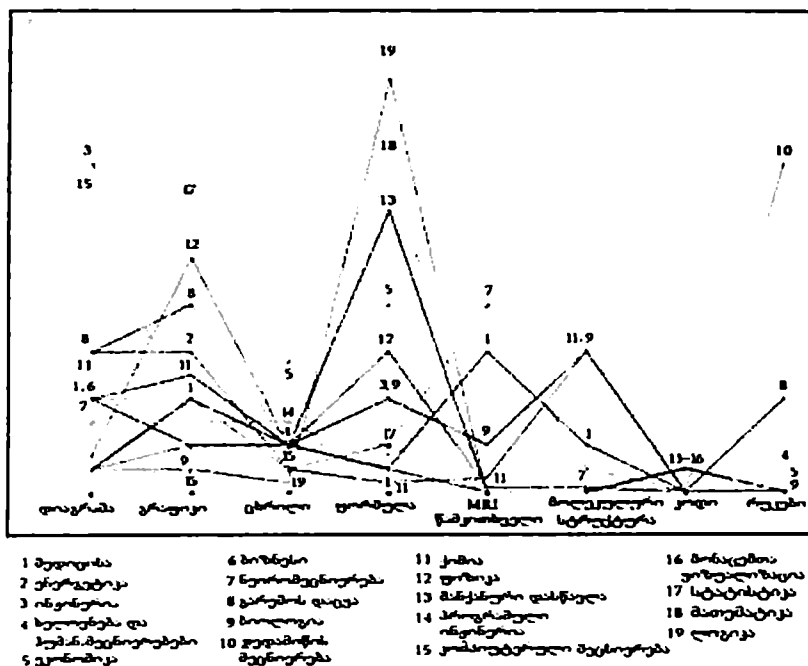
ხშირად დიაგრამებია გამოყენებული. ამასთან ერთად და ჩვენი სხვა კვლევებიდან გამომდინარეც, პროექტში დიაგრამების აგების ფუნქციონალური ერთ-ერთი ყველაზე მნიშვნელოვანი აღმოჩნდა. აგრეთვე, ძალზე მნიშვნელოვანია გრაფიკული დიაგრამების (გრაფიკების) აგებაც.

კიდევ ერთხელ უნდა აღინიშნოს ფაქტი, რომ დღევანდელ ტექსტურ რედაქტორებში, დიაგრამების და გრაფიკული დიაგრამების აგების ფუნქციონალური საკმაოდ მწირია და განვითარებას მოითხოვს. აღნიშნულ ფუნქციებთან ერთად მნიშვნელოვანია ფორმულებისა და ცხრილების აგების ფუნქციონალიც (ნახ.4.12).

დიაგრამები (1)	გრაფიკები (2)
4 ხელოვნება და ჰუმან. მეცნიერებები	16 მონაცემთა ვიზუალიზაცია
3 ინჟინერია	17 სტატისტიკა
15 კომპიუტერული მეცნიერება	12 ფიზიკა
14 პროგრამული ინჟინერია	8 გარემოს დაცვა
2 ენერგეტიკა	2 ენერგეტიკა
და ა.შ.	და ა.შ.
ცხრილები (3)	ფორმულები (4)
6 ბიზნესი	19 ლოგიკა
5 ეკონომიკა	18 მათემატიკა
4 ხელოვნება და ჰუმან. მეცნიერებები	13 მანქანური დასწავლა
14 პროგრამული ინჟინერია	5 ეკონომიკა
15 კომპიუტერული მეცნიერება	12 ფიზიკა
და ა.შ.	და ა.შ.

ნახ. 4.12. 1-4 ფუნქციონალური გამოყენება სფეროების მიხედვით, მოწესრიგებული გამოყენების სიხშირის კლუბადობით

იმის დასადგენად, თუ რამდენად მნიშვნელოვანია თითოეული გასაავითარებელი ფუნქცია, ყოველი კატეგორიის სამეცნიერო ნაშრომისათვის, წარმოვადგინოთ 4.13 ნახაზი.



ნახ. 4.13. ფორმების გამოყენების მაჩვენებელი თითოეული სამეცნიერო კატეგორიის ნაშრომისთვის

როგორც ნახაზიდან ჩანს, სამეცნიერო ნაშრომთა კატეგორიის უმეტესობას ყველაზე ხშირად გამოყენებული ფუნქციონალი 60%-ის ქვემოთაა განლაგებული, რაც მეტყველებს იმაზე, რომ ამ კატეგორიებისთვის, მნიშვნელოვანია არა მხოლოდ რომელიმე ერთი სახეობის ფუნქცია, არამედ რამდენიმე მათგანი.

მოცემულ თავში გაკეთებული კვლევების შედეგების შესაბამისად, მოხდა ტექსტური ინფორმაციის შექმნის პირველი პროტოტიპის აგება, რომელსაც დეტალურად განვიხილავთ მომდევნო თავში.

თავი 5

საგამომცემლო სფეროს ახალი პროგრამული პროდუქტის პროტოტიპის შექმნა

5.1. ტექსტური ინფორმაციის შექმნა	144
5.2. ტექსტის წამკითხველი	145
5.3. ახალი პროდუქტის შექმნის სისტემის პროტოტიპი	146
5.4. უსაფრთხოება	151

5.1. ტექსტური ინფორმაციის შექმნა

ტექსტურ დოკუმენტებთან სამუშაოდ დღეისათვის არაერთი პროგრამული პროდუქტი არსებობს. მათგან ყველაზე ფართოდ გამოყენებადი და პოპულარული პროგრამაა Ms Word. როგორც უკვე აღვნიშნეთ, მაიკროსოფტის ეს პროდუქტი ერთდროულად განკუთვნილია ყველანაირ ტექსტურ დოკუმენტებთან სამუშაოდ და არაა სპეციალიზებული, მაგალითად, სამეცნიერო ნაშრომების შექმნისა თუ თარგმნისათვის. შესაბამისად, სისტემა ამ მოთხოვნების დასაკმაყოფილებლად საჭირო არაერთი ფუნქციის ნაკლებობას განიცდის.

კვლევის დაწყებამდე, იდეის ვალიდაციის მიზნით ჩატარებული კვლევების მიხედვით, შეგვიძლია ვთქვათ, რომ დღეისათვის პროგრამული ინდუსტრიის ბაზარზე არ გვხვდება ისეთი სისტემა, რომელიც სხვადასხვა კატეგორიის სამეცნიერო ნაშრომის შექმნისთვის იყოს გათვალისწინებული, და ამ კატეგორიაზე მომუშავე მეცნიერებს, მათზე მორგებულ ინტერფეისსა და ფუნქციონალს სთავაზობდეს.

ამასთანავე არსებული სისტემები ვერ ეხმარება ავტორებს მათი წიგნის ან სტატიის შემოწმებისა და მისი დასრულების ავტომატიზაციაში, თანაავტორებთან ურთიერთობასა და ჩასწორებების ისტორიის შენახვაში და ა.შ.

არ არსებობს სისტემა, რომელიც სამეცნიერო ნაშრომის დაწყებისას პირველი წინადადებების შეყვანის შემდეგ (რაც, როგორც წესი, რეზიუმეა), მანქანური დასწავლის მეთოდებით მიღებული მოდელის საშუალებით გამოიცნობს ნაშრომის თემას და

შესთავაზებს ავტორ-მწერალს სპეციალურად მასზე მორგებულ იტერფეისს.

მიუხედავად ზემოხსენებული სისტემების არარსებობისა, პროგრამები, მაგალითად, Microsoft Word საკმაოდ განვითარებულია და მასზე მომხმარებელთა მიჩვევადობის მაღალი მაჩვენებლით გამოირჩევა. აქედან გამომდინარე, გათვალისწინებულ უნდა იქნას ისეთი ინტერფეისი, რომ მომხმარებელს მასზე მუშაობისას მიჩვევადობის გამო არ შეექმნას პრობლემა.

5.2. ტექსტის წამკითხველი

მსოფლიოში ყველაზე გავრცელებული და, ალბათ, წიგნის საუკეთესო წამკითხველი პროგრამა Amazon კომპანიის აპლიკაციაა Amazon-Kindle-ს სახით. იგი გამოირჩევა სიმარტივითა და საჭირო ფუნქციონალით. დღეისათვის, ელექტრონული დოკუმენტების წამკითხველი სისტემები განვითარებული და თითქმის უნაკლოა.

ჩვენ პროექტში, მწერლისა და მკითხველის გაერთიანების და მკითხველისთვის საგამომცემლო მარკეტინგის განვითარების გამო, ტექსტის წამკითხველი ფუნქციის ინტეგრაცია, ძალიან მნიშვნელოვანია. ასეთ შემთხვევაში, საჭიროა გათვალისწინებულ იქნას დღევანდელი სისტემები და მოხდეს მაქსიმალურად კომფორტული და მარტივი წამკითხველის შექმნა ან ინტეგრაცია.

ერთ-ერთი უპირატესობა, რაც ჩვენ პროექტს გააჩნია, არის ის, რომ სისტემაში შექმნილი დოკუმენტები ვებ-ფორმატითაა წარმოდგენილი და მოწყობილობის ზომის შესაბამისად კარგ მოქნილობას გვაძლევს.

აღსანიშნავია, რომ არსებობს არაერთი ღია კოდის მქონე ბიბლიოთეკა, რომელიც წიგნის წამკითხველი სისტემისთვისაა შექმნილი. სისტემაში საჭიროა გამოყენებულ იქნას ასეთი ღია

კოდის მქონე ბიბლიოთეკა და გამოირჩეოდეს ისეთი ფუნქციონალით, რომელიც ყველაზე მნიშვნელოვანი და საჭიროა.

ხაზი გავუსვათ რამდენიმე ისეთ ფუნქციას, რომლებიც წიგნის წასაკითხი აპლიკაციებიდან მნიშვნელოვანია.

ერთ-ერთ ფუნქციად შეიძლება მივიჩნიოთ სისტემაში სიტყვათა ლექსიკონის ინტეგრაცია.

ასევე მნიშვნელოვანია ფუნქციონალი, რომლის მეშვეობითაც მკითხველი აყენებს დიზაინს წიგნის კითხვისათვის. ასეთი დიზაინის ელემენტები შეიძლება იყოს ეკრანის სიკაშკაშე, ტექსტის ფონტი, ფერი, ზომა, ჩარჩოს ზომა და სხვ.

5.3. ახალი პროდუქტის შექმნის სისტემის პროტოტიპი

ჩატარებული კვლევების საფუძველზე, ჩამოყალიბებულ იქნა დასაპროექტებელი სისტემის პროტოტიპი, რომელიც სხვადასხვა სფეროში მომუშავე პირებს, სპეციალურად მათთვის მორგებულ მომხმარებლის ინტერფეისსა და ფუნქციებს შესთავაზებს.

კვლევებიდან გამომდინარე, სისტემაში საჭირო იყო დიაგრამების, გრაფიკების, მოლეკულური სტრუქტურების, რუკების და სხვა მნიშვნელოვანი ფუნქციების არსებობა. ამისთვის, მოძიებულ იქნა ღია კოდის მქონე ბიბლიოთეკები, რომლებიც მოცემული დავალების შესრულებაში დაგვხმარა.

გრაფიკებისთვის აღსანიშნავია Chart.js ბიბლიოთეკა. ის ძალზე მრავალფუნქციურია და მისი ინტეგრაცია სისტემაში საკმაოდ მარტივი აღმოჩნდა. ტექსტურ რედაქტორში მისმა არსებობამ, საგრძნობლად გაამარტივა მუშაობის პროცესი და შედეგების წარდგენის საშუალება [84].

რაც შეეხება *დიაგრამების აგებას*, ამ შემთხვევაში არჩევანი გაკეთდა, ორ ბიბლიოთეკაზე. ეს ინტეგრირებული ბიბლიოთეკებია

gojs და bpmn-js [85, 86]. აგრეთვე საინტერესოა draw2d, რომლის ინტეგრაციაც პოტენციურად განიხილება [87].

უნდა აღინიშნოს ფაქტი, რომ დღევანდელ ტექსტურ რედაქტორებს არ აქვს დიაგრამების აგების კომფორტული საშუალება და გრაფიკების ასაგებად იყენებენ დამხმარე სისტემებს, რაც მათთან მუშაობას უფრო რთულს ხდის. აღწერილი ბიბლიოთეკების ინტეგრაციის შემდეგ კი მათი გამოყენების ხარჯზე ნაშრომის შექმნის პროცესი ბევრად უფრო კომფორტული, მრავალმხრივი და სწრაფია.

ზემოაღნიშნული ფუნქციების ინტეგრაციასთან ერთად, საჭიროა *მათემატიკური ფორმულების* კომფორტული შეტანის შესაძლებლობაც. ამ დავალებას mathjax ბიბლიოთეკა თავს არაჩვეულებრივად ართმევს [101]. მისი საშუალებით, ფორმულის შეყვანა ბევრად უფრო სწრაფი და მოხერხებული ხდება, ვიდრე ჩვეულებრივ რედაქტორებში.

ამის გარდა, სისტემაში ინტეგრირებულია *მათემატიკური ამოცანების* გაშვების საშუალებაც. ამის შესაძლებლობას ბიბლიოთეკა math.js გვძლევს [102].

გეოგრაფიული სფეროს ნაშრომებისთვის *რუკების* ინტეგრაციის მიზნით, google-ის რუკები უკონკურენტოა, მისი API-ის მეშვეობით შესაძლებელია რუკებზე ანოტაციების გაკეთება, ადგილების მონიშვნა და ა.შ.

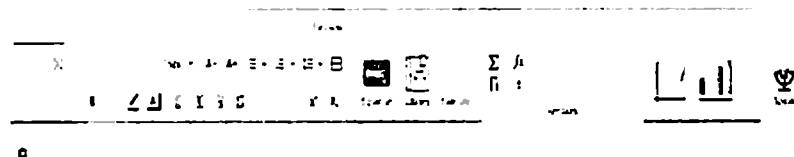
რაც შეეხება *მოლეკულურ სტრუქტურებს*, ამ შემთხვევაში გამოყენებული იქნა smilesDrawer ბიბლიოთეკა [88]. მისი გამოყენების შემთხვევაშიც, სტრუქტურის ხაზვა ძალზე სწრაფი და კომფორტულია.

სამედიცინო ფაილებთან მუშაობისთვის კი NIFTI-Reader-JS ბიბლიოთეკის ინტეგრაცია გახდა საჭირო [89].

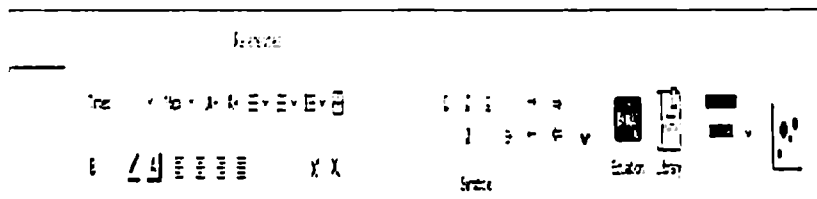
ზემოაღწერილი ბიბლიოთეკების ინტეგრაციის შედეგად, მიღებულ იქნა სისტემა იმ ფუნქციონალთა ნაკრებით, რომელიც მიზნად გვეყონდა დასახული. შემდეგი ეტაპი სხვადასხვა მიმართულების სამეცნიერო ნაშრომთათვის სპეციალურად მორგებული ინტერფეისის შექმნაა.

წარმოვადგინოთ რამდენიმე მაგალითი ჩვენი პროტოტიპიდან იმის შესახებ თუ როგორ გამოიყურება ეს ინტერფეისი.

5.1 და 5.2 ნახაზებზე ნაჩვენებია რედექტორები მათემატიკისა და ლოგიკის სფეროს ნაშრომთა ასაგებად.



ნახ. 5.1. მომხმარებლის ინტერფეისი - კატეგორია მათემატიკური ნაშრომისთვის



ნახ. 5.2. მომხმარებლის ინტერფეისი - კატეგორია ლოგიკის ნაშრომისთვის

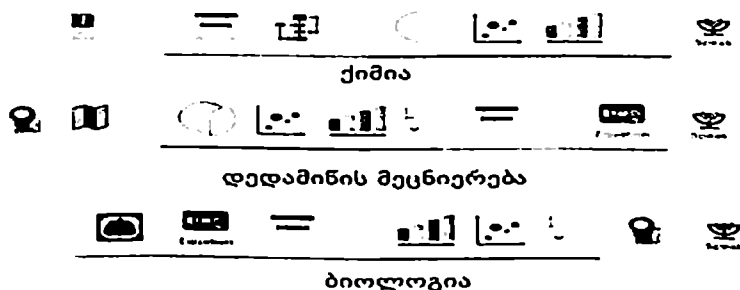
როგორც ნახაზებიდან ჩანს, მომხმარებლის ინსტრუმენტთა ნაწილის მარცხნივ, ყველა ტიპის ნაშრომისთვის საჭირო ინსტრუმენტებია განლაგებული. აღნიშნული ნაწილი სტატიკურია და ის უცვლელი რჩება ყველა ტიპის სამეცნიერო ნაშრომის შესრულებისას. თუმცა მარჯვენა მხარეს ვხედავთ განსხვავებული ფუნქციების შეთავაზებას.

პირველ რიგში ხაზი უნდა გაესვას იმას, რომ ორივე სამეცნიერო კატეგორიის შემთხვევაში, საკმაოდ დიდი ადგილი უკავია სპეციალურ სიმბოლოებს. როგორც ხედავთ, ეს სიმბოლოები ორივე შემთხვევისთვის განსხვავებული და თემის შესაბამისია.

მათემატიკაში ვხედავთ ჯამის, ფესვისა და სხვა სიმბოლოებს, მაშინ, როცა ლოგიკაში განსხვავებული, ამ თემისათვის რელევანტური სიმბოლოებია ნაჩვენები.

ყოველი ფუნქციის ჩვენების თანმიმდევრობა და ინტერფეისში მის მიერ დაკავებული ადგილის პროცენტული მაჩვენებელი დაკავშირებულია პროექტში ჩატარებულ კვლევებთან.

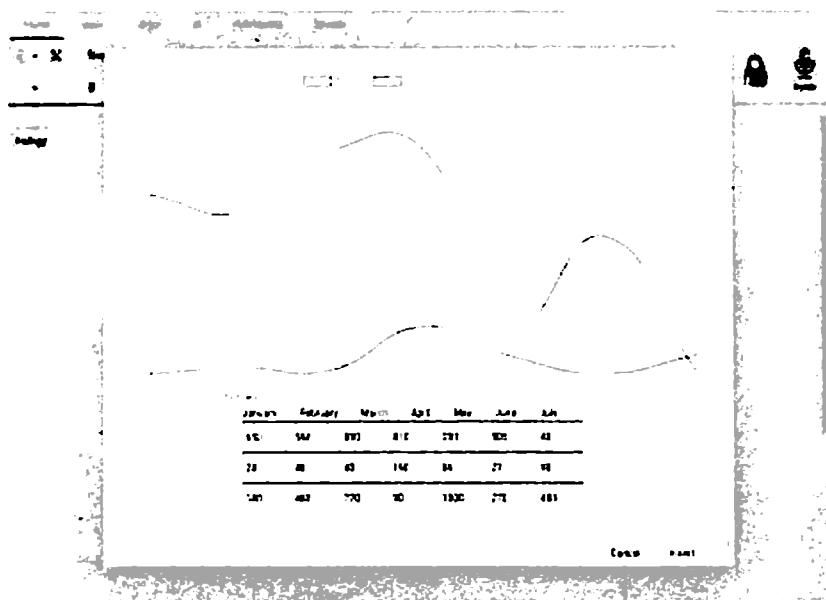
5.3 ნახაზზე წარმოდგენილია სხვა მაგალითები სფეროზე მორგებული ინტერფეისისათვის.



