



Kauno technologijos universitetas

Informatikos fakultetas

**Sekų analizės taikymas informacinių sistemų naudotojų
patirties analitikai**

Baigiamasis magistro projektas.

Laura Kačinskienė

Projekto autorė

Prof. Evaldas Vaičiukynas

Vadovas

Kaunas, 2021



Kauno technologijos universitetas

Informatikos fakultetas

Sekų analizės taikymas informacinių sistemų naudotojų patirties analitikai

Baigiamasis magistro projektas

Informacinių sistemų inžinerija (6211BX009)

Laura Kačinskienė

Projekto autorė

Prof. Evaldas Vaičiukynas

Vadovas

Mokslo darbuotojas Paulius

Danėnas

Recenzentas

Kaunas, 2021



Kauno technologijos universitetas

Informatikos fakultetas

Laura Kačinskienė

Sekų analizės taikymas informacinių sistemų naudotojų patirties analitikai

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Laura Kačinskienė

Patvirtinta elektroniniu būdu

Kačinskienė, Laura. Sekų analizės taikymas informacinių sistemų naudotojų patirties analitikai. Magistro baigiamasis projektas / vadovas prof. dr. Evaldas Vaičiukynas; Kauno technologijos universitetas, informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Informacijos sistemos, Informatikos mokslai.

Reikšminiai žodžiai: sekų analizė, vizualizavimas, informacinių sistemų analitika, įvykių žurnalas, analitinės hierarchijos procesas.

Kaunas, 2021. 73 p.

Santrauka

Šio tyrimo tikslas yra palengvinti elgsenos analizę naudotojų patyrimo ekspertams skirtingų tipų informacinėse sistemose, apibrėžiant sekų analizės procesą ir realizuojant jam skirtą įrankį. Informacinių sistemų naršymo duomenys yra saugomi įvykių žurnale, o šių duomenų gavyba ir analizavimas gali atskleisti naudotojų patirtį. Darbe išgryninami sekų analizės iš įvykių žurnalo duomenų aspektai iki esminių kriterijų ir atitinkamų vizualizavimo galimybių. Išskirti kriterijai: praleisto laiko analizė, dažnų sekų kombinacijos ar taisyklės, nubyrežimų analizė, vizualizavimas ir klasterizavimas. Tikslui pasiekti sukuriama metodika atskleidžianti pagrindinius žingsnius, kuriuos reikia atlikti norint taikyti sekų analizę informacinių sistemų įvykių žurnalo duomenims. Apibendrinus 15 vertinimų, gautų iš ekspertų, analitinės hierarchijos proceso dėka, išskiriami analizės kriterijų rekomendaciniai prioritetai skirtingo tipo informacinėms sistemoms. Nustatyta, jog elektroninės komercijos sistemoms rekomenduojama atlikti nubyrežimų analizę ir realių sekų vizualizavimą. Interneto aplikacijų informacinėms sistemoms ir socialinių tinklų, forumų informacinėms sistemoms atlikti praleisto laiko tyrimą ir klasterizavimą. Reprezentacinės veiklos sistemoms atlikti nubyrežimų analizę ir dažnų sekų taisyklių radimą, o naujienų portalų ir interneto tinklaraščio IS rasti dažnų sekų kombinacijas ir dažnų sekų taisykles. Tyrimo metu, pagal sudarytą metodiką, realizuota prototipinė sistema trijų skirtingų tipų informacinių sistemų duomenims ir sudaromos rezultatus apibendrinančios ataskaitos.

Kačinskienė, Laura. Application of Sequence Analysis for Information Systems' User Experience Analytics. Master's Final Degree Project / supervisor prof. dr. Evaldas Vaičiukynas; The Faculty of Informatics, Kaunas University of Technology.

Study field and area (study field group): Information Systems, Computing.

Keywords: sequence analysis, visualizations, information systems analytics, web log data, analytic hierarchy process.

Kaunas, 2021. 73 p.

Summary

The aim of this study is to facilitate behavior analysis in different type information systems for user experience experts. This could be done by defining the process of sequence analysis and implementing the tool for it. Information systems browsing data is stored in the web logs. Extraction and analysis of this data can reveal the user experience. The paper refines an aspects of sequence analysis from web log data to essential criteria and provides suitable visualization options. Distinguished criteria: time spent analysis, frequent sequential patterns or sequential pattern rules, funnel analysis, visualizations and clustering. To achieve this goal the methodology was developed, that reveals the key steps that need to be done in order to apply sequence analysis to information systems web log data. Summarizing the 15 evaluations received from the experts, the recommended priorities of the analysis criteria for different types of information systems are distinguished using analytical hierarchy process. It has been found that in e-commerce systems it is recommended to perform funnel analysis and visualization of real sequences. Perform clustering and time spent analysis for web application or forum, social networks information systems. For corporate sites perform funnel analysis and find sequential rules, and for the news portals and blogs IS, find combinations of frequent sequences and sequential rules. According to the created methodology, the prototype system has been created, that uses three different type information systems web log data. Different reports has been created for each system where summarized results are displayed.

Turinys

Lentelių sąrašas	8
Paveikslų sąrašas	9
Santrumpų ir terminų sąrašas	10
Įvadas.....	11
1. Sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai sprendimų įvertinimas.....	13
1.1. Analizės tikslas	13
1.2. Tyrimo objektas, sritis ir problema	13
1.3. Naudotojų patirties analitika.....	14
1.3.1. Naudotojų patirties analizavimo svarba	14
1.3.2. Informacinių sistemų naudotojų patirties analizavimo būdai.....	14
1.4. Žiniatinklio gavyba.....	15
1.5. Informacinių sistemų tipai	17
1.6. Sekų analizės apibrėžimas ir panaudojimas	18
1.7. Sekų vizualizavimas	19
1.7.1. Sankio diagrama	20
1.7.2. Praleisto laiko dažnio ir apimtys diagrama.....	20
1.7.3. Piltuvėlio diagrama.....	21
1.7.4. Tinklo ir akordo diagramos	23
1.8. Sekų analizės taikymo naudotojų analizė.....	24
1.9. Esamų sekų analizės taikymo įrankių analizė	25
1.10. Darbo tikslas, uždaviniai ir siekiamas sprendimas	27
1.11. Analizės išvados	28
2. Sekų analizės sprendimo reikalavimų specifikacija ir projektas, formalus aprašas	29
2.1. Sekų analizės posistemio reikalavimų specifikacija.....	29
2.2. Sekų analizės posistemio dalykinės srities modelis	38
2.3. Formalus sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai sprendimo aprašas	39
2.4. Reikalavimų analizės apibendrinimas	41
3. Sekų analizės posistemio eksperimentinės realizacijos projektas.....	42
3.1. Apklauskos sudarymas	42
3.2. Ekspertų apklauskos rezultatai	42
3.2.1. Elektroninės komercijos sistemų analizės kriterijų prioritetai	42
3.2.2. Naujienų portalų ir interneto tinklaraščių analizės kriterijų prioritetai	43
3.2.3. Reprezentacinis veiklos tinklapių analizės kriterijų prioritetai	43
3.2.4. Interneto