ნახ. 5.3. მომხმარებლის ცვალებადი ინტერფეისის შეთავაზებები სხვა სფეროებისთვის

როგორც სურათებიდან ჩანს, ინსტრუმენტებში გვხვდება “Speak” ლილაკი, რომელიც speech-to-text ტექნოლოგიას იყენებს. რაც შეეხება ნაშრომში კონკრეტული ფორმების ჩასმას, კომფორტულად მუშაობისთვის, მათი დამუშავება დიალოგურ ფანჯარაში შესაძლებელი.

ამის შემდეგ კი „ჩამატების“ ლილაკის გამოყენებით ის ნაშრომში დაემატება (ნახ. 5.4).



ნახ. 5.4. Cartas.js მუშაობის პროცესი დიალოგურ ფანჯარაში

საბოლოოდ, 5.5 ნახაზზე წარმოდგენილია სპეციალურად შექმნილი ინტერფეისი მთარგმნელთათვის.

ფუნქციონირებით. ტექსტური ინფორმაციის შექმნის, მისი გამოცემებისა და მართვის უამრავი შესაძლებლობა არსებობს გვერდზე უამრავი წიგნების ბიბლიოთეკა თუ ელექტრონული წიგნბანია. მათი წარმოადგენელი და დამუშავების შექმნის ხელშეწყობით სპეციალური პლატფორმები თუბრა, საგამომცემლო სფეროში მათი უამრავი პრობლემაა. მაგალითად, ამ არსებობს ტექსტური ინფორმაციის შექმნის, მისი გამოცემებისა და გავრცელების ერთიანი სისტემა ამ არსებობს ინფორმაციის შექმნის იდენტურ საშუალება, რომელიც მოერგება სხვადასხვა ტიპის სამეცნიერო მიმართულებებზე მომუშავე პირებს. მათთვის საინტერესო ფუნქციონირებას და ინტერფეისის შეთავაზებით

Nowadays, there are many opportunities to create, publish and read textual information. We have a lot of book libraries or e-shops, their readers and special applications for creating documents. However, in the publishing field

სადოქტორო პროცესზე შემოხილს ფაქსი, ჩვენი მთავარი მიზანია. მივიღოთ ისეთი შედეგი, რომელიც სულს შევსოვს, შევსოვრებს, შევსოვრებსა და მთავრდება. როგორც ტექსტური ინფორმაციის შექმნაში, აგრეთვე მარკეტინგის პროცესში, ამასთან ურთოდ, გვინდა, სული შევსოვით, სხვადასხვა ტიპის სამეცნიერო თუ არასამეცნიერო ლიტერატურით დაინტერესებულ პირებს, იმ ნაშრომების მოძიებაში, რაც მათთვის საინტერესოა.

სხვადასხვა ტიპის სამეცნიერო ნაშრომებში მომუშავე პირებისა და მათი მეთვლელების მიხედვით მოთხოვნების ურთი პროცესში, პროდუქტის საშუალებით დაკმაყოფილება. სადაც აგრეთვე იწეება ამ მომხმარებლების დაკავშირების ისეთი გზები, რეკონსტრუქცია, წიგნის გამოცემა, მათთვის

ნახ. 5.5. სპეციალურად მთარგმნელთათვის შექმნილი მომხმარებლის ინტერფეისი

5.4. უსაფრთხოება

მომხმარებელთა უსაფრთხოება ნებისმიერი სისტემის უმნიშვნელოვანესი ნაწილია და მისი ვებ აპლიკაციაში უზრუნველყოფის მრავალი საშუალება არსებობს. თუმცა, სისტემაში უსაფრთხოების მოდელის აგებისას, განსაკუთრებით საყურადღებოა, რომ ამ საშუალებებიდან საუკეთესო იქნას არჩეული.

ვებ სამყაროში, მომხმარებლის უსაფრთხოების უზრუნველყოფის სამი ძირითადი გზა არსებობს, ესენია:

- 1) მომხმარებლის რეგისტრაცია სისტემაში;

- 2) მომხმარებლის იდენტიფიკაცია და
- 3) ავტორიზაცია.

სისტემაში რეგისტრაცია მომხმარებლის პირადი ინფორმაციისა და იდენტიფიკაციის დამატებას გულისხმობს აპლიკაციის მონაცემთა ბაზებში.

იდენტიფიკაცია პროცესია, როდესაც მომხმარებელი სისტემაში წვდომის მოპოვებას საკუთარი მონაცემების (ელ-ფოსტა და პაროლი) შეყვანის შედეგად ახდენს.

ავტორიზაცია კი, შეიძლება ითქვას, უსაფრთხოების პროტოკოლია, რომელიც მომხმარებლის მიერ მოთხოვნილი მონაცემებისთვის, ამავე მომხმარებლის წვდომის უფლებას ამოწმებს.

ჩვენ პროექტში მოდელირებული სისტემის უსაფრთხოება Node.js ბიბლიოთეკის მეშვეობით იქნა უზრუნველყოფილი.

სისტემაში რეგისტრაცია ელ-ფოსტისა და პაროლის მეშვეობითაა შესაძლებელი. მას შემდეგ, რაც სისტემა ვალიდაციას გაატარებს სწორი ელექტრონული მისამართისა და დამაკმაყოფილებელი პაროლის მნიშვნელობებს, პირველ რიგში ის მოახდენს მომხმარებლის შერჩეული პაროლის დაშიფვრას, ისე, რომ, ერთის მხრივ, შეუძლებელი იყოს მისთვის შიფრის მოხსნა, ხოლო მეორეს მხრივ, იგივე პაროლის ხელახალი დაშიფვრის შემთხვევაში, ყოველთვის ერთიდაიგივე შედეგი მივიღოთ.

ამ შიფრაციის დროს კი გვჭირდება დამატებითი საიდუმლო კოდი, რომელიც მხოლოდ ჩვენ ბაზებში იქნება შენახული. მეტი უსაფრთხოებისთვის, ამ საიდუმლო კოდს ყოველი მომხმარებლისთვის, მათ სისტემაში რეგისტრაციის დროს ვაგენერირებთ.

ამ კოდის გენერირებისა და მომხმარებლის რეგისტრაციის ამსახველი კოდი იხილეთ 5.6 ნახაზზე.

სისტემაში რეგისტრაციის შემდეგ, უსაფრთხოების შემდეგი ეტაპი აუთენტიფიკაციაა. სპეციალურ ფორმაში, ელ-ფოსტისა და

პაროლის შეყვანის შემდეგ, სისტემა მონაცემთა ბაზიდან მომხმარებლის საიდუმლო კოდს ამოიღებს და მას პაროლის ხელახალი შიფრაციისათვის გამოიყენებს.

```
bcrypt.hash(userPassword, salt, 12)
  .then(password => {
    return User.create({email: email,...})
  })
  .then(async () => {
    const userSecret = new UserSecret({
      secret: makeSecret(1000000, 30),
      refreshToken: makeSecret(1000000, 30),
      userId: result.id
    })
    userSecret.save()
    .catch(err => {...})
    userId = result.id
  }).then(async () => {
    return UserDetails.create({email: {
      firstName,
      lastName,
      userId
    }}
  })
```

ნახ. 5.6. მომხმარებლის საიდუმლო კოდის გენერირება და სისტემაში დამატება

ამის შემდეგ, მოხდება ახლად დაშიფრული პაროლის მონაცემთა ბაზაში შენახულ შიფრთან შედარება. თუ შედეგები ერთმანეთის ტოლია, მომხმარებლის მიერ მითითებული მონაცემები სწორია და ის აუთენტიფიკაციას გაივლის. ამისთვის, მომხმარებლისთვის სპეციალურად დავაგენერირებთ ორ jwt-ს (Json ვებ ტოკენი) [90, 91], ერთი ავტორიზაციისთვის (Auth token), მეორე კი ამ ტოკენის განახლებისთვისაა საჭირო (Refresh token).

უნდა აღინიშნოს, რომ jwt-ს აქვს ამოწურვადი ვადა, შესაბამისად, მისი გენერირებისას აუთენტიფიკაციის ტოკენის ვადა ძალზე მცირეა, ხოლო ავტორიზაციისა შედარებით ხანგრძლივი.

გენერირებული ტოკენები კლიენტს ეგზავნება და ბრაუზერში "httpOnly cookie" სახით ინახება. რაც მონაცემის შენახვის ყველაზე უსაფრთხო გზაა, რადგან მისი ნახვა ვებსაიტის javascript ფაილებიდან შეუძლებელია [92], შესაბამისად, ტოკენების ამ ადგილას შენახვის დროს, ბრაუზერის აპლიკაციები ვერ შეძლებს მის მოპოვებას.

ჩვენ პროექტში კლიენტის მხრიდან სერვერისკენ ყოველი გაგზავნილი მოთხოვნა აუთენტიფიკაციის ტოკენს შეიცავს. თუ ტოკენს ვადა არ აქვს გასული, სერვერი გადადის ავტორიზაციის შემდეგ ეტაპზე, წინააღმდეგ შემთხვევაში კი აგზავნის შეტყობინებას ვადის გასვლასთან დაკავშირებით. ამის შემდეგ, კლიენტი სერვერს უკვე განახლების ტოკენს გაუგზავნის და მისი ვალიდურობის შემთხვევაში, ავტორიზაციის შემდეგ ეტაპზეც გადავა და ახალ ავტორიზაციის ტოკენსაც მიიღებს.

რაც შეეხება ავტორიზაციის ეტაპებს, ეს nodejs-ის შუალედური კოდებით ხდება (Middleware). ყოველ მოთხოვნაზე დარეგისტრირებულია მისი http მისამართი, მისამართთან ერთად კი მითითებულია მასზე წვდომის შეზღუდვა, რომელიმე როლის მქონე მომხმარებლებისათვის.

მაგალითად, როდესაც მომხმარებელი ცდილობს ცვლილება შეიტანოს მის ნაშრომში, ყოველი ასეთი ცვლილების დროს, სერვერულ აპლიკაციას ამის შესახებ ინფორმაცია ეგზავნება.

ამ დროს სერვერი ამოწმებს ინფორმაციას, რომ მოთხოვნა, სწორედ ნაშრომის შემქმნელის მიერაა გაგზავნილი და არა რომელიმე სხვა პირის მიერ.

დასკვნა

XXI საუკუნე გახდა ახალი ცივილიზაციის ჩამოყალიბების ზღურბლი, რომელიც პოსტინდუსტრიული საზოგადოებიდან საინფორმაციო საზოგადოებაზე გადასვლას გულისხმობს, ციფრული ტექნოლოგიების, ხელოვნური ინტელექტის, პროგრამული ინჟინერიის, ინტერდისციპლინური და დისტანციური სწავლების საფუძველზე. საინფორმაციო საზოგადოების ფორმირება და ინფორმატიკის დიდაქტიკის ინტენსიფიკაცია ქვეყნის მდგრადი განვითარების და ეროვნული, მაღალზნეობრივი მენტალიტეტის ჩამყალიბების ერთ-ერთი მთავარი კრიტერიუმია;

მნიშვნელოვანი წვლილი ამ საქმეში თანამედროვე მედია ტექნოლოგიებმა და საინფორმაციო კომპიუტერულმა სისტემებმა უნდა შეიტანოს. ამიტომაც განსაკუთრებული ყურადღება უნდა მიექცეს მათი სიმბიოზის - მედიაინფორმატიკის და ინფორმატიკის დიდაქტიკის საგანმანათლებლო-სამეცნიერო მიმართულებებს. შეიძლება თამამად ითქვას, რომ საინფორმაციო საზოგადოება მაღალგანვითარებული ინფორმაციული და კომუნიკაციური ტექნოლოგიების (ICT) პირმშოა;

საგამომცემლო ინდუსტრია – ის სფეროა, სადაც იქმნება ლიტერატურული და სამეცნიერო წიგნები და ნაშრომები, გადის რედაქტირებას, ხარისხობრივ სრულყოფას და გამოიცემა ტრადიციული ან ელექტრონული პროდუქციის სახით, გადაეცემა ბიბლიოთეკებს, ფიზიკურ და ონლაინ მაღაზიებს, დაინტერესებულ მკითხველს, ანუ ეს პროდუქტი რჩება ისატორიას, თანამედროვე საზოგადოებას და მომავალ თაობებს;

როგორც ყველა კარგ საქმეს, საგამომცემლო ბიზნეს-პროცესებსაც შეიძლება ჰქონდეს სენსიტიური (წყლულოვანი) ადგილები, რაც სიფრთხილესა და „მკურნალობის“ ეტაპის არსებობას ითხოვს ასეთი ანომალიების პრევენციის მიზნით. ლიტერატურულ და სამეცნიერო შედეგებთან, ეთიკური ნორმებით შესრულებულ

ნაშრომებთან ერთად, არსებობს საფრთხეებიც, მაგალითად, პლაგიატის, ასლების უკანონო გადაღება, გავრცელება და ა.შ. ასეთი პროცესები ექვემდებარება ეთიკურ და იურიდიულ პასუხისმგებლობას. ამიტომ, წესიერი ავტორებისათვის მნიშვნელოვანია მათი ახალი ნაშრომის დაწყების ეტაპზე „ინოვაციურმა სისტემამ“, რომელიც აშკარად ხელოვნური ინტელექტისა და პროგრამული ინჟინერიის ერთობლივი პროდუქტია, ხელი შეუწყოს მწერალ-ავტორებს – სასტარტო ეტაპზე შესაბამისი „თემის ამოცნობის“ საფუძველზე, მიაწოდოს აუცილებელი და საკმარისი მოცულობის ინფორმაცია (ინტერნეტში არსებული პირველწყაროები);

მანქანური დასწავლის ალგორითმების საფუძველზე შესაძლებელია ტექსტური ინფორმაციის დამუშავების ინოვაციური მოდელების აგება, მაგალითად, ნაშრომთა რეზიუმეების და სრული ტექსტების ლინგვისტური ანალიზით. მათი სინქრონულად გამოყენების შემთხვევაშიც, რომელიმე სამეცნიერო ნაშრომით დაინტერესებული ადამიანისთვის, როგორც თემატურად, ასავე მეთოდოლოგიურად საინტერესო ნაშრომების შეთავაზება ხდება შესაძლებელი;

დღეისათვის არსებულ პოპულარულ ტექსტურ რედაქტორებს სამეცნიერო ნაშრომის შექმნისათვის მნიშვნელოვანი ფუნქციონალი არ აქვს. გამოკთხვის შედეგად, გაანალიზებულ იქნა ინფორმაცია ამ ფუნქციონალის საჭიროებებზე, შემდეგ კი გამოკვლეულ იქნა მათი შესრულების მეთოდები და ავტომატიზაციის მოდელები;

დღევანდელი ტექსტური რედაქტორები, არაა მორგებული კონკრეტულ სამეცნიერო სფეროზე მომუშავე პირების საჭიროებებზე. გამოკვლეული იქნა სხვადასხვა კატეგორიის მქონე სამეცნიერო ნაშრომებში გამოყენებული ფუნქციებისა და ფორმების რაოდენობრივი მაჩვენებლები. შედეგად შექმნილ იქნა პროტოტიპი, რომელიც სხვადასხვა სამეცნიერო სფეროში მომუშავე ტექსტური ინფორმაციის შემქმნელ ადამიანებს, სპეციალურად მათზე მორგებულ სამუშაო გარემოს სთავაზობს.

გამოყენებული ლიტერატურა

1. ჩოგოვაძე გ. საქართველოს ტექნიკური უნივერსიტეტის იუნესკოს კათედრა „ინფორმაციული საზოგადოება“ 15 წლისაა (შესავალი სიტყვა კონფერენციის გახსნისას). სტუ-ს შრ. კრებ. „მას“, N2(26), თბ., გვ. 11-14
2. ჩოგოვაძე გ., ფრანგიშვილი ა., ჯაგოდნიშვილი თ., სურგულაძე გ. (2018). ინფორმაციული საზოგადოება მულტი-დისციპლინური განათლების თანამედროვე გამოწვევა. სტუ-ს შრ. კრებ. „მას“, N2(26), თბ., გვ. 19-24
3. სურგულაძე გ. სტუ-ის UNESCO-ს კათედრის ისტორიული როლი ინფორმაციული საზოგადოების ფორმირების პროცესში. სტუ, შრ.კრ. „მას“, N 1(35). თბ., 2023. გვ. 5-10
4. ჩოგოვაძე გ., ფრანგიშვილი ა., სურგულაძე გ., პეტრიაშვილი ლ., ამილახვარი ნ. „ინფორმატიკის“ – მულტიდისციპლინური მეცნიერების თანამედროვე გამოწვევები და საქართველოს ტექნიკური უნივერსიტეტის სადოქტორო პროგრამის სტრატეგიული მიზნები. სტუ-ს შრ.კრ., „მას“, No 1(30), 2020. გვ. 7-10
5. ჩოგოვაძე გ., სურგულაძე გ., თოფურია ნ., ხარიტონაშვილი მ. სასწავლო პროცესის სრულყოფა ინტერდისციპლინური დიდაქტიკის ინტენსიფიკაციის საფუძველზე. სტუ-ს შრ.კრ., „მას“, No 1(28), 2019. გვ. 7-16
6. World Summit on the Information Society Forum 2023. Internet resource: <https://www.itu.int/net4/wsis/forum/2023/en>
7. Information society. Internet resource: https://en.wikipedia.org/wiki/Information_society#cite_note-19 (20.09.23)
8. Publishing Industry: The importance of marketing for publishing houses. Internet resource: <https://www.doxee.com/blog/digital-marketing/the-importance-of-marketing-for-publishing-industry/>

9. Publication. Internet resource: <https://en.wikipedia.org/wiki/Publication> (16.09.23)

10. Blythe S. A guide to jobs in the publishing industry. Australia. Internet resource: <https://au.prosple.com/career-planning/a-guide-to-jobs-in-the-publishing-industry> (2.08.23)

11. Publisher Assotiations, UK publishing. Internet resource: <https://www.publishers.org.uk/about-publishing/how-publishing-works/> (11.08.23)

12. Publishing Industry. 2018. Internet resource: <https://www.encyclopedia.com/literature-and-arts/journalism-and-publishing/journalism-and-publishing/publishing-industry> (2.08.23)

13. Crawford S. The origin and development of a concept: the information society. Bulletin of the Medical Library Association. 71 (4): pp. 380–385. Internet resource: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC227258/> (14.09.23)

14. კოტლერი ფ. მარკეტინგის საფუძვლები. თარგმანი. ინგლ. „მაცნე“, თბ., 1993.

15. სურგულაძე გ., ქრისტესიაშვილი ხ., სურგულაძე გიორგი. საწარმოო რესურსების მენეჯმენტის ბიზნესპროცესების მოდელირება და კვლევა. ISBN 978-9941-20-557-6. სტუ. „ტექნიკ. უნივერსიტეტი“, თბ., 2015 -216 გვ.

16. გოგშელიძე დ. კვლევის მიზნები და მეთოდოლოგია „ტექსტური ინფორმაციის შექმნისა და საგამომცემლო მარკეტინგის ერთიანი სისტემის“ დამუშავებისას. სტუ-ს შრ. კრებ. „მას“, N2(26), თბ., ISSN 1512-2174, გვ. 261-264

17. სურგულაძე გ., თურქია ე., ოხანაშვილი მ., სურგულაძე გ. მარკეტინგული პროცესების მართვის ერთი მოდელის შესახებ ფერადი პეტრის ქსელებით- სტუ-ს შრ. კრებ. „მას“, N2(5), 2008. ISSN 1512-2174, გვ. 9_16.

18. გოგშელიძე დ., სურგულაძე გ., თურქია ე. ტექსტური ინფორმაციის შექმნისა და საგამომცემლო მარკეტინგის ერთიანი სისტემის ბიზნესპროცესების მოდელირება. შრომები, მართვის ავტომატიზებული სისტემები. 2018, No 1(25), EISSN 1512-2174, გვ. 41-48

19. თურქია ე. ბიზნეს-პროექტების მართვის ტექნოლოგიური პროცესების ავტომატიზაცია. სტუ. „ტექნიკური უნივერსიტეტი“, თბ., 2010. ISBN 978-9941-14-784-5, - 311 გვ.

20. OA Books Toolkit. Internet resource: <https://www.oabooks-toolkit.org/> (10.08.23)

21. How does the Linear Support Vector Classifier Auto Update integration work? Internet resource: <https://info.orcid.org/ufaqs/how-does-the-crossref-auto-update-integration-work/>

22. ERP for Publishers & Publishing Industry - A Complete Guide. Internet resource: <https://www.deskera.com/blog/erp-for-publishing-industry-publishers/> (23.06.23)

23. Reisig W. Elements of Distributed Algorithms : Modeling and Analysis with Petri Nets. Berlin ; Heidelberg ; New York et al : Springer, 1998

24. CPN Tools. www.daimi.au.dk/CPNTools/. gadamowm. 1.10.08

25. Jensen K., Kristensen M.L., Wells L. Coloured Petri Nets and CPN Tools for Modelling and Validation of Concurrent Systems. University of Aarhus. Denmark. 2007

26. Murphy KP. Naive bayes classifiers. University of British Columbia, 2006 Oct;18(60):1-8.

27. Rish I. An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001, August, Vol. 3, No. 22, pp. 41-46.

28. სურგულაძე გ., კივილაძე გ. შესავალი NoSQL მონაცემთა ბაზებში. ISBN978-9941-0-9642-6. სტუ. „IT-კონსალტინგ ცენტრი“. თბ., 2017. -152 გვ.

29. სურგულაძე გ., პეტრიაშვილი ლ. მონაცემთა საცავის დაპროექტება და აგება ინტერნეტული ბიზნესისათვის. ISBN 978-9941-8-0623-0, მონოგრაფია. სტუ-ს „IT-კონსალტინგ სამეცნიერო ცენტრი“. თბ., 2022, - 200 გვ.

30. Blei D.M., Ng, A.Y., Jordan M.I. Latent Dirichlet Allocation. J. Mach. Learn. 2003, Res., 3, 993--1022

31. Lee D., Seung H. Learning the parts of objects by non-negative matrix factorization. Nature, 1999, 401, 788–791.

32. სურგულაძე გ., თოფურია ნ., ამილახვარი ნ. ინფორმაციული საზოგადოება და ინფორმატიკის დიდაქტიკა. IX საერთაშ. კონფ. „ინტერნეტი და საზოგადოება“ ქუთაისის აკ.წერეთლის სახ. უნივ.-ის 90 წლის იუბილისადმი მიძღ. კონფ. INSO-2023. გვ. 5-9

33. სურგულაძე გ., ბიტარაშვილი მ., ქრისტესიაშვილი ხ. პროგრამული აპლიკაციების დეველოპმენტის საფუძვლები /საკურსო პროექტებისთვის/ (C#, MsAccess, ADO & ASP.NET). ISBN 978-9941-20-251-3. სტუ. „ტექნიკ.უნივერს.“, თბ., 2013 -96 გვ.

34. Bizagi Modeler for Windows (Free download). v.4.0. Internet resource: <https://bizagi-process-modeler.en.softonic.com/> (10.09.23)

35. ბოლხი გ., გოგიჩაიშვილი გ., სურგულაძე გ., პეტრიაშვილი ლ. მართვის ავტომატიზებული სისტემების ობიექტ-ორიენტირებული დაპროექტების და მოდელირების ინსტრუმენტები (MsVisio, WinPepsy, PetNet, CPN). ISBN 99940-56-77-8. სტუ. „ტექნიკ.უნივერს.“, თბ., 2013 -232 გვ.

36. A tool for editing, simulating, and analyzing Colored Petri nets. Internet resource: <https://cpntools.org/> (11.08.23)

37. From CPN Tools via Access/CPN to CPN IDE. Internet resource: <https://cpnide.org/> (11.08.23)

38. Paatero P., & Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. In Environmetrics. Wiley, 1994, Vol. 5, Issue 2, pp. 111–126.

39. Speech-to-Text. Internet resource: <https://cloud.google.com/speech-to-text> (10.05.23)

40. Speech to Text. Internet resource: <https://www.duplichecker.com/speech-to-text.php> (10.05.23)

41. Artificial intelligence and Machine learning. Columbia University, Engineering. USA, New York. Internet resource: <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>

42. Shruti M. AI vs Machine Learning vs Deep Learning: Know the Differences. Internet resource: <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/ai-vs-machine-learning-vs-deep-learning> (7.11.23)

43. სურგულაძე გ. კორპორაციის ავტომატიზებული სამუშაო ადგილების ქსელის აგების ტექნოლოგია (პირველი ქართული ERP სისტემა). ISBN 978-9941-8-5109-4, მონოგრაფია. სტუ-ს „IT-კონსალტინგ სამეცნიერო ცენტრი“ თბ., 2023, 331 გვ. https://old.gtu.ge/book/ims/GisSurg_ERP-1990.pdf

44. Berdzenishvili I., Surguladze G. lean management, devops, agile softwaredevelopment for robotic process automation. Transactions of GTUniv., Automated Control Systems, No 1(33), vol.1, 2022. pp. 51-57. DOI.org/10.36073/1512-3979

45. Deep learning. Internet resource: https://en.wikipedia.org/wiki/Deep_learning

46. Deng L., Yu D. Deep Learning: Methods and Applications. Microsoft Research. USA, 2014. <https://www.microsoft.com/en-us/rese>

-arch/wp-content/uploads/2016/02/DeepLearning-NowPublishing-Vol7-SIG-039.pdf

47. Hebb D.O. The organization of behavior; a neuropsychological theory. Wiley, 1949.

48. Atsuhiko S. and Yasuhito T. Maximin and Minimax Strategies in Two-Players Game with Two Strategic Variables, 2018.

49. Willett, P. The Porter stemming algorithm: then and now. Program: Electronic Library and Information Systems, 40 (3). pp. 219-223, 2006, ISSN 0033-0337.

50. The Turing Test. Stanford Encyclopedia of Philosophy. First published Wed Apr 9, 2003; substantive revision Mon Oct 4, 2021. Internet resource: <https://plato.stanford.edu/entries/turing-test/> (6.10.23)

51. Grün, B., & Hornik, K. topicmodels: An R Package for Fitting Topic Models. Journal of Statistical Software, 2011, 40(13), 1–30

52. Lee, J., Kang, J.-H., Jun, S., Lim, H., Jang, D., & Park, S. Ensemble Modeling for Sustainable Technology Transfer. In Sustainability, 2018, Vol. 10, Issue 7, p. 2278. MDPI AG.

53. Hassan S. U., Ahamed J., & Ahmad K. Analytics of machine learning-based algorithms for text classification. In Sustainable Operations and Computers, 2022, Vol. 3, pp. 238–248. Elsevier BV.

54. Hearst M. A., Dumais S. T., Osuna E., Platt J., Scholkopf B. Support vector machines. IEEE Intelligent Systems and their applications, 1998 Jul;13(4):18-28.

55. Steinwart I., Christmann A. Support vector machines. Springer Science & Business Media; 2008 Sep 15.

56. სურგულაძე გ., გულუა დ., კახელი ბ. პროგრამული აპლიკაციების აგება ვირტუალიზაციის პირობებში. ISBN 978-9941-8-0627-4. სტუ. „IT-კონსალტინგ ცენტრი“. თბ., 2019. -159 გვ. https://gtu.ge/book/SurGuluaKakheli_Virtualization.pdf

57. Amazon Simple Storage Service. Developer Guide 2006
<https://docs.aws.amazon.com/AmazonS3/latest/dev/s3-dg.pdf>

58. Using AWS Lambda with Amazon DynamoDB. <https://docs.aws.amazon.com/lambda/latest/dg/with-ddb.html>

59. Amazon Virtual Private Cloud User Guide. Copyright © 2019 Amazon Web Services, Inc. and/or its affiliates. <https://docs.aws.amazon.com/vpc/latest/userguide/vpc-ug.pdf>

60. სურგულაძე გ., თოფურია ნ. მონაცემთა ბაზების მართვის სისტემები: ობიექტ-როლური მოდელირება (ORM / ERM, SQL Server). ISBN 99940-995-3-7. სტუ. „ტექ.უნივერს.“, თბ., 2007 -56 გვ.

61. მეიერ-ვეგენერი კ., სურგულაძე გ., ბასილაძე გ. საინფორმაციო სისტემების აგება მულტიმედიური მონაცემთა ბაზებით. ISBN 978-9941-20-468-5. სტუ. „ტექნიკ.უნივერსიტეტი“, თბ., 2014 345 გვ.

62. ჩოგოვაძე გ., ფრანგიშვილი ა., სურგულაძე გ., მართვის საინფორმაციო სისტემების დაპროგრამების ჰიბრიდული ტექნოლოგიები და მონაცემთა მენეჯმენტი. ISBN 978-9941-20-790-7. სტუ. „ტექნიკ.უნივერსიტეტი“, თბ., 2017. -1001 გვ.

63. სურგულაძე გ., პეტრიაშვილი ლ. მონაცემთა მენეჯმენტის თანამედროვე ტექნოლოგიები (Oracle, MySQL, MongoDB, Hadoop). ISBN 978-9941-27-176-2. სტუ. „IT-კონსალტ.ცენტრი“. თბ., 2017. -202 გვ.

64. 10 Essential Firebase Tutorials for Beginners. Dev. Mrcflorian. 2023. Internet resource: <https://dev.to/mrcflorian/10-essential-firebase-tutorials-for-beginners-4i39> (15.08.2023)

65. pickle – Python object serialization. Internet resource: <https://docs.python.org/3/library/pickle.html> (20.08.23)

66. Sharma H. Topic Model Visualization using - pyLDavis. 2021. Internet resource: <https://towardsdatascience.com/topic-model-visualization-using-pyldavis-fecd7c18fbf6> (20.08.23)

67. Matplotlib: Visualization with Python. Internet resource: <https://matplotlib.org/> (20.08.23)
68. სურგულაძე გ., თურქია ე. ინფორმატიკა - „პროგრამული ინჟინერია“ (საბაკალავრო ნაშრომის მეთოდური მითითებანი). ISBN 978-9941-8-2927-7. სტუ. თბ., 2021, -42 გვ.
69. სურგულაძე გ. გამოყენებითი პროგრამული ინჟინერია (საპროექტო სილაბუსის მეთოდური მითითებანი). ISBN 978-9941-8-2871-3. სტუ. თბ., 2020, 48 გვ.
70. სურგულაძე გ., ბიტარაშვილი მ., ქრისტესიაშვილი ხ. პროგრამული აპლიკაციების დეველოპმენტის საფუძვლები /საკურსო პროექტებისთვის/ (C#, MsAccess, ADO & ASP.NET). ISBN 978-9941-20-251-3. სტუ. „ტექნიკ.უნივერს.“, თბ., 2013 -96 გვ.
71. სურგულაძე გ. ობიექტ-ორიენტირებული დაპროგრამების მეთოდი (C++-ის ბაზაზე). ISBN 99940-56-18-2. სტუ. „ტექნიკ.უნივერს.“, თბ., 2008 -164 გვ.
72. სურგულაძე გ., ს. დოლიძე. მომხმარებლის ინტერფეისის დაპროგრამება (AngularJS, ReactJS). ISBN 978-9941-8-0625-4. სტუ. „IT-კონსალტინგ ცენტრი“. თბ., 2019. -106 გვ.
73. სურგულაძე გ., გულიტაშვილი მ., კივილაძე ნ. Web-აპლიკაციების ტესტირება, ვალიდაცია და ვერიფიკაცია. ISBN 978-9941-0-7682-4. სტუ. „IT-კონსალტინგ ცენტრი“. თბ., 2015. -205 გვ.
74. სურგულაძე გ., ბულია ი. კორპორაციულ Web-აპლიკაციათა ინტეგრაცია და დაპროექტება. ISBN 978-9941-20-165-3. სტუ. „ტექნიკ.უნივერს.“, თბ., 2012 -324 გვ.
75. სურგულაძე გ., ბულია დ., თურქია ე. Web-აპლიკაციების დამუშავება მონაცემთა ბაზების საფუძველზე (ADO.NET, ASP.NET, C#). ISBN 978-9941-14-289-5. სტუ. „ტექნიკ.უნივ.“, თბ., 2009 -189 გვ.
76. Pete LePage, Sam Richard. What are Progressive Web Apps ? Internet resource: <https://web.dev/what-are-pwas/> (09.01.2022)

77. Sen J., Mehtab S., Sen R., Dutta A., Kherwa P., Ahmed S., Berry P., Khurana S., Singh S., Cadotte D.W.W., Anderson D.W., Ost K.J., Akinbo R.S., Daramola O.A., Lainjo B. Machine Learning: Algorithms, Models and Applications. Internet resource: <https://arxiv.org/abs/2201.01943> (14.12.22)

78. Clement C.B., Bierbaum M., O'Keeffe K.P., Alemi A.A. On the Use of ArXiv as a Dataset (Version 1). arXiv, 2019

79. National Library of Medicine. National Center for Biotechnology Information. Internet resource: <https://pubmed.ncbi.nlm.nih.gov/> (13.05.22)

80. Neri Van Otten. What is CountVectorizer in NLP ? 2023. Internet resource: <https://spotintelligence.com/2023/05/17/countvectorizer/>

81. Tapson J., Greene J.R. Plant data visualization using non-negative matrix factorization. In IFAC Proceedings Volumes, 2005, Vol. 38, Issue 1, pp. 73–78. Elsevier BV

82. History of Microsoft Word. 2022. Internet resource: https://en.wikipedia.org/wiki/History_of_Microsoft_Word. (12.10.23)

83. გოგშელიძე დ. ტექსტური ინფორმაციის შექმნის ბიზნეს-პროცესების მოდელირების მხარდაჭერი პლატფორმა სერვის ორიენტირებული მიდგომით. სტუ-ს შრ.კრ., „მას“, 2022, No 1(33), vol.2. 41-44. EISSN 1512-2174, DOI.org/10.36073/1512-3979

84. 15 Best JavaScript Chart Libraries in 2023. Internet resource: <https://www.atatus.com/blog/javascript-chart-libraries/>

85. Build Diagrams for the Web in JavaScript and TypeScript. 2023. Internet resource: <https://gojs.net/latest/index.html> (10.10.23)

86. bpmn-js. BPMN 2.0 viewer and editor. Internet resource: <https://bpmn.io/toolkit/bpmn-js/> (20.10.23)

87. Draw2D JavaScript graph and diagramming lib. Internet resource: <https://freegroup.github.io/draw2d/#/api/draw2d>

88. Draw2D JavaScript graph and diagramming lib. Internet resource: <https://freegroup.github.io/draw2d/#/api/draw2d>

89. NIFTI-Reader-JS. Internet resource: <https://www.npmjs.com/package/nifti-reader-js> (22.10.23)

90. Jánoky L. V., Levendovszky J., & Ekler P. An analysis on the revoking mechanisms for JSON Web Tokens. In International Journal of Distributed Sensor Networks, 2018, Vol. 14, Issue 9, p. 155014771880153. SAGE Publications.

91. გოგშელიძე დ., გოგიშვილი ა. სხვადასხვა კომპონენტების განსაზღვრა ტექსტური ინფორმაციის შექმნისა და საგამომცემლო მარკეტინგის ბიზნეს პროცესების მოდელირებისას. სტუ-ს შრ.კრ., „მართვის ავტომატიზებული სისტემები“, No 2(29), 2019, გვ. 178-182. EISSN 1512-2174,

92. Cahn A., Alfeld S., Barford P., Muthukrishnan S. An Empirical Study of Web Cookies. In Proceedings of the 25th International Conference on World Wide Web. WWW '16: 2016.

93. Turian J., Ratnoff L., and Bengio Y. "Word representations: a simple and general method for semi-supervised learning." Proceedings of the 48th annual meeting of the association for computational linguistics, 2010.

94. Ando R., Zhang T. A Highperformance Semi-Supervised Learning Method for Text Chunking. ACL. 2005.

95. Bengio Y. Neural net language models. Scholarpedia, 2008, 3, 3881

96. k-nearest neighbors algorithm. Internet resource: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

97. ჩოგოვაძე გ., სურგულაძე გ., გულიტაშვილი მ., დოლიძე ს. პროგრამული აპლიკაციების ხარისხის მართვა: ტესტირება და ოპტიმიზაცია. ISBN 978-9941-20-629-2. სტუ. „IT-კონსალტინგ ცენტრი“. თბ., 2020. -363 გვ.

98. Costa G., Ortale R. Effective interrelation of Bayesian nonparametric document clustering and embedded-topic modeling. In Knowledge-Based Systems, Elsevier BV. 2021, Vol. 234, 107591

99. Brownlee J. How to Implement Bayesian Optimization from Scratch in Python. Internet resource: 2019. <https://machinelearningmastery.com/what-is-bayesian-optimization/>

100. გოგშელიძე დ. საგამომცემლო მარკეტინგის ბიზნეს პროცესების ავტომატიზაცია სერვის ორიენტირებული არქიტექტურით. დოქტ. აკად. ხარისხის მისანიჭ. დისერტ. ავტორეფ., (ხელმძღვ. გ. სურგულაძე). სტუ, თბ., 2022, 32 გვ.

101. Beautiful and accessible math in all browsers. Internet resource: <https://www.mathjax.org/> (1.10.23)

102. An extensive math library for JavaScript and Node.js. Internet resource: <https://mathjs.org/> (5.10.23)

103. სურგულაძე გ., ოხანაშვილი მ., ქრისტესიაშვილი ბ. მარკეტინგული პროცესების მოდელირება და ანალიზი UML დიაგრამებით და პეტრის ქსელებით, Intern. Science Conf. "Automat.Control Syst. &new IT" 20-22 Mai, 2013, გვ. 96.

104. ჩოგოვაძე გ., გოგიჩაიშვილი გ., სურგულაძე გ. მართვის ავტომატიზებული სისტემების" კათედრა 50 წლისაა – მისი როლი განათლებისა და მეცნიერების განვითარების სფეროში. სტუ-ს შრ. კრებ. „მას“, N1(32), ტ.1.1, 2021. თბ., გვ. 13-20. DOI.org/10.36073/1512-3979

105. სურგულაძე გ., ოხანაშვილი მ., სურგულაძე გიორგი. მარკეტინგის ბიზნეს-პროცესების უნიფიცირებული და იმიტაციური მოდელირება. ISBN 978-9941-14-377-9. სტუ. „ტექნიკური უნივერსიტეტი“, თბ., 2009 -170 გვ.

106. ზოგოვაძე გ. InforNation: ინფორმაცია, საზოგადოება, ხალხი. თბილისი, "ნეოსტუდია", 2003

107. ზოგოვაძე გ. გლობალური ბალანსი. მოსკოვი. „ზ. წერეთლის შემოქმედებითი ვორქშოფები" 2006

108. ზოგოვაძე გ. ბიოსფერო. მოსკოვი. „ზ. წერეთლის შემოქმედებითი ვორქშოფები". 2009

109. ზოგოვაძე გ. ფიქრები მომავლისთვის. თბილისი. -198 გვ. 2015

110. Memeti S., Pllana S., Binotto A., Kołodziej J., Brandic I. Using meta-heuristics and machine learning for software optimization of parallel computing systems: a systematic literature review. Springer. 2018. <https://link.springer.com/article/10.1007/s00607-018-0614-9>

111. Applications of artificial intelligence. Internet resource: https://en.wikipedia.org/wiki/Applications_of_artificial_intelligence.

112. Shitut N. Most Popular Regression Algorithms in Machine Learning. Internet resource: <https://translate.google.com/>

113. რეგრესიული ანალიზი. ლექსიკონი-ცნობარი სოცმეცნიერებებში. 2016. <http://dictionary.css.ge/content/analysis-regression>

114. Linear regression. Internet resource: https://en.wikipedia.org/wiki/Linear_regression (11.09.23)

115. Logistic regression. Internet resource: https://en.wikipedia.org/wiki/Logistic_regression (14.09.23)

116. Decision tree learning. Internet resource: https://en.wikipedia.org/wiki/Decision_tree_learning (14.09.23)

117. Willems K. Python Machine Learning: Scikit-Learn Tutorial. 2019. Internet resource: <https://www.datacamp.com/community/tutorials/machine-learning-python> (14.02.23)

118. Bayesian Methods for Machine Learning. MEDIUM. 2019. Internet resource: <https://towardsdatascience.com/bayesian-methods-for-machine-learning/home> (16.02.23).

119. Wong P. Predicting vs. Explaining. And Why Data Science Needs More "Half-Bayesians" <https://towardsdatascience.com/predicting-vs-explaining-69b516f90796> (16.02.23)

120. Brownlee J. How to Implement Bayesian Optimization from Scratch in Python. Internet resource: 2019. <https://machinelearningmastery.com/what-is-bayesian-optimization/> (1.03.23)

121. სურგულაძე გ., გულუა დ. განაწილებული სისტემების ობიექტ-ორიენტირებული მოდელირება უნიფიცირებული პეტრის ქსელებით. ISBN 9940-48-07-4. სტუ. „ტექნ.უნივ.“. თბ., 2005. -210 გვ.

122. Ray S. Understanding Support Vector Machine(SVM) algorithm from examples (along with code). 2017. Internet resource: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>

123. k-nearest neighbors algorithm. Internet resource: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

124. Machine Learning Made in a Minute. Internet resource: <http://accord-framework.net/>

125. Kriegel H., Schubert E., Zimek A. The (black) art of runtime evaluation: Are we comparing algorithms or implementations. Knowledge and Information Systems. 52 (2), Munich. Gernany 2016. 341–378. doi:10.1007/s10115-016-1004-2. ISSN 0219-1377

126. Classification Algorithms - Random Forest. Internet resource: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm

127. Artificial neural network. Internet resource: https://en.wikipedia.org/wiki/Artificial_neural_network

128. Pelk H. Machine Learning, Neural Networks and Algorithms. <https://chatbotsmagazine.com/machine-learning-neural-networks-and-algorithms-5c0711eb8f9a>

129. Deep learning. Internet resource: https://en.wikipedia.org/wiki/Deep_learning

130. Schulz H., Behnke S. Deep Learning. KI Künstliche Intelligenz. 26 (4), 2012, pp. 357–363. doi:10.1007/s13218-012-0198-z. ISSN 1610-1987. https://www.researchgate.net/publication/230690795_Deep_Learning_Layer-wise_Learning_of_Feature_Hierarchies

131. Deng L., Yu D. Deep Learning: Methods and Applications. Microsoft Research. USA, 2014. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DeepLearning-NowPublishing-Vol7-SIG-039.pdf>

დანართები:

დანართი-1: მარკეტინგული მენეჯმენტის ბიზნესპროცესები	172
დანართი-2: პეტრის ქსელების თეორიული საფუძვლები	179
დანართი-3: ERP-საწარმოო რესურსების მართვის სისტემა	186
დანართი-4: პროგრამების ოპტიმიზაციის მანქანური დასწავლის ალგორითმები	201
დანართი-5: DevOps მეთოდოლოგია და ინსტრუმენტული საშუალებები	213

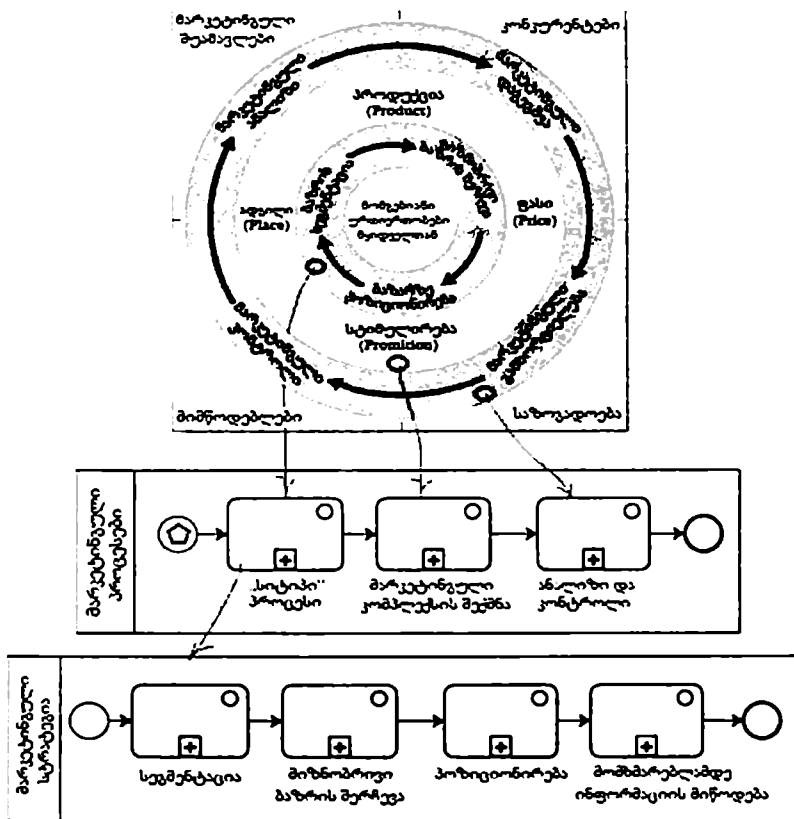
დ-1. მარკეტინგული მენეჯმენტის ბიზნესპროცესები

კორპორაციის ორგანიზაციული მართვის (მენეჯმენტის) საკითხებში მნიშვნელოვანი ადგილი უკავია მარკეტინგს, როგორც „ბიზნესის ფილოსოფიას“ – ჩიკაგოს უნივერსიტეტის ეკონომიკის პროფესორის, ფილიპ კოტლერის (Philip Kotler) განმარტება [14].

მარკეტინგი, როგორც მმართველობითი კონცეფცია, ჩვენს ქვეყანაში, საბაზრო რეფორმების განვითარებასთან ერთად, საკმაოდ ფართოდ გავრცელდა ინდუსტრიის, განათლების, სოფლის მეურნეობის და სხვ. დარგების ბიზნესისა და კომერციის ობიექტებზე. მისი ერთ-ერთი ძირითადი ამოცანაა ბაზრის მოთხოვნილებების შესაბამისი პროდუქციის შეთავაზება მყიდველზე.

კოტლერის მარკეტინგული საქმიანი ფუნქციონალობის ზოგადი მოდელი, ჩვენი ინტერპრეტაციით BPMN-ის ბაზაზე, წარმოდგენილია დ.1.1 ნახაზზე. მარკეტინგის მოდელზე ნათლად ჩანს ის ფაქტორები, რომლებიც ზემოქმედებას ახდენს მის სტრატეგიაზე [15].

მარკეტინგის პროცესის არსი ჩვენი წიგნის გადმოსახედიდან, ასეთია: (საგამომცემლო) ფირმა აწარმოებს პროდუქციას (წიგნებს ან სხვა ნაშრომებს) გარკვეული ფასით, გავრცელების მეთოდებით და სტიმულირების სხვადასხვა მექანიზმით ცდილობს ის მიიტანოს მიზნობრივ მყიდველამდე. მარკეტინგი მოიცავს პროდუქციის წარმოების დაგეგმვის, საქონლის შექმნისა და მისი ბაზარზე რეალიზაციის ეტაპებს. ამიტომაც მარკეტინგული პროცესების მართვის სისტემა შედგება ოთხი ქვესისტემისგან: *მარკეტინგული ანალიზი, მარკეტინგული დაგეგმვა, მარკეტინგული განხორციელება და მარკეტინგული კონტროლი*, რომლებიც ერთმანეთთან მჭიდრო კავშირშია (ნახ. დ.1.1 - გარე წრე).



ნახ. დ1.1. მარკეტინგის კოტლურისეული მოდელის ასახვა BPMN ნოტაციით

მყიდველი მოთავსებული ცენტრში. სისტემის მიზანია ძლიერი და მომგებიანი ურთიერთობების ჩამოყალიბების ხელშეწყობა მყიდველებთან. შემდეგ მოდის მარკეტინგული სტრატეგია – მარკეტინგული ლოგიკა, რომლის საშუალებით კომპანია ამ მომგებიანი ურთიერთობების ჩამოყალიბებას იმედოვნებს. ბაზრის

სეგმენტაციის, მიზნობრივი ბაზრის შერჩევისა და პოზიციონირების საშუალებით კომპანია წყვეტს, თუ რომელ მყიდველს როგორ მოემსახურება. იგი მთლიანი ბაზრის იდენტიფიცირებას ახდენს, შემდეგ მას სეგმენტებად ყოფს. მათ შორის ყველაზე პერსპექტიულს ირჩევს და მის მომსახურებასა და დაკმაყოფილებაზე ამახვილებს ყურადღებას.

მარკეტინგული სტრატეგიის ხელმძღვანელობით იგი ქმნის მარკეტინგულ კომპლექსს, რომელიც შედგება კომპანიის მიერ კონტროლირებადი ფაქტორებისგან: პროდუქტი, ფასი, ადგილი და სტიმულირება.

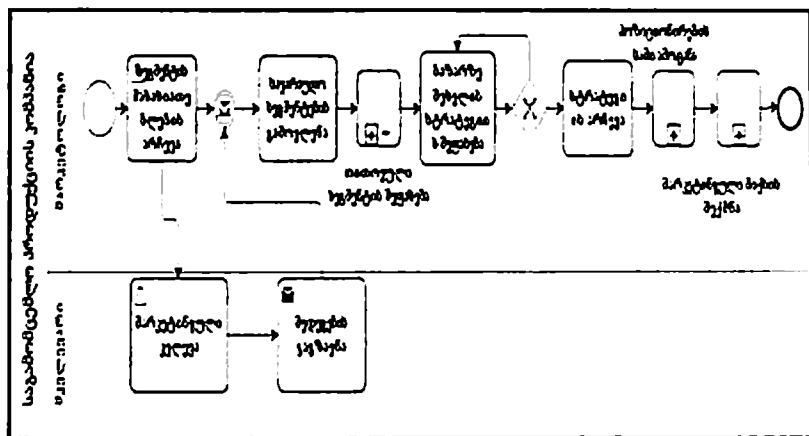
საუკეთესო მარკეტინგული სტრატეგიისა და კომპლექსის მისაღწევად კომპანია ჩაბმულია მარკეტინგული ანალიზის, დაგეგმვის, რეალიზაციისა და კონტროლის საქმიანობებში. ეს კი კომპანიას საშუალებას აძლევს თვალყური ადევნოს საბაზრო გარემოს და მის მონაწილეებს.

❖ მარკეტინგული ანალიზის STP-პროცესი

განვიხილოთ საგამომცემლო პროდუქციის მწარმოებელი კომპანიების მაგალითზე ბაზრის სეგმენტირების, მიზნობრივი ბაზრის არჩევის და პოზიციონირების პროცესები ე.წ. „სი-ტი-პი“ პროცესი (STP – Segment, Target, Position) დ1.1 ნახაზზე ცენტრის პირველი გარე წრე.

დ1.2 ნახაზზე მოცემულია STP-პროცესის BPMN დიაგრამა [15, 103].

➤ **ბაზრის სეგმენტაცია** – მთავარი მარკეტინგული გადაწყვეტილება – როგორ შეიძლება ბაზრის დაყოფა მყიდველთა ჯგუფების გათალისწინებით. ანუ, მყიდველთან სწორი ურთიერთობის ჩამოყალიბების სტრატეგია უნდა იქნას შემუშავებული.



ნახ. დ1.2. STP პროცესის BPMN დიაგრამა

ბაზრის სეგმენტაციის საშუალებით კომპანიები მსხვილ, პეტეროგენულ ბაზარს პატარა სეგმენტებად ყოფენ, რომელთა მოცვა უფრო ეფექტურად შეიძლება იმ პროდუქტებისა და მომსახურების საშუალებით, რომლებიც მყიდველთა უნიკალურ მოთხოვნილებებს პასუხობს.

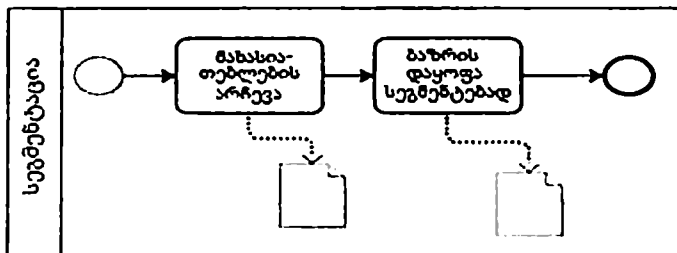
არსებობს სეგმენტაციის სამი მნიშვნელოვანი ტიპი:

- 1) სამომხმარებლო ბაზრების სეგმენტაცია;
- 2) ბიზნესბაზრების სეგმენტაცია და
- 3) საერთაშორისო ბაზრების სეგმენტაცია.

ბაზრის სტრუქტურის უკეთ დასანახად მარკეტოლოგმა უნდა მოსინჯოს სეგმენტაციის სხვადასხვა საშუალება: გეოგრაფიული და დემოგრაფიული (ინდუსტრია, კომპანიის ზომა) ფაქტორები, სარგებელი, მოხმარების სტატუსი, მოხმარების სიხშირე და ერთგულების სტატუსი, მყიდველის საოპერაციო მახასიათებლები, შესყიდვისადმი მიდგომა, სიტუაციური ფაქტორები და პიროვნული მახასიათებლები.

მათ საფუძველზე ხდება ბიზნესბაზრების სეგმენტაცია. შესაძლებელია, როგორც ერთი, ასევე რამდენიმე მახასიათებლის კომბინირებული გამოყენებაც.

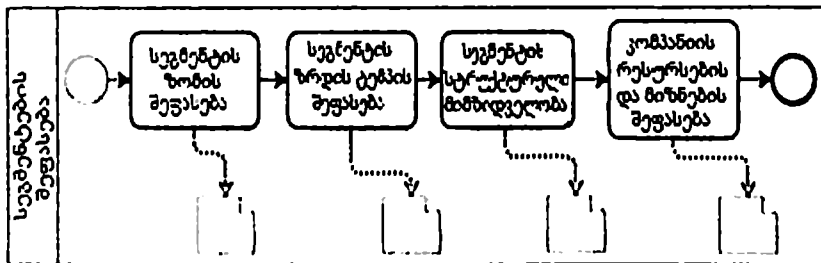
მარკეტოლოგი არსებული მახასიათებლებიდან თავიდან ირჩევს რამდენიმეს, რომ გამოავლინოს საგარაუდო სეგმენტები და შექმნას სეგმენტაციის მატრიცა (ნახ.დ1.3).



ნახ. დ1.3. სეგმენტაციის BPMN მოდელი

ფუნქტური სეგმენტაციისთვის მან უნდა გაითვალისწინოს სამი მთავარი მოთხოვნა: ყველა სეგმენტი უნდა იყოს ადვილად გაზომვადი, ხელმისაწვდომი, მნიშვნელოვანი.

შესაძლო სეგმენტების გამოვლენასთან ერთად უნდა განისაზღვროს საგამომცემლო კომპანიის მიზნობრივი ბაზარი. ამისათვის კი საჭიროა თითოეული სეგმენტის შეფასება (ნახ.დ1.4).



ნახ. დ1.4. სეგმენტის შეფასების პროცესის BPMN დიაგრამა

სეგმენტების შეფასება ხდება სამი კრიტერიუმის მიხედვით:

1) *სეგმენტის ზომა და ზრდის ტემპი* – კომპანიამ უნდა შეაგროვოს მონაცემები თითოეული სეგმენტის მსყიდველობით უნარიანობაზე, ზრდის ტემპსა და მოსალოდნელ მომგებიანობაზე;

2) *სეგმენტის სტრუქტურული მიმზიდველობა* – უნდა შემოწმდეს სტრუქტურული ფაქტორები, კონკურენტების და შემცველი პროდუქტების რაოდენობა;

3) *კომპანიის მიზნები და რესურსები* – იმ შემთხვევაშიც კი, თუ სეგმენტი ხასიათდება საჭირო ზომით, ზრდის ტემპითა და სტრუქტურული მიმზიდველობით, კომპანიამ საკუთარი მიზნები და რესურსები მაინც არ უნდა დაივიწყოს. ზოგიერთი მიმზიდველი სეგმენტი შესაძლოა დაწუნებული იყოს იმ მიზეზით, რომ ის კომპანიის გრძელვადიან მიზნებს არ შეესაბამება. კომპანიამ უნდა გაითვალისწინოს, ფლობს თუ არა იგი ამ სეგმენტში წარმატების მიღწევისათვის საჭირო კვალიფიკაციასა და რესურსებს.

➤ *მიზნობრივი ბაზრის შერჩევა* – სეგმენტების შეფასების შემდეგ კომპანია წყვეტს დაუკავებელი და მიმზიდველი სეგმენტების განსაზღვრის საკითხს, საკუთარი რესურსების და მიზნების გათვალისწინებით. ამის საფუძველზე კი ხდება *მიზნობრივი ბაზრის შერჩევა*.

➤ *ბაზარზე პოზიციონირება* – კომპანიამ, მიზნობრივი ბაზრის არჩევის შემდეგ, უნდა აირჩიოს თუ რა *პოზიციის* დაკავება სურს მას ამ ბაზარზე.

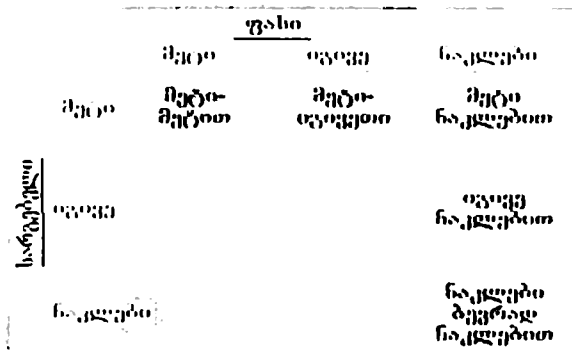
პოზიციონირების ამოცანა სამი საფეხურისგან შედგება:

- დადგინდეს შესაძლო კონკურენტულ უპირატესობათა ჩამონათვალი;

- ალტერნატივებიდან ამოირჩეს სწორი კონკურენტული უპირატესობა;

- შეირჩეს საერთო პოზიციონირების სტრატეგია.

ამის შემდეგ კომპანიამ ეფექტურად უნდა აწარმოოს ურთიერთობები და თავისი პოზიცია მიზნობრივ ჯგუფს წარუდგინოს. არსებობს 9 პოზიცია, აქედან მომგებიანია – 5 (ნახ. დ1.5).



ნახ. დ1.5. პოზიციონირების ცხრა პოზიცია

საგამომცემლო მარკეტინგის პროგრამული უზრუნველყოფა ეხმარება ბიზნეს-კომპანიებს მარკეტინგული მიზნების შესრულებაში. მისი გამოყენება შესაძლებელია მარკეტინგული მენეჯმენტის პროცესების ავტომატიზაციისთვის. იგი ეხმარება ბიზნესს სრულყოფილ თავისი მარკეტინგული ამოცანების გადაჭრა და მიაღწიოს მიზნებს უფრო სწრაფად, ჰკვიანურად და შეცდომების გარეშე.

პროგრამული უზრუნველყოფა უზრუნველყოფს მონაცემთა ინტელექტუალურ (BI) ანალიზს და შიგა კონტროლს, ეყრდნობა ცენტრალურ მონაცემთა ბაზას, რომელიც აგროვებს მონაცემებს განყოფილებებიდან, მათ შორის ბუღალტერია, წარმოება, მიწოდების ჯაჭვის მენეჯმენტი, გაყიდვები, მარკეტინგი, ადამიანური რესურსები და სხვ. კომპლექსური ავტომატიზაციის ასეთი ტექნოლოგია თამამად შეიძლება შევადაროთ ERP სისტემას.

მომდევნო დანართში ამ საკითხებს შევხებით.

დ-2. ინფორმაციული სისტემები (IS) და საწარმოო რესურსების მართვა (ERP, ისტორიულ ფაქტებით)

ჩვენი წიგნის 1.4. პარაგრაფში მოკლედ განვიხილეთ საკითხი *ERP სიტემა საგამომცემლო ინდუსტრიაში*. მისი, როგორც ერთ-ერთი კომპლექსური ავტომატიზაციის ტიპის ინფორმაციული სისტემის როლი პროდუქციის მწარმოებელი კორპორაციებისთვის, უდავოდ მნიშვნელოვანია.

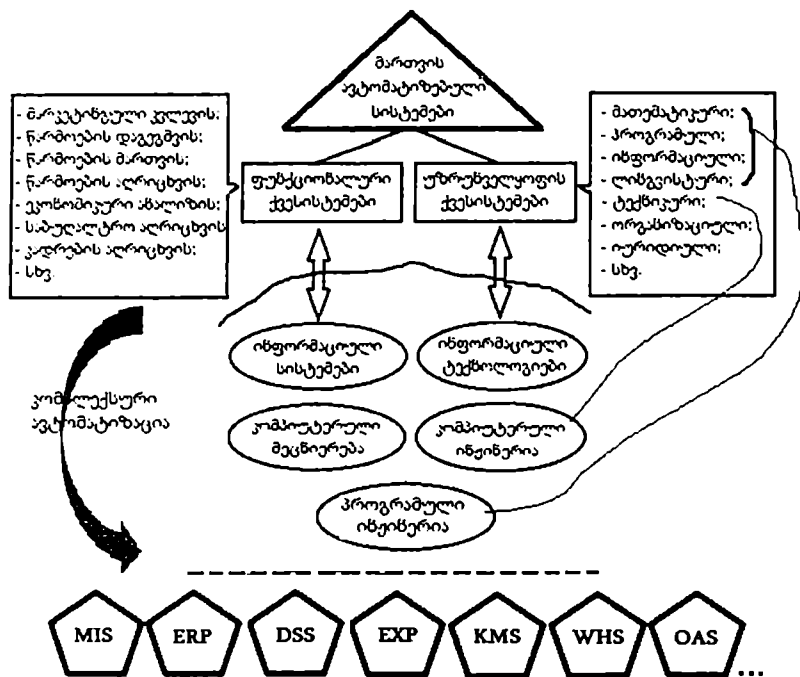
დანართში გვინდა დამატებით წარმოვადგინოთ ის ინოვაციური სამუშაოები და შედეგები ამ სფეროში, რომლებიც სტუ-ს მართვის ავტომატიზებული სისტემების კათედრაზე შესრულდა მისი არსებობის 50 წლის მანძილზე (1971-2021), აკადემიკოს გოჩა ჩოგოვადის ხელმძღვანელობით (ან მხარდაჭერით, მისი UNESCO-ში მუშაობის და პარიზში ცხოვრების პერიოდში 1980-2012 წლებში) [106-109].

ყველაფერი დაიწყო იმით, რომ სტუ-ს მართვის ავტომატიზებული სისტემების კათედრის (დაარსებიდან - 1971) აკადემიური და სამეცნიერო მიმართულება იყო „*ადმას*“ – ინფორმაციის დამუშავებისა და მართვის ავტომატიზებული სისტემები (*გამოსაშვებ სტუდენტთა სპეციალობა*), ანუ ეს გახლდათ ინფორმაციული, პროგრამული, ტექნიკური და ორგანიზაციული უზრუნველყოფების ერთობლიობა, თითქმის ყველა დარგის კონკრეტული ობიექტისათვის. ცოტა ხანში (1976) ჩამოყალიბდა ახალი სამეცნიერო მიმართულება მონაცემთა ინფორმაციული ბაზების, საწარმოო პროცესებისა და დაპროგრამების ავტომატიზაციის ამოცანების გადასაწყვეტად.

გ. ჩოგოვადის ოპერატიული მართვის პროცესების ავტომატიზაცია და გ. გოგიჩაიშვილის სიტუაციური მართვის პროცესების

კვლევა სემანტიკური ქსელებით ხელოვნური ინტელექტის მეთოდების საფუძველზე, 1975-1985 წლებში გახდა ჩვენი კათედრის პრიორიტეტული სამეცნიერო მიმართულება [104].

ახალგაზრდა მეცნიერ-მკვლევართა რამდენიმე ჯგუფი აქტიურად მუშაობდა მართვის ავტომატიზებული სისტემების (*მას*) ფუნქციონალური და უზრუნველყოფის ქვესისტემების პროექტირებისა და პროგრამირების მიმართულებით. *მას*-ის კლასიკური სტრუქტურა ასეთია (ნახ. დ2.1 - ზედა ნაწილი).



➤ **ფუნქციონალური ქვესისტემა** ორგანიზაციის (კორპორაციის) ფუნქციურ განყოფილებათა (ან დეპარტამენტების) მართვის ავტომატიზებული ქვესისტემაა. ტერმინი „ავტომატიზებული“ ნიშნავს ადამიანის უშუალო მონაწილეობას მართვის პროცესში (გადაწყვეტილების მისაღებად). მის გარეშე გვაქვს „ავტომატიური“ მართვის სისტემა (ანუ მართვა ტექნიკურ სისტემებში, სადაც რეგულირების თეორია დომინირებს).

➤ **უზრუნველყოფის ქვესისტემა** ახორციელებს ფუნქციონალური ქვესისტემების მხარდაჭერას, რათა მათ შეასრულონ ავტომატიზებულ რეჟიმში თავიანთი ბიზნეს-ამოცანები.

დღევანდელი გადმოსახედიდან (კომპიუტერული სისტემების ამერიკული მოდელისა და შესაბამის აკადემიურ დისციპლინათა ანალიზის საფუძველზე) შეიძლება ასეთი სურათი წარმოვადგინოთ (ნახ. დ.3.1 - ქვედა ნაწილი). აქ ჩანს ინფორმაციულ სისტემათა ძირითადი ტიპები:

- MIS – Management Information System (მენეჯმენტის საინფორმაციო სისტემები). მენეჯმენტი - ორგანიზაციული მართვა;
- ERP – Enterprise Resource Planing (საწარმოო რესურსების დაგეგმვა (მართვა);
- DSS – Decision Support System (გადაწყვეტილების მიღების მხარდამჭერი სისტემა);
- EXP – Expert Systems (AI, ექსპერტული სისტემები (ხელოვნური ინტელექტის ქვემიმართულება);
- KMS – Knowledge Management System (ცოდნის მართვის სისტემა);
- WHS – Warehousing Systems (OLAP, Data Mining, BigData) (მონაცემთა საცავების სისტემები);
- OAS – Office Automation Systems (ოფისის ავტომატიზაციის სისტემები, მაგ., Office-365) და სხვ.

სტუ-ის ინფორმატიკის ფაკულტეტის *მას* კათედრამ განსაკუთრებული როლი ითამაშა ERP – საწარმოს რესურსების დაგეგმვის (მართვის) რეალური სისტემების დამუშავებაში (თუმცა ტერმინი ERP არ მოიხსენიებოდა 80-იან წლებში. იგი 90-იანის შემდეგ გამოჩნდა). კონკრეტულად ამ მიმართულებას კათედრაზე პროფ. გ. სურგულაძე (წიგნის თანაავტორი) უძღვებოდა.

საწარმოო პროცესების მართვის ავტომატიზებული სისტემა (პირველი ქართული ERP-ს პროტოტიპი) დაინერგა თბილისის N1 ტრიკოტაჟის ფაბრიკაში (გენერალური დირექტორი თამაზ ცინცაძე 50-იანი წლების გვი-ს სტუდენტი და მისი პირველი ჯაზის კვარტეტის მომღერალი) [43].

ERP-ს პროტოტიპის სისტემაში რეალიზებული იყო მარკეტინგის, საწარმოო და ტექნოლოგიური პროცესების მენეჯმენტის, დაპროექტების ავტომატიზაციის და ოპერატიული მართვის ამოცანები კლასიკური და ხელოვნური ინტელექტის მეთოდების გამოყენებით.

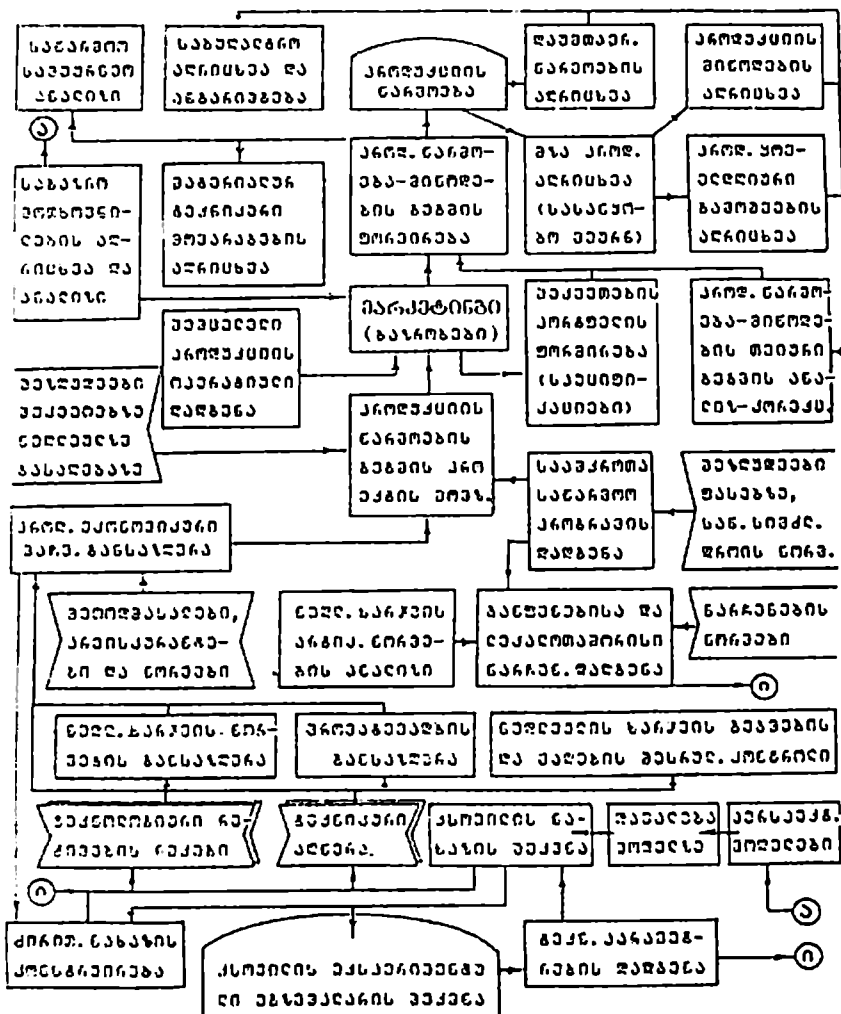
მარკეტინგული პროცესების ავტომატიზებული სისტემის საწარმოო-ექსპერიმენტული გამოცდა (პრაქტიკული დანერგვა) ჩატარდა საქართველოს მსუბუქი მრეწველობის საწარმოო ობიექტების დარგობრივ ბაზრობაზე (თბილისის დიდუბის საგამოფენო პავილიონში Expo Georgia). დარგობრივი სამინისტროს ხელმძღვანელობით ჩატარდა საქართველოს 20-ზე მეტი საქსოვი, სამკერვალო და სხვა ფაბრიკების საკუთარი პროდუქციის გამოფენა-გაყიდვის პროცესი (წლიური კონტრაქტების გაფორმება სავაჭრო ორგანიზაციებთან). 1987 წელს ჩვენი სისტემის ტირაჟირების საფუძველზე მომზადდა საქართველოს სხვა საქსოვი და სამკერვალო ფაბრიკების მარკეტინგული პროცესების ავტომატიზაცია. ჩვენი ERP-სისტემა ამავე წელს შემოდგომით მონაწილეობდა

მოსკოვში, საკავშირო მსუბუქი მრეწველობის სამინისტროს ობიექტების პროდუქციის ბაზრობაზე (ამოცანები: კონტრაქტების გაფორმების პროცესის ავტომატიზაცია, მონაცემთა ბაზების დინამიკური ცვლილებების სწრაფი ასახვა, გადაწყვეტილების მიღებისთვის ხელმძღვანელებზე საჭირო ინფორმაციის დროული მიწოდება და სხვ.). ჩვენ ვემსახურებოდით საქართველოს დელეგაციას (10-ზე მეტი ფაბრიკა).

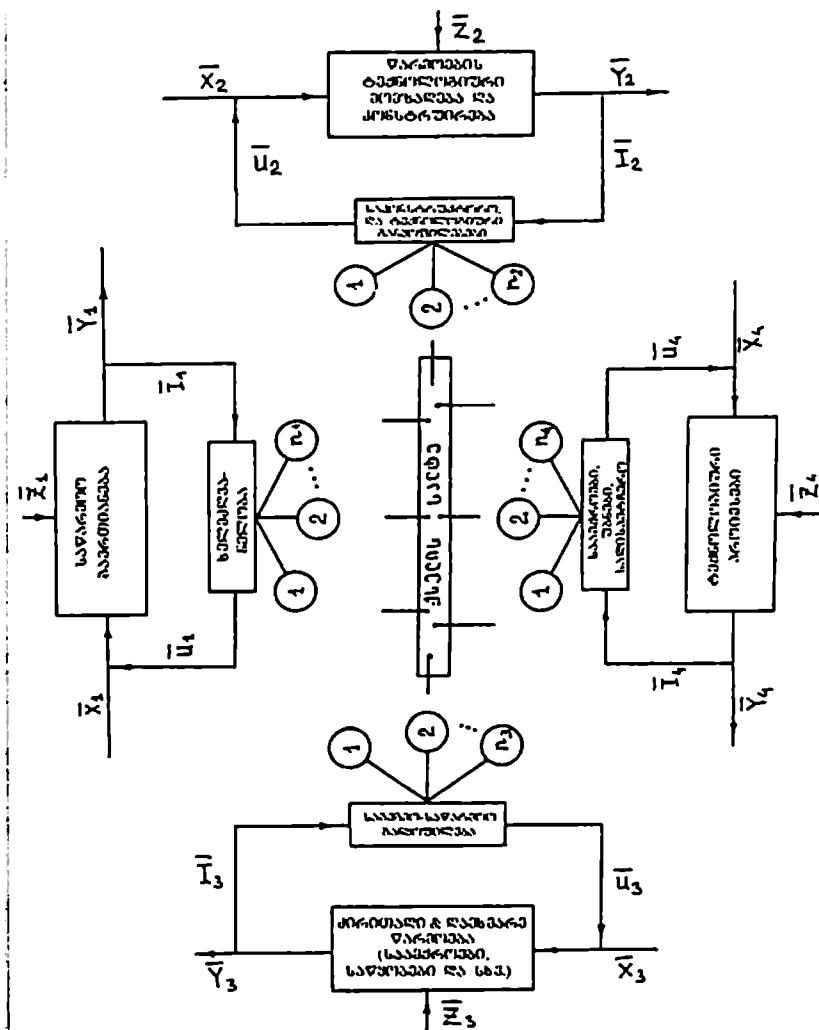
დ.2.2 ნახაზზე წარმოდგენილია ფაბრიკის საწარმოო პროცესების ინფორმაციულ-ტექნოლოგიური სქემა, რომლისთვისაც აგებულ იქნა ERP-სისტემის პროტოტიპი. აგებული მოდელი შესაძლებელია გამოყენებულ იქნას სხვა პროდუქციის მწარმოებელი კომპანიებისთვისაც, მაგალითად, საგამომცემლო ბიზნესის ობიექტებისთვის, მცირე ადაპტაციით. ამ თვალსაზრისით, ჩვენ შეიძლება 1.3 პარაგრაფში (წიგნების გამოცემის პროცესის სასიცოცხლო ციკლი) აღწერილი ეტაპები (დაგეგმვის, წარმოების, რეალიზაციის, მარკეტინგული კვლევის და ა.შ.) და მათი ამოცანები გადავიტანოთ პროგრამულ უზრუნველყოფაში.

დ.2.3 ნახაზზე წარმოდგენილია ERP-პროტოტიპის სისტემის სტრუქტურა. იგი კომპლექსური სისტემაა და შედგება, ძირითადად, ოთხი ტიპის პროცესების მართვის ბლოკისაგან, რომლებიც კავშირშია ერთმანეთთან ლოკალური ქსელის საშუალებით.

ბლოკების შიგნით განთავსებულია ფუნქციური ავტომატიზებული სამუშაო ადგილები (ლოკალური ქსელური კავშირებით მონაცემთა ცენტრალიზებულ ბაზასთან და ერთმანეთთან). ინტეგრირებული ERP-სისტემის ფუნქციური სამუშაო ადგილები ოთხი ძირითადი კლასისაა: „ხელმძღვანელის“ (მმართველობითი ამოცანების გადასაწყვეტად); „დაპროექტების“ (საპროექტო და საკონსტრუქტორო ამოცანებისთვის); „ტექნოლოგიურის“ და „საგეგმო-ეკონომიკური გაანგარიშებების“ [დეტალები იხ. 43-ში].



ახ. დ2.2. საწარმოო პროცესების ინფორმაციულ-ტექნოლოგიური სქემა



ნახ. დ2.3. ERP- ტიპის კომპლექსური ავტომატიზაციის სისტემის სქემა

დ-3. პეტრის ქსელების თეორიული საფუძვლები

პეტრის ქსელების ისტორია 1962 წლიდან იწყება, როცა გერმანელმა ინჟინერმა, კარლ ადამ პეტრიმ (Carl Adam Petri) დარმშტადტის ტექნიკურ უნივერსიტეტში დაიცვა სადოქტორო დისერტაცია თემაზე „კომუნიკაცია ავტომატებით“

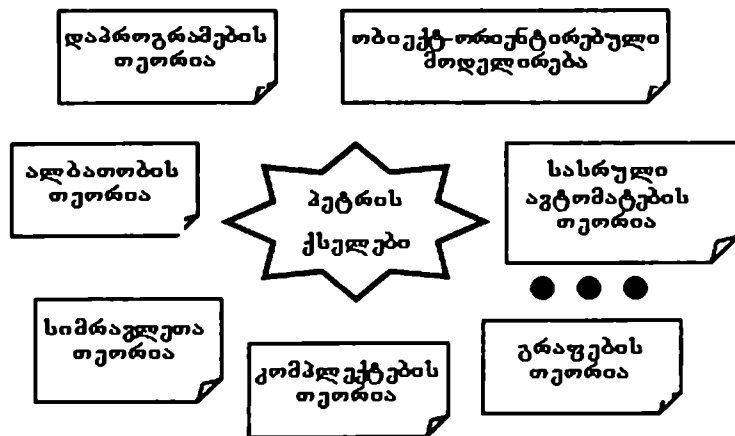
ამ ნაშრომში მან პირველად ჩამოაყალიბა და დაასაბუთა იდეა ორი განსხვავებული ტიპის: *კვანძებით (პოზიციებით)* და მათი დამაკავშირებელი მიმართული *რკალებით (გადასასვლელ-ლებით)* აგებული მუშა ქსელების შესახებ, რომლებიც ერთი მოდელის ფარგლებში გააერთიანებდა პროცესებს და მონაცემებს.

თავისი თეორიის საფუძვლად კ. პეტრიმ სასრული ავტომატების, სიმრავლეთა და გრაფების თეორიების ელემენტები გამოიყენა.

პეტრის ქსელები გრაფო-ანალიზური ინსტრუმენტია მათემატიკური შიგთავსის აპარატი. იგი გამოიყენება თითქმის ყველა სფეროს და ყველა დარგის ავტომატიზებული კვლევის პროცესებში, როგორც სიმულაციის და ანალიზის ძლიერი და მოქნილი საშუალება.

ერთის მხრივ პეტრის ქსელები ფართოვდება თეორიულად, სულ უფრო მძლავრი ხდება მისი მათემატიკური აპარატი, იქმნება ახალი თეორიული კლასები.

მეორეს მხრივ მატულობს პეტრის ქსელების პრაქტიკული გამოყენების სიხშირე ინფორმატიკის მოწინავე მიმართულებებთან მჭიდრო თანამშრომლობის შედეგად, რაც სქემატურად დ3.1 ნახაზზეა გამოსახული.



ნახ. დ3.1. პეტრის ქსელები და მისი გარემოცვა

➤ სიმრავლეები

სიმრავლეთა თეორია პეტრის ქსელების ერთერთი ბაზისია. განვიხილოთ მთელმთელი მისი ძირითადი ელემენტები.

საწყისი აღნიშვნები:

$N = \{0, 1, \dots\}$ - ნატურალურ რიცხეთა სიმრავლე;

$Z = \{\dots, -1, 0, 1, \dots\}$ - მთელ რიცხეთა სიმრავლე

$Boolean = \{true, false\}$ - ბულის სიმრავლე

სიმრავლე არაერთგვაროვან ობიექტთა ერთობ-ღიობაა. მათ სიმრავლის **ელემენტები** ეწოდება.

a არის **A**-სიმრავლის ელემენტი, თუკი ფლობს თვისებას $a \in A$ („მიეკუთვნება“). სიმრავლე მოიცემა ელემენტთა ნამონათვალით $A = \{a_1, a_2, \dots, a_n\}$, ან გარკვეულ $p(a)$ ფუნქციასე დაყრდნობით, რომლის შედეგი სიმრავლის ელემენტისთვის აუცილებელ პირობას აკმაყოფილებს:

$$A = \{a \mid p(a)\}.$$

ცარიელი სიმრავლე $\emptyset$ -სიმბოლოთი აღინიშნება და გამოისახება პირობით $\emptyset = \{a \mid a \neq a\}$, რადგან პირობა $a \neq a$ ყოველთვის მცდარია.

სიმრავლეთა თეორიაში განისაზღვრება შემდეგი ძირითადი დამოკიდებულებები და ოპერაციები: **ქვესიმრავლე** ($A \subseteq B$), **ჯეშმარიტი ქვესიმრავლე** ($A \subset B$), **გაერთიანება** ($A \cup B$), **თანაკვეთა** ($A \cap B$), **სხვაობა** ($A \setminus B$), სადაც:

$A \subseteq B$, როცა ნებისმიერი $a \in A$ -თვის მართებულია $a \in B$

$A \subset B$, როცა $A \subseteq B$ და $A \neq B$

$A \cup B = \{a \mid a \in A \text{ ან } a \in B\}$

$A \cap B = \{a \mid a \in A \text{ და } a \in B\}$

$A \setminus B = \{a \mid a \in A \text{ და } a \notin B\}$

ცარიელი სიმრავლე ნებისმიერი არაკარიელი სიმრავლის ქვესიმრავლეა: $\emptyset \subset A$. A და B სიმრავლეებს **განცალკევებული სიმრავლეები** ეწოდება, თუ $A \cap B = \emptyset$.

სიმრავლე შეიძლება შეიცავდეს ელემენტებს, რომლებიც თავადაა სიმრავლეები. A -სიმრავლის ყველა შესაძლო ქვესიმრავლეთა სიმრავლე $\Pi(A)$ -თი აღინიშნება, ნატურალურ რიცხვთა სიმრავლე $\mathbb{N}$ -ის ნათელით $\mathbb{N}$ -ით, ლოგიკურ მნიშვნელობათა (ჯეშმარიტი ან მცდარი) სიმრავლე – B -თი.

$A_1, A_2, \dots, A_n$ ($n \in \mathbb{N}$) სასრულ სიმრავლეთა **პროდუქტი** (დეკარტული ნამრავლი) განისაზღვრება შემდეგნაირად:

$$A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n) \mid a_i \in A_i \text{ } i=1, \dots, n\}$$

პროდუქტის ცალკეულ ელემენტს n -კორტეჟი ეწოდება. ყოველი i -სთვის, სადაც $1 \leq i \leq n$, a_i -ს კორტეჟის i -ური ელემენტი ეწოდება $(a_1, \dots, a_n)$. წყვილი განისაზღვრება, როგორც n -კორტეჟის კერძო შემთხვევა, 2-კორტეჟი (ბინარული).

თუ ყველა სიმრავლე $A_1 = A_2 = \dots = A_n = A$ მსგავსია, პროდუქტი ნაიწერება A^n სახით. გარდა ამისა, $A^1 = A$ და $A^0 = \emptyset$.

➤ მულტისიმრავლეები (კომპლექტები)

მულტისიმრავლე განისაზღვრება სიმრავლეში ერთი და იმავე ელემენტის რამდენჯერმე ასახეისთვის. მაგალითად, პეტრის ქსელის პოზიციაში რამდენიმე მსგავსი მარკერის აღსაწერად.

მულტისიმრავლე B არაცარიელ საბაზო A სიმრავლეზე, ეწოდება ფუნქციას:

$$B:A \rightarrow N$$

სადაც საბაზო **A** სიმრავლის ყოველი $a \in A$ ელემენტის **სიხშირე B** მულტისიმრავლეში აისახება ფორმით **B(a)**. სიხშირის სიდიდე შეიძლება 0-ის ტოლიც იყოს. სიმრავლე მულტისიმრავლის სპეციალური შემთხვევაა, სადაც სიხშირის მნიშვნელობებია 0 ან 1.

მულტისიმრავლის ასახვის გაფართოებული ფორმა შემდეგია: $[a, a, \dots, a, b, \dots, b, \dots]$, სადაც ყოველი ელემენტი თავისი სიხშირის მიხედვით მეორდება.

მულტისიმრავლეთა სიმრავლე საბაზო **A** სიმრავლეზე აღინიშნება μA -თი. მულტისიმრავლე ასევე შეიძლება გამოისახოს სიმბოლური ჯამის სახით, რომელიც $a \in A$ ელემენტის სიხშირეს და სახელს შეიცავს:

$$B = \sum_{a \in A} B(a)a$$

თუ $B(a) = 1$, მაშინ ჯამურ ასახვაში საერთოდ გამოიტოვება და იწერება მხოლოდ **a**.

$B \in \mu A$ მულტისიმრავლეში, $a \in A$ ელემენტს ეწოდება **B-ს** წევრი და ნაიწერება $a \in B$, თუ $B(a) > 0$ და პირიქით, თუ $B(a) < 0$, მაშინ $a \notin B$. ცარიელი მულტისიმრავლე $\emptyset$ წევრებს არ შეიცავს: $\forall a \in A, \emptyset(a) = 0$.

მულტისიმრავლის **სიმბლავრე (კარდინალურობა)** მისი ყველა ელემენტის **სიხშირეთა ჯამს** ეწოდება და განისაზღვება შემდეგნაირად:

$$|B| = \sum_{a \in A} B(a)$$

თუ $|B|$ სასრულია, მაშინ მულტისიმრავლე B -ს სასრული მულტისიმრავლე ეწოდება.

ორი მულტისიმრავლე, B_1 და B_2 ტოლია ($B_1=B_2$), თუ $\forall a \in A, B_1(a)=B_2(a)$.

B_1 ნაკლებია ან ტოლია B_2 -ის (ანუ B_2 მოიცავს B_1 -ს) თუ $\forall a \in A, B_1(a) \leq B_2(a)$.

მულტისიმრავლეებზე ძირითად ოპერაციებს წარმოადგენს:

შეკრება: $B=B_1+B_2$, თუ $\forall a \in A, B(a) = B_1(a) + B_2(a)$

გამოკლება: $B=B_1-B_2$, თუ $\forall a \in A, ((B_1(a) \geq B_2(a)) \wedge ((B(a) = B_1(a) - B_2(a)))$

სკალარული ნამრავლი: მულტისიმრავლის $B_1 \in \mu A$ და ნატურალური რიცხვის $n \in \mathbb{N}$ სკალარული ნამრავლი განისაზღვრება როგორც $B=n \cdot B_1$, თუ $\forall a \in A, B(a) = n \cdot B_1(a)$, სადაც
*** არითმეტიკული გამრავლების ოპერაციაა.

➤ ძირითადი ცნებები [121]:

- საბაზო სიმრავლე (*Basis Set*). ობიექტების საწყისი სიმრავლე მულტი-სიმრავლეების (კომპლექტების) შესაქმნელად.

- მულტისიმრავლე ანუ კომპლექტი (*Multiset*) ობიექტების ნაკრები, სადაც ერთგვაროვანი ობიექტების განმეორება შესაძლებელია.

- მულტისიმრავლის კარდინალურობა (*Cardinality*). მულტი-სიმრავლის ელემენტების საერთო რაოდენობა.

- პოზიცია (*Place*) ქსელის ტიპიზებული კვანძი. ქსელის გრაფში წრით ან ელიფსით გამოისახება.

- გადასასვლელი (*Transition*). ქსელის არატიპიზებული კვანძი, რომელიც მართკუთხედით გამოისახება.

- რკალი (*Arc*). ქსელის მიმართული კავშირის ხაზი, რომელიც აერთებს პოზიციებს გადასასვლელებთან (შემაჯავლი რკალი) ან პირიქით (გამომავალი რკალი).

- *შემავალი პოზიცია (Input Place)*. გადასასვლელთან შემავალი რკალით შეერთებული პოზიცია.
- *გამომავალი პოზიცია (Output Place)*. გადასასვლელთან გამომავალი რკალით შეერთებული პოზიცია.
- *პოზიციის ტიპი (Place Type)*. პოზიციასთან დაკავშირებულ მონაცემთა ელემენტების არაღარიფი სიმრავლე.
- *მარკერი (Marker)*. პოზიციასთან დაკავშირებული და შესაბამისი პოზიციის ტიპის მონაცემთა ელემენტი.
- *მარკირება (Marking)*. ეველა პოზიციაში შემავალ მარკერთა ერთობლიობა.
- *საწყისი მარკირება (Initial Marking)*. ეველა პოზიციაში შემავალ მარკერთა ერთობლიობა ქსელის მუშაობის დასაწყისში.
- *პოზიციის მარკირება (Place Marking)*. პოზიციაში მოთავსებულ მარკერთა მულტისიმრავლე.
- *გადასასვლელის გახსნის პირობა (Transition Condition)*. გადასასვლელთან დაკავშირებული ლოგიკური (ბულის) ტიპის გამოსახულება.
- *გადასასვლელის გახსნის რეჟიმი (Transition Mode)*. ცვლადების დაკავშირება გადასასვლელის გახსნის პირობასთან ისე, რომ გადასასვლელის გახსნა ნებადართული გახდეს.
- *გადასასვლელის გახსნის ნებადართვა (Enabling a Transition)*. გადასასვლელი რომ გაიხსნას, ყოველი პოზიციის მარკირება უნდა აკმაყოფილებდეს მისი და გადასასვლელის დამაკავშირებელი რკალის მოთხოვნას (რკალის გამოსახულება). რაც ნიშნავს, რომ მარკირება შეიცავს მარკერების მინიმუმ იმავე მულტისიმრავლეს, რაც რკალის გამოსახულებაზეა ახასუელი.
- *გადასასვლელის გახსნა (Transition Occurrence)*. თუ გადასასვლელის გახსნა ნებადართულია, იგი შეიძლება გაიხსნას. ამ დროს გადასასვლელის ყოველი შემავალი პოზიციიდან მოიხსნება მარკერები გახსნის რეჟიმის შესაბამისად, ხოლო ყოველ გამომავალ პოზიციაში გამომავალი რკალების გამოსახულებათა შესაბამისი მარკერები ჩაემატება. პოზიცია შეიძლება ერთდროულად შემავალი და გამომავალი იყოს (მარყუქი)

- გადასასვლელის ცვლადები (*Transition Variables*). რეაქციისა და გადასასვლელის გახსნის პირობაში შემაჯავლი ცვლადების ერთობლიობა.
- რეაქციის ანოტაცია (*Arc Annotation*). გამოსახულება, რომელიც შეიძლება შეიცავდეს კონსტანტებს, ცვლადებს და ოპერატორებს რეაქციის დაკავშირებული პოზიციის ტიპის მუდგისიმრავლიდან.
- მიღწევათი მარკირება (*Reachable Marking*). მარკირება, რომელიც მიიღება ქსელის საწყისი მარკირებიდან გადასასვლელთა გარკვეული მიმდევრობის გახსნის შემდეგ.
- მიღწევათი მარკირებათა სიმრავლე (*Reachability Set*). საწყისი მარკირებიდან მიღწევათი მარკირებათა სიმრავლე თვით საწყისი მარკირების ნათესავით.
- ალგებრა (*Algebra*). მათემატიკური სტრუქტურა, რომელიც შეიცავს სიმრავლეთა სიმრავლეს და ფუნქციითა სიმრავლეს, რომლებიც ამ სიმრავლეთა დომენებსა და ქვედომენებზე მოქმედებს.
- ტიპი (*Sort*). მონაცემთა სტრუქტურის სახელი.
- არგუმენტის ტიპი (*Argument Sort*). ოპერატორის არგუმენტის ტიპი.
- გამოშვებითი ტიპი (*Output Sort*). ოპერატორის შედეგის ტიპი.
- არეკლობა (*Arity*) - ფუნქციაში შემავალი (არგუმენტები) და გამოშვებითი (შედეგი) ტიპები (მაგ., ბინარული, n-არული).
- ტიპიზაცია (*Typisation*). ტიპის დაკავშირება პოზიციასთან.
- აღწერები (*Declarations*). გამოსახულებათა სიმრავლე სიმრავლეთა, კონსტანტების, პარამეტრების მნიშვნელობათა, ტიპისებული ცვლადებისა და ფუნქციების განსაზღვრისათვის, რომლებიც მაღალი დონის პეტრის ქსელებზე აისახება.
- ოპერატორი (*Operator*). სიმბოლოთა ერთობლიობა (აბრევიატურა) ფუნქციის სახელის წარმოსადგენად.
- პარამეტრი (*Parameter*). მუდმივა (კონსტანტა), რომელიც სიმრავლეში განსაზღვრულ სიდიდეთა არეს შეიცავს.
- მინიჭება (*Assignment*). მნიშვნელობის მინიჭება ცვლადების სიმრავლის კონკრეტული ცვლადისათვის.
- სიგნატურა (*Signature*). ალგებრული სტრუქტურა, რომელიც ტიპების და ოპერატორების სიმრავლეებისგან შედგება.

- **ბულის სიგნატურა (Bool Signature).** სიგნატურა, რომელიც ბულის (ლოგიკურ) ტიპს შეიცავს.

- **მრავალტიპური სიგნატურა (Many-sorted Signature).** სიგნატურა, სადაც ტიპების სიმრავლის კარდინალურობა ერთზე მეტია.

- **ცვლადიანი სიგნატურა (Signature with Variables).** სიგნატურა, რომელიც შეიცავს ცვლადების სახელებს, ტიპებს და ოპერატორებს.

- **თერმი (Term).** სიგნატურის საფუძველზე შედგენილი გამოსახულება, რომელიც შეიცავს მუდმივებს, ცვლადებს და ოპერატორებს.

- **დახურული თერმი (Closed Term).** თერმი, რომელიც შეიცავს კონსტანტებს და ოპერატორებს, მაგრამ არა ცვლადებს.

- **თერმის მნიშვნელობა (Term Evaluation).** შედეგი, რომელიც მიიღება თერმის ცვლადებითვის მნიშვნელობების მინიჭებისა და ფუნქციითა შედეგების გამოთვლის შემდეგ.

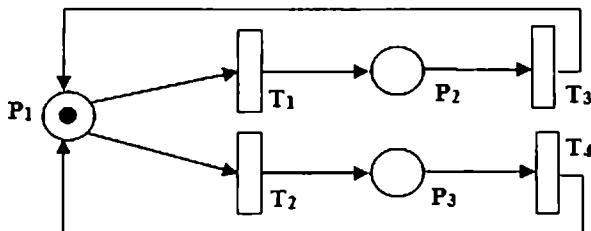
- **მაღალი დონის პეტრის ქსელი (High Level Petri Net).** ალგებრული სტრუქტურა, რომელიც შეიცავს: პოზიციების სიმრავლეს; გადასასვლელთა სიმრავლეს; ტიპების სიმრავლეს; ტიპების პოზიციებზე და ტიპების გადასასვლელებზე დამაკავშირებელ ფუნქციებს; პრეფუნქციებს შემავალი და პოსტფუნქციებს გამომავალი მარკირებების განსაზღვრისათვის; საწყის მარკირებას.

- **პეტრის ქსელის გრაფი (Petri Net Graph).** მიმართული გრაფი ორი ტიპის კვანძებითა (პოზიციები და გადასასვლელები) და მათი დამაკავშირებელი რკალებით. დაშვებულია კავშირები „პოზიცია-გადასასვლელი“ ან „გადასასვლელი-პოზიცია“. მაგრამ არა „პოზიცია-პოზიცია“ ან „გადასასვლელი-გადასასვლელი“.

- **მაღალი დონის პეტრის ქსელის გრაფი (High Level Petri Net Graph).** ქსელის გრაფისა და ანოტაციების (წარწერების) ერთობლიობა, რომელიც შეიცავს პოზიციითა ტიპებს, რკალების ანოტაციებს, გადასასვლელთა გახსნის პირობებს, შესაბამის განსაზღვრებებს განსაზღვრებათა სიაში და ქსელის საწყის მარკირებას.

- **მიღწევადობის გრაფი (Reachability Graph).** მიმართული გრაფი, სადაც კვანძები მიღწევად მარკირებებს შეესაბამება, რკალები - გადასასვლელთა გახსნის ოპერაციას.

➤ დაბალი დონის პეტრის ქსელის მაგალითი



ნახ. დ3.2. პეტრის ქსელის გრაფი სამი პოზიციით (P_1 , P_2 და P_3),
ოთხი გადასასვლელით (T_1, T_2, T_3 და T_4), ერთი მარკერი P_1 -ში

მუშაობის პრინციპი ასეთია: თუ გაიხსნა T_1 გადასასვლელი, მაშინ მარკერი გადაადგილდება P_2 პოზიციაში (P_1 - ში წაიშლება). ამ დროს გააქტიურდება T_3 გადასასვლელი, რადგანაც მის შესასვლელ რკალზე გაჩნდა მარკერი. შემდეგ იხსნება T_3 გადასასვლელი და მარკერი გადადის ისევ P_2 პოზიციაში. ახლა თუ გაიხსნა T_2 გადასასვლელი, მაშინ იმუშავებს გრაფის ქვედა შტო.

➤ მაღალი დონის პეტრის ქსელის მაგალითი

პეტრის ასეთ ქსელებს სისტემურსაც უწოდებენ [35]. ჩვენ მიერ 1.5 პარაგრაფში განხილული საგამომცემლო მარკეტინგის პროცესის პეტრის ფერადი ქსელი (Coloured Petri Net – CPN, ნახ.1.3) არის მაღალი დონის [25].

იმიტაციური მოდელირების (სიმულაციის) და ანალიზის ეს ინსტრუმენტი თავისუფალი წვდომის (Open access) პროდუქტია და შესაძლებელია მისი უფასოდ ჩამოწერა და დაინსტალირება. ამერიკის, ევროპის, ჩინეთის და სხვა უნივერსიტეტებში იგი აქტიურად გამოიყენება [36].

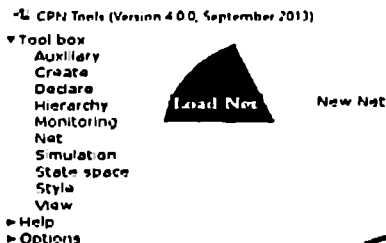
CPN Tools-ის განვითარების შედეგად შესაძლებელი გახდა CPN IDE-ს გამოიყენება. IDE-ის მთავარი უპირატესობა ისაა, რომ

იგი არის გაფართოებადი ინსტრუმენტი, რაც საჭიროა მისი მუშაობის შესანარჩუნებლად და ახალი ფუნქციების დასამატებლად [37].

დ.2.3 ნახაზზე ნაჩვენებია Tools-ის ლოგო და პეტრის ქსელის აგების სასტარტო ფრაგმენტი. იგი სხვადასხვა ფერის მარკერებს შეიცავს (რომლებიც ნატურალური რიცხვებით გამოისახება).

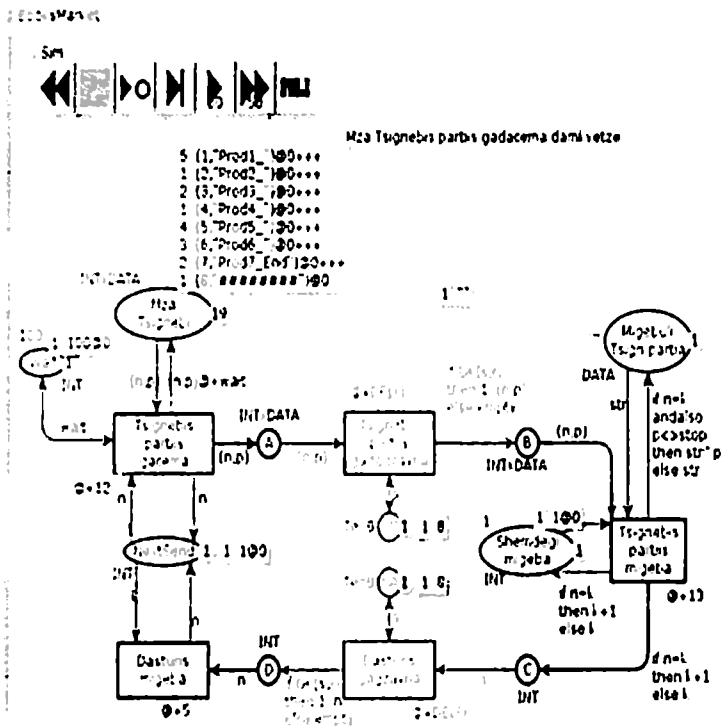


Tsign._gadacema



ნახ. დ.2.3. CPN Tools ლოგო და ქსელის აგების სასტარტო ფანჯარა

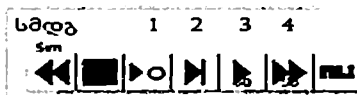
დ.3.4. ნახაზზე მოცემულია საგამომცემლო მარკეტინგის CPN ქსელის ფრაგმენტი პროდუქციის (საგამომცემლო სტამბაში დაბეჭდილი წიგნების პარტიის) მიწოდების იმიტაციური მოდელი.



**ნახ. დ.3.4. პროდუქციის გადაცემის CPN იმიტაციური მოდელი
(საწყისი მდგომარეობა)**

ნახაზის ზედა მარცხენა კუთხეში მოთავსებულია ქსელის პროცესების (მართკუთხედების) მუშაობის მართვის (ნავიგაციის) „სიმულაციის“ დილაკები (ნახ. დ3.5).

მათი საშუალებით შესაძლებელია პროცესების მართვა ხელით, თითო-ბიჯით, მითითებული ბიჯების რაოდენობით (ვიზუალური დაკვირვებისათვის) ან ავტომატურად (ბოლო შედეგების სწრაფად გამოსატანად).



ნახ. დ3.5. CPN-ის იმიტაციური პროცესის ნავიგატორი

დ3.6 ნახაზზე ნაჩვენებია CPN ქსელის Tool box (მაგალითად, აქედან Simulation-ის გადატანით სამუშაო ფორმაზე იხსნება ნავიგატორი).

Help და Options -ის გარდა, ქვემოთ გამოიტანება ჩვენი ქსელის (წიგნების პარტიის გადაცემის) იმიტაციური პროცესის პარამეტრები (საკონტროლო ბიჯები, დრო), ქსელის გამოცხადებული ცვლადები, ფუნქციები და ა.შ.

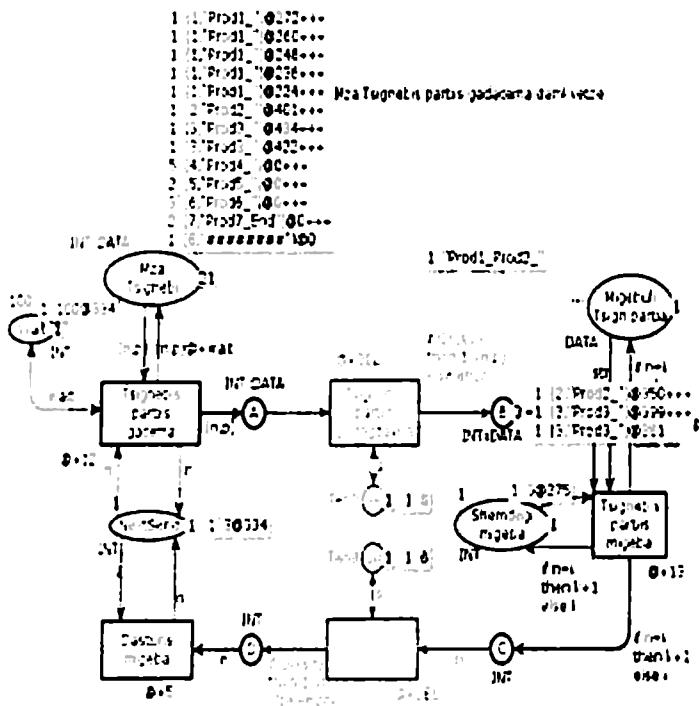
დ3.7. ნახაზზე მოცემულია, მაგალითად, ჩვენი ქსელის მუშაობის პროცესის იმიტაცია: დაბეჭდილი წიგნების პარტიების გადაგზავნა დამკვეთებთან, პროცესის კონტროლით (ბიჯი ავიღეთ 50).

- ▼ Tool box
 - Auxiliary
 - Create
 - Declare
 - Hierarchy
 - Monitoring
 - Net
 - Simulation
 - State space
 - Style
 - View
- Help
- Options
- ▼ TsionPart გადაცემა
 - Step: 0
 - Time: 0
- Options
- History
- ▼ Declarations
 - colset INT
 - colset DATA
 - colset INT+DATA
 - var n k wait
 - var p str
 - val stop
 - colset Ten0
 - colset Ten1
 - var i
 - var r
 - fun OK
 - colset NetDelay
 - fun DEL
- Monitors
 - BoolsMant et

ნახ. დ3.6. CPN ToolBox

ნახ. 3.7

ნახ. 3.7



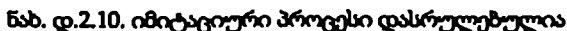
ნახ. დ.3.7. 50 ზოგის (Step) შემდეგ მიღებული მდგომარეობა

აქ ჩანს, რომ დამკვეთთან (მაგალითად, წიგნების მაღაზია, ბიბლიოთეკები და ა.შ.) მისულია 1 და 2 პროდუქტი, მე-3 პროდუქტი გზაშია. სისტემა პროცესებს გვერდით უწერს დროის ერთეულებს (ნახ. დ3.8).

Marlowe

100

12a Tsionebis parvis gadacema damkvetze



200

დ-4. პროგრამების ოპტიმიზაციის მანქანური დასწავლის ალგორითმები

პროგრამული უზრუნველყოფის ოპტიმიზაციის მიზნით, განსაკუთრებით პარალელური კომპიუტინგის სფეროში, აქტიურად გამოიყენება მეტა-ევერისტიკული და მანქანური დასწავლის ალგორითმები [97, 110]. ამჯერად განვიხილოთ მანქანური დასწავლის მეთოდები. კერძოდ, პროგრამული სისტემების ოპტიმიზაციის მიზნით გამოყენებადი ალგორითმები.

მანქანური დასწავლის პროგრამები მუშაობს პროგნოზირების მოდელის შექმნის საფუძველზე მასწავლი მონაცემების ერთობლიობიდან. ეს მოდელები შემდეგ გამოიყენება პროგნოზირების მიზნით მონაცემთა საფუძველზე. მანქანური დასწავლის ზოგიერთი პოპულარული ალგორითმებია: რეგრესიული, გადაწყვეტილების ხე, ვექტორული მანქანის მხარდამჭერი, ბაიესური დასკვნა, შემთხვევითი ტყე და ხელოვნური ნეირონული ქსელები (ნახ. დ4.1).

რეგრესიის ალგორითმი
გადაწყვეტილების ხე
ვექტორული მანქანის მხარდამჭერი
K-უახლოესი მეზობელი
კლასტერიზაცია
ბაიესური დასკვნა
შემთხვევითი ტყე
ნეირონული ქსელები
ღრმა სწავლება
და სხვ. ...

პროგნოზირების მოდელის დასწავლის მეთოდისაგან დამოკიდებულებით, მანქანური დასწავლა შეიძლება იყოს მართვადი (კონტროლირებადი) ან არამართვადი (არაკონტროლირებადი).

ნახ. დ4.1. მანქანური დასწავლის ალგორითმები

მართვადი მანქანური დასწავლის შემთხვევაში პროგნოზირების მოდელი სწავლობს არსებულ მაგალითებზე, ანუ შემავალი და გამომავალი მონაცემები დასწავლის მონაცემთა ერთობლიობაში ცნობილია. მართვად დასწავლაში გამოიყენება კლასიფიკაციის მეთოდები ცალკეული პასუხების პროგნოზირებისათვის, აგრეთვე რეგრესიის მეთოდები – უწყვეტი პასუხებისთვის (მაგალითად, ტემპერატურის ცვლილება). უკეთესი ალგორითმის შერჩევა დამოკიდებულია შემავალ მონაცემთა რაოდენობასა და ტიპზე, აგრეთვე სასურველ შედეგზე და მათი გამოყენების სახეზე.

მანქანური დასწავლის არამართვადი მოდელებისათვის პრაქტიკულად უცნობია თუ როგორი სახე აქვს შედეგებს. აქ სწორი შედეგები (მასწავლი მონაცემების ერთობლიობა) არ გამოიყენება მოდელის დასწავლისათვის. მოდელის მიზანია მონაცემებში ფარული შაბლონების პოვნა მასწავლ მონაცემთა ერთობლიობის სტატისტიკური თვისებების საფუძველზე. არამართვადი დასწავლა შეიძლება გამოყენებულ იქნას, მაგალითად, მონაცემთა კლასტერიზაციის პრობლემის გადასაწყვეტად სხვადასხვა სფეროში.

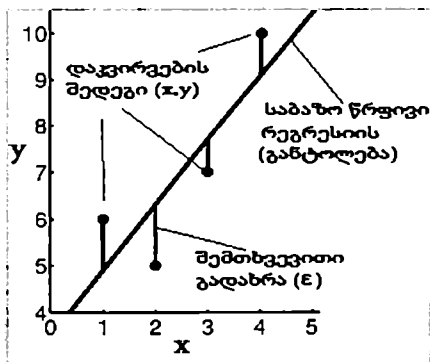
მანქანური დასწავლა (*Machine learning – ML*) – არის სამეცნიერო მიმართულება, რომელიც კომპიუტინგის სისტემებს, სრულყოფის მიზნით, აძლევს თვითგანსწავლის შესაძლებლობას ავტომატიზებულ რეჟიმში. მისი მეთოდები და ალგორითმები ორიენტირებულია ისეთი პროგრამული სისტემების შექმნაზე, რომლებსაც გამოცდილების საფუძველზე და ხელმისაწვდომი მონაცემების ბაზაზე შეუძლია ახალი ცოდნის შეძენა, ანუ თვითსრულყოფა. იგი ხელოვნური ინტელექტის პრაქტიკული გამოყენების სფეროა, ჰკვიანი პროგრამული აპლიკაციების ასაგებად [110,111].

ამ მიმართულებით, როგორც ლიტერატურულ წყაროებშია აღნიშნული, პერსპექტიულად ითვლება შემდეგი მეთოდები ან ალგორითმები [110] (ნახ. დ4.1):

- რეგრესიის ალგორითმები (*Regression algorithms RA*) – მიეკუთვნება მართვადი მანქანური დასწავლის ალგორითმების ოჯახს. რეგრესიის ალგორითმები აპროგნოზირებს გამომავალ მნიშვნელობებს (შედეგებს), სისტემის საწყისი (შესატანი) მონაცემების საფუძველზე. იგი გამოიყენება მრავალ სფეროში: ჯანდაცვა, განათლება, ფინანსები, ტრანსპორტი და ა.შ. [112].

მაგალითისათვის განვიხილოთ წრფივი რეგრესიის და ლოგისტიკური რეგრესიის ალგორითმები.

- **წრფივი რეგრესია** (*Linear Regression - LR*). – სტატისტიკაში გამოიყენება რეგრესიული მოდელი ერთი დამოკიდებული y



ცვლადისა ერთ (ან რამდენიმე) სხვა დამოუკიდებელ x ცვლადზე (ფაქტორები, რეგრესორები) (ნახ. დ4.2) [113, 114].

ნახ. დ4.2 წრფივი რეგრესია

წრფივი დამოკიდებულების ფუნქცია ჩაიწერება შემდეგი გამოსახულებით:

$$y = b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon$$

სადაც b და ε კოეფიციენტები გაითვლება უმცირეს კვადრატთა მეთოდით. იგი მდგომარეობს შემდეგში, რომ ყოველი

დამუკიდებელი (x) და დამოკიდებული (y) ცვლადების წყვილებისთვის (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) ვექტორ წრფივი რეგრესიის გამოყენებით ისეთ $y = \varepsilon + bx$ წრფეს, რომელშიც x დაკვირვებული მნიშვნელობების შეტანის შემდეგ, y -ის შესაბამისი მნიშვნელობებიდან გადახრათა კვადრატების ჯამი იქნება მინიმალური:

$$\sum_{i=1}^n (y_i - \varepsilon - bx_i)^2 \rightarrow \min$$

• **ლოგისტიკური რეგრესია** ან **ლოგისტიკური მოდელი** (*Logistic regression - Logit model*) – არის სტატისტიკური მოდელი, რომელიც გამოიყენება მოვლენის ალბათობის პროგნოზირებისთვის, მისი შედარებით ლოგისტიკურ მრუდთან [115]. ეს რეგრესია იძლევა პასუხს ბინარული მოვლენის ალბათობის სახით (1 ან 0).

$y=1$ ხდომილობის განხორციელების ალბათობა გამოითვლება ფორმულით:

$$p = \frac{1}{1 + e^z}$$

სადაც

$$z = b_1x_1 + b_2x_2 + \dots + b_nx_n$$

ლოგისტიკური რეგრესიის ამოცანა სწორედ ამ b_i -ური კოეფიციენტების გამოთვლაა.

გადაწყვეტილების ხე (*Decision tree DT*) და **გადაწყვეტილების ხის სწავლება** (*Decision tree learning*) – არის პროგნოზული მოდელირების ერთ-ერთი მიდგომა, რომლის გამოყენების სფეროებია: სტატისტიკა, მონაცემთა მოპოვება (ან ინტელექტუალური ანალიზი) და მანქანური დასწავლა. ეს მიდგომა იყენებს გადაწყვეტილების ხეს როგორც პროგნოზირების მოდელს, რათა ელემენტზე დაკვირვებებიდან (ასახული ხის შტოებში) გადავიდეს დასკვნებამდე ელემენტის მიზნობრივი მნიშვნელობის შესახებ

(ასახული ხის ფოთლებში). ხისმაგვარ მოდელებს, რომლებშიც მიზნობრივ ცვლადს შეუძლია მიიღოს მნიშვნელობათა დისკრეტული ერთობლიობა, უწოდებენ კლასიფიკაციის ხეებს. ამ ხისმაგვარ სტრუქტურებში ფოთლები არის კლასის ჭდეები (class labels), ხოლო შტოები – ფუნქციითა მახასიათებლების ერთობლიობა, რომლებსაც მიყვავართ ამ კლასის ჭდეებთან.

გადაწყვეტილების ხეებს, რომლებშიც მიზნობრივ ცვლადს შეუძლია მიიღოს უწყვეტი მნიშვნელობები (ნამდვილი რიცხვები), ეწოდება რეგრესიული ხეები.

ალტერნატიულ გადაწყვეტილებათა ანალიზის პროცესში გადაწყვეტილების ხე გამოიყენება როგორც ვიზუალური და ცხადი ინსტრუმენტული საშუალება საბოლოო შედეგის (გადაწყვეტილების) მისაღებად.

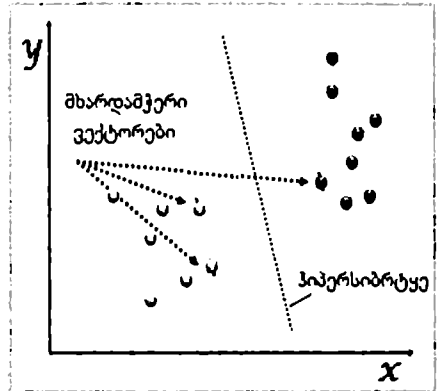
ხე შეიძლება „განსწავლილ“ იქნას სიმრავლის დაყოფით ქვესიმრავლეებად, შესაბამის ატრიბუტთა მნიშვნელობების შემოწმების საფუძველზე. ეს პროცესი მეორდება რეკურსიულად თითოეული ქვესიმრავლისთვის და მას ეწოდება რეკურსიული სექტორი (დანაყოფი). რეკურსია წყდება მაშინ, როდესაც ქვესიმრავლეს ხის კვანძში აქვს მიზნობრივი ცვლადის იგივე მნიშვნელობა (აღარ ემატება პროგნოზის ახალი მნიშვნელობა). ასეთი ინდუქციური top-down მეთოდი ხშირად გამოიყენება მონაცემთა გადაწყვეტილების ხეების სწავლების პროცესში [115, 116].

• **მხარდაჭერი ვექტორული მანქანა** (Support Vector Machine – SVM) – არის მანქანური დასწავლის მართვადი ალგორითმი, რომელიც შეიძლება გამოყენებულ იქნას როგორც კლასიფიკაციის (უმეტესად), ასევე რეგრესიული ამოცანებისთვის. SVM ალგორითმში მონაცემთა თითოეული ელემენტი აიგება როგორც წერტილი n -განზომილებიან სივრცეში (სადაც n მახასიათებლების

რაოდენობა). ამასთანავე, თითოეული მახასიათებლის მნიშვნელობა არის კონკრეტული კოორდინატის მნიშვნელობა. შემდეგ, სრულდება კლასიფიკაცია იმ ჰიპერ-სიბრტყის პოვნით, რომელიც კარგად განასხვავებს ორ კლასს (ნახ. დ4.3) [122].

მხარდამჭერი ვექტორები – ინდივიდუალური დაკვირვების კოორდინატებია. SVM კლასიფიკატორი არის საზღვარი, რომელიც უკეთესად გამოყოფს ორ კლასს:

ჰიპერ – სიბრტყე / ხაზი.



ნახ. დ4.3. მხარდამჭერი ვექტორები

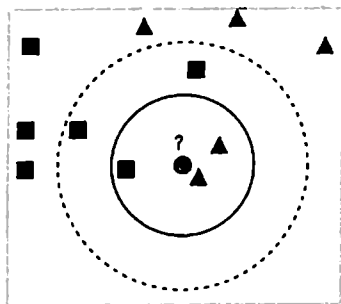
მანქანური დასწავლის ალგორითმების განსახორციელებლად ფართოდ გამოიყენება Python ენაში scikit-Learn ბიბლიოთეკა. SVM ალგორითმისათვისაც ხელმისაწვდომია ეს ბიბლიოთეკა [117].

ბაიესის დასკვნა (Bayesian inference – BI) – არის სტატისტიკური დასკვნის მეთოდი, რომლის დროსაც ბაიესის თეორემა გამოიყენება ჰიპოთეზის ალბათობის განახლების მიზნით, როდესაც მეტი მტკიცებულება ან ინფორმაციაა მიღებული. ბაიესის დასკვნის მნიშვნელოვანი ნაწილია პარამეტრების და მოდელების დადგენა [118, 119]. ბაიესის ოპტიმიზაციის საკითხი პროგრამული კოდებისთვის მეტად მნიშვნელოვანია, როგორც დეველოპმენტის,

ასევე ტესტირების პროცესში [120]. მაგალითად, როგორ შეიძლება გამოვიყენოთ ბაიესის მეთოდი რთული ოპტიმიზაციის პრობლემების კვლევაში. კერძოდ, გლობალური ოპტიმიზაციაში. როგორ განვახორციელოთ ბაიესის ოპტიმიზაცია და როგორ გამოვიყენოთ იგი ღია პროგრამული კოდისთვის.

• **K-უახლოესი მეზობელი (*K*-nearest neighbor *K*/N)** არის მეტრიკული ალგორითმი ობიექტების ავტომატური კლასიფიკაციის ან რეგრესიისათვის.

კლასიფიკაციის მეთოდის გამოყენების შემთხვევაში, ობიექტი მიეკუთვნება იმ კლასს, რომელიც ყველაზე გავრცელებულია ამ ელემენტის *K* მეზობლებს შორის, რომელთა კლასები უკვე ცნობილია (ნახ. დ4.4).



რეგრესიის მეთოდის გამოყენების შემთხვევაში, ობიექტს ენიჭება საშუალო მნიშვნელობა *K* უახლოეს ობიექტებზე, რომელთა მნიშვნელობებიც ცნობილია.

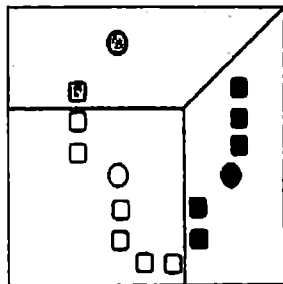
ნახ. დ4.4. *K*-უახლოესი მეზობელი

ალგორითმის გამოყენება შესაძლებელია დიდი რაოდენობით ატრიბუტების მქონე (მრავალგანზომილებიან) ნიმუშებზე.

ამისათვის, გამოყენებამდე უნდა დადგინდეს მანძილის ფუნქცია; ასეთი ფუნქციის კლასიკური ვერსია ევკლიდეს მეტრიკაა - ესაა მანძილი ორ წერტილს შორის ევკლიდეს სივრცეში, რომელიც გაითვლება პითაგორას თეორემით.

• **k-საშუალოების კლასტერიზაცია** (*k-means clustering*) – არის ვექტორული კვანტირების (შეკუმშვის) მეთოდი, რომელიც გამოიყენება მონაცემთა მოპოვების (Data Mining) კლასტერულ ანალიზში (მაგალითად, მონაცემთა ბაზაში მსგავსი სტრუქტურების აღმოჩენის მეთოდი) [115]. k-საშუალოების კლასტერიზაციის მიზანია n დაკვირვებების დაყოფა k კლასტერად, რომლებშიც თითოეული დაკვირვება მიეკუთვნება კლასტერს უახლოესი სშუალო მნიშვნელობით და წარმოადგენს კლასტერის პროტოტიპს. შედეგად მონაცემთა სივრცე იყოფა ვორონოის უჯრედებად (Voronoi cells) (ნახ. დ4.5).

k-საშუალოებს მინიმუმამდე დაჰყავს დისპერსიები კლასტერის შიგნით (ევკლიდური მანძილის კვადრატი). საშუალო მნიშვნელობა ასდენს კვადრატული შეცდომების ოპტიმიზაციას, ხოლო გეომეტრიული მედიანა ამცირებს ევკლიდურ მანძილებს.



ნახ.4.5. k-საშუალოები
კლასტერიზაცია

k-საშუალოების კლასტერიზაციის ალგორითმი პროგრამულად რეალიზებულია სხვადასხვა ენაზე, მათ შორის არის Free/Open Source Software ვერსიებიც [123]. მაგალითად, Accord.NET ფრეიმვორკი, რომელშიც k-means და სხვა ვერსიები რეალიზებულია C# ენაზე. ისინი გამოიყენება სამეცნიერო გამოთვლებისათვის .NET-ში.

პროგრამები ძირითადად განსხვავდება მწარმოებლობით, (შესრულების სწრაფქმედებით), მაგალითად, 10 წამიდან – რამდენიმე საათამდე დიაპაზონში [124, 125]. განსხვავება არის, რა თქმა უნდა, შესრულების ხარისხშიც და შედეგების სიზუსტეშიც.

• **შემთხვევითი ტყე (Random forest)** – არის მანქანური დასწავლის მართვადი ალგორითმი, რომელიც გამოიყენება როგორც კლასიფიკაციის, ასევე რეგრესიისთვის. ანალოგიურად, ეს ალგორითმი ქმნის გადაწყვეტილების ხეებს მონაცემთა სახეების ამოსარჩევად, შემდეგ კი თითოეული მათგანისგან იღებს პროგნოზს და, საბოლოოდ, კენჭისყრით ირჩევს საუკეთესო შედეგს [126].

შემთხვევითი ტყის ალგორითმის მუშაობა ხორციელდება, ზოგადად, შემდეგი ბიჯებით:

1-ბიჯი: შემთხვევითი ნიმუშების შერჩევის დაწყება მოცემულ მონაცემთა ერთობლიობიდან;

2-ბიჯი: ეს ალგორითმი ააგებს გადაწყვეტილების ხეს ყველა ნიმუშისათვის. შემდეგ ის მიიღებს პროგნოზირების შედეგს ყველა გადაწყვეტილების ხისგან;

3-ბიჯი: ხმის მიცემა განხორციელდება ყველა პროგნოზირებული შედეგისთვის;

4-ბიჯი: აირჩევა ყველაზე მეტი ხმით პროგნოზირებული შედეგი, როგორც საბოლოო.

➤ **უპირატესობები:**

– დიდი რაოდენობის მახასიათებლებისა და კლასების მქონე მონაცემთა ეფექტურად დამუშავების უნარი;

– მახასიათებლების (ატრიბუტის) მნიშვნელობათა მასშტაბირებისადმი (და სხვა გარდაქმნებისადმი) არამგრძნობიარობა;

– დისკრეტული და უწყვეტი მახასიათებლების თანაბრად კარგი დამუშავების უნარი;

– მოდელში ცალკეული მახასიათებლის მნიშვნელობის შეფასების მეთოდების არსებობა;

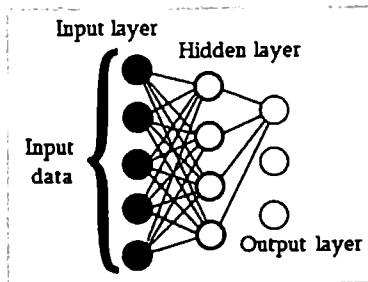
– მოდელის შესაძლებლობის შინაგანი შეფასების უნარი განზოგადებისადმი;

– მაღალი პარალელიზმი და მასშტაბურობა.

➤ **ნაკლოვანებები:**

- სირთულე არის „შემთხვევითი ტყის“ ალგორითმების მთავარი მინუსი;
- შემთხვევითი ტყეების აგება გაცილებით რთული და შრომატევადია, ვიდრე გადაწყვეტილების ხეებისა;
- შემთხვევითი ტყის ალგორითმის რეალიზაციისთვის საჭიროა მეტი გამოთვლითი რესურსი;
- პროგნოზირების პროცესი შემთხვევითი ტყეების გამოყენებით გაცილებით შრომატევადია სხვა ალგორითმებთან შედარებით.

• **ხელოვნური ნეირონული ქსელები** (*Artificial Neural Networks* -ANN) – მანქანური დასწავლის მეთოდია. ხელოვნური ინტელექტის AI) ერთ-ერთი პროგრამა. მანქანური დასწავლის ალგორითმებს შეუძლია ცხადად გაუმჯობესება, დაპროგრამების გარეშე. სხვა სიტყვებით რომ ვთქვათ, მათ შეუძლია კანონზომიერებათა პოვნა მონაცემებში და გამოიყენონ ისინი მომავალში ახალი ამოცანების გადასაწყვეტად [127]. ANN ქსელი მარტივი პროცესორების (ხელოვნური ნეირონების) დაკავშირებული და ერთმანეთთან ურთიერთმოქმედი სისტემაა (ნახ. დ4.6). ასეთი ქსელის თითოეული პროცესორი მხოლოდ იმ სიგნალებს ეხმიანება, რომლებსაც იგი პერიოდულად იღებს და რომლებსაც პერიოდულად უგზავნის სხვა პროცესორებს. ამ დიდ ქსელში, რომელსაც აქვს მართვადი ურთიერთქმედება, ეს ინდივიდუალურად მარტივი პროცესორები ერთად ახერხებს საკმაოდ რთული დავალებების შესრულებას.



ნახ. დ4.6. მარტივი ნეირონული ქსელის მოდელი

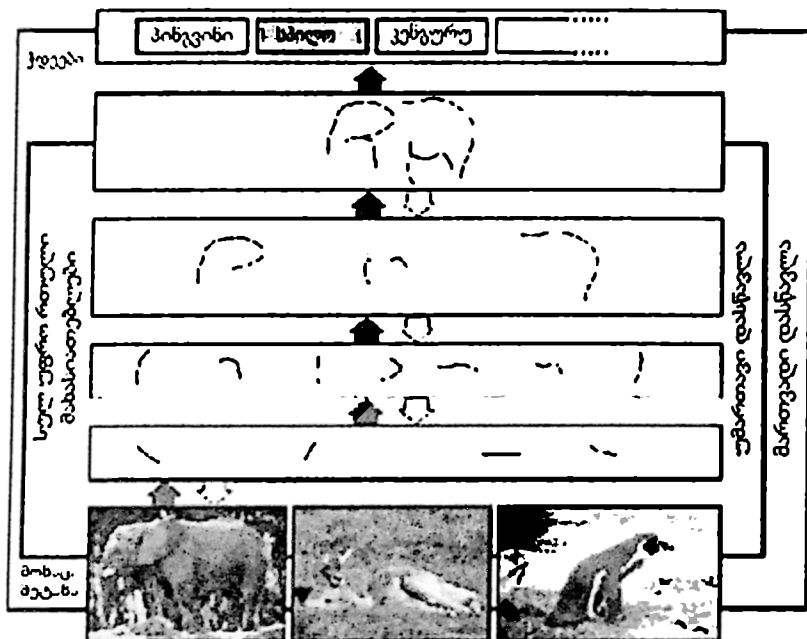
დასწავლის ალგორითმების შემუშავების შემდეგ შესაძლებელია მიღებული მოდელების პრაქტიკული მიზნებისთვის გამოყენება: პროგნოზირების, სახეთა ამოცნობის, მართვის და სხვა სახის ამოცანებში.

ხელოვნური ნეირონული ქსელის ალგორითმები მთაგონებულია ადამიანის ტვინის მიერ. ხელოვნური ნეირონები ურთიერთდაკავშირებულია და ერთმანეთთან ურთიერთობს [128]. თითოეული კავშირი შეწონილია წინა დასწავლის მოვლენებთან და მონაცემთა ყოველი ახალი შეყვანის შედეგად ხდება მეტი განსწავლა. მრავალი განსხვავებული ალგორითმი ასოცირდება ხელოვნურ ნეირონულ ქსელთან და ერთ-ერთი ყველაზე მნიშვნელოვანია „ღრმა დასწავლა“ (Deep Learning). ეს განსაკუთრებით ეხება ბევრად უფრო რთული ნეირონული ქსელების აგებას.

- *ღრმა სწავლება (Deep Learning - DL)* არის მანქანური დასწავლის მეთოდების ფართო ოჯახის ნაწილი, რომელიც ბაზირებულია ხელოვნურ ნეირონულ ქსელებზე, წარმოდგენის დასწავლით. დასწავლა შეიძლება იყოს მართვადი, ნახევრად-მართვადი ან უმართავი [129].

მთავარი განსხვავება მანქანურ დასწავლასა და ღრმა დასწავლას შორის მდგომარეობს სისტემაში მონაცემთა ასახვის მეთოდებსა და ხერხებში. მანქანური დასწავლის ალგორითმები თითქმის ყოველთვის მოითხოვს სტრუქტურირებულ მონაცემებს, ხოლო ღრმა დასწავლის ქსელები კი ეყრდნობა ხელოვნურ ნეირონულ ქსელებს (ANN - Artificial Neural Networks).

ღრმა დასწავლა, როგორც მანქანური დასწავლის ალგორითმების კლასი, იყენებს რამდენიმე შრეს, რათა საწყისი მონაცემებიდან (ქვედა დონე) თანდათანობით ამოიღოს ზედა დონისთვის მახასიათებლები (წიბოები) [130,131] (ნახ.დ4.7).



ნახ. დ4.7. იერარქიული ობიექტების ფუნქციური დასწავლის სტრუქტურული მომხილვა [123]

მაგალითად, გამოსახულებათა დამუშავებისას ქვედა შრეებს შეუძლია წიბოების იდენტიფიცირება, ხოლო ზედა შრეებს – ადამიანისთვის გასაგები ცნებების იდენტიფიცირება, როგორიცაა რიცხვები, ასოები და სახეები [123].

დ-5. DevOps მეთოდოლოგია და ინსტრუმენტული საშუალებები

Agile და DevOps მეთოდოლოგიები ასრულებს ურთიერთ-დამატებით როლებს. მაგალითად, პროგრამული სისტემების ავტომატიზებული კონსტრუირება და ტესტირება, უწყვეტი ინტეგრაცია და უწყვეტი მიწოდება [10, 134, 135]. Agile შეიძლება ჩაითვალოს, როგორც დამკვეთებსა და დეველოპერებს შორის საკომუნიკაციო ხარვეზების გადაჭრის მექანიზმი, ხოლო DevOps კი ორიენტირებულია დეველოპერებსა და IT ოპერაციებს / ინფრასტრუქტურებს შორის არსებული ხარვეზების აღმოფხვრაზე.

DevOps მეთოდოლოგია აერთიანებს პროგრამული უზრუნველყოფის დეველოპმენტის (Dev) და ინფორმაციული ტექნოლოგიების ოპერაციებს (Ops) (ნახ.3.12). მისი მიზანია პროგრამული სისტემების შექმნის ციკლის დროის შემცირება და მაღალი ხარისხის პროგრამული უზრუნველყოფის უწყვეტი მიწოდება [133]. DevOps ასევე, ყურადღებას ამახვილებს აგებული პროგრამული უზრუნველყოფის განთავსების (deployment) ამოცანებზე,

DevOps არის გუნდური მუშაობის კონცეფცია, ამიტომაც არ არსებობს ერთი კონკრეტული ინსტრუმენტი მის განსახორციელებლად. პირიქით, არსებობს რამდენიმე ინსტრუმენტის ერთობლიობა („DevOps ინსტრუმენტების ჯაჭვი“), რომლებიც კარგად ასახავს პროგრამული სისტემების დეველოპმენტის და დამკვეთებზე მიწოდების ასპექტებს [133]:

1. Coding – კოდის დეველოპმენტი და ანალიზი, საწყისი კოდის მენეჯმენტის (Source Code Management) ინსტრუმენტი, კოდების შერწყმა [136];

2. Building – უწყვეტი ინტეგრაციის (Continuous Integration) ინსტრუმენტი, კონსტრუირების სტატუსი [137];

3. Testing – უწყვეტი ტესტირების (Continuous Testoing) ინსტრუმენტი, რომელიც უზრუნველყოფს სწრაფ და დროულ უკუკავშირის ბიზნეს რისკების მიხედვით [138];

4. Packaging – არტეფაქტების რეპოზიტორია (მონაცეთა საცავი - data warehouse), აპლიკაციის წინასწარი განთავსება. საცავში ინახება პროგრამული პაკეტები და მათი მეტამონაცემები, ცვლილებათა ჟურნალი სისტემის მენეჯერისათვის [139];

5. Releasing – ცვლილებათა მენეჯმენტი, რელიზის (ვერსიის) გამოცემის დამტკიცება, აპლიკაციის ვერსიის ავტომატიზაცია Application-release automation (ARA) [140];

6. Configuring – ინფრასტრუქტურის კონფიგურაცია და მენეჯმენტი, ინფრასტრუქტურა, როგორც კოდის ინსტრუმენტი (Infrastructure as code – IaC) [141];

7. მონიტორინგი – აპლიკაციების მწარმოებლურობის (სწრაფ-ქმედების) მონიტორინგი, მუშაობის გამოცდილება საბოლოო მომხმარებელთან (Application Performance Management – APM). APM ცდილობს კომპლექსური პროგრამების შესრულების პრობლემების გამოვლენას (იხ. პარაგრაფი 2.3 და 3.2.1 – პროფაილერი). პროგრამული სისტემის მწარმოებლურობის ოპტიმიზაციის საკითხს ჩვენ კიდევ დავუბრუნდებით მომდევნო პარაგრაფში – React Hook-ის საფუძველზე.

8. და სხვ.

ავტორთა შესახებ:

გია სურგულაძე – ტექნიკის მეცნიერებათა დოქტორი, გაეროსთან არსებული „ინფორმატიზაციის საერთაშორისო აკადემიის“ წამდვილი წევრი (1994 წლიდან, IIA). საქ. ტექნიკური უნივერსიტეტის პროფესორი. ინფორმატიკის ფაკულტეტის „პროგრამული ინჟინერიის“ აკადემიური დეპარტამენტის უფროსი. 100-ზე მეტი წიგნის ავტორი (მათ შორის 27 მონოგრაფია, 22 სახელმძღვანელო, 55 დამხმარე და მეთოდური სახელმძღვანელო), 80 ელ-სახელმძღვანელო მოთავსებულია სტუ-ს ვებ გვერდზე, 50-ზე მეტი პარლამენტის ციფრულ ბიბლიოთეკაში „ივერიელი“ 300-მდე სამეცნიერო ნაშრომის ავტორი მართვის საინფორმაციო სისტემების პროგრამული ინჟინერიის, მონაცემთა რელაციური და NoSQL ბაზების, დაპროგრამების ჰიბრიდული ტექნოლოგიების, იმიტაციური მოდელირების (პეტრის ქსელებით) და სხვ. სფეროში. **წვლილი:** შეიმუშავა „საინფორმაციო საზოგადოების“ ფორმირების და განვითარების პროცესში საგამომცემლო ბიზნესის და მარკეტინგის როლი ახალი ციფრული ტექნოლოგიების ბაზაზე; ინფორმაციული სისტემის ბიზნესპროცესების დაპროექტება და კვლევა უნიფიცირებული და იმიტაციური მოდელებით: UML, ERP და CPN ბაზაზე.

დავით გოგშელიძე – აკადემიური დოქტორი ინფორმატიკაში. სტუ-ის წარჩინებული ბაკალავრი და მაგისტრი, „ინფორმატიკის“ პროგრამის დოქტორანტი (2019-2022). ერთი მონოგრაფიის, 5 სამეცნიერო სტატიის ავტორი. 2 საერთაშორისო კონფერენციის მონაწილე-მომხსენებელი. ვებ-სისტემების პროგრამისტი-დეველოპერი. 2019 წლიდან სწავლობს და მუშაობს იულისის უნივერსიტეტში (გერმანია), ასრულებს სხვადასხვა სამუშაოს ვებ-სისტემების და ღრუბლოვანი ტექნოლოგიების მიმართულებით. **წვლილი:** საგამომცემლო მარკეტინგის მომხმარებელთათვის ინოვაციური სამუშაო ინტერფეისების შემუშავების კონცეფცია, მათი მხარდაჭერი პროგრამული სისტემის (Python-ზე) აგება მანქანური დასწავლის ალგორითმების ბაზაზე. დაიცვის თემა: „საგამომცემლო მარკეტინგის ბიზნეს პროცესების ავტომატიზაცია“ (მეცნ. ხელმძღვ. პროფ. გ. სურგულაძე).

გოჩა დალაქიშვილი – სტუ-ის საინფორმაციო სისტემების დეპარტამენტის ასოც. პროფესორი. აკად. დოქტორი. 15 წიგნის და 20 სამეცნიერო სტატიის ავტორი: ენერგოსისტემების მდგრადობის პრობლემებსა და ინფორმაციული სისტემების პრაქტიკული გამოყენების და დანერგვის საკითხებზე. **წვლილი:** კონსულტანტი საგამომცემლო და სარედაქციო ბიზნესის საკითხებში ციფრული ტექნოლოგიების ბაზაზე.

რედაქცია - პროფ. გ. სურგულაძე
კომპიუტერული დაკაბადონება - ავტორების მიერ.
სტამბური გამოცემა - გოჩა დალაქიშვილი

გადაუცვარშრომობას 25.11 2023. ოფსეტური ქაღალდის ზომა
60X84 1/16. პირობითი ნაბეჭდი თაბახი 13.2, ტირაჟი 50 ეგზ.



სტუ-ს „IT კონსალტინგის სამეცნიერო ცენტრი“,
თბილისი, მ.კოსტავას 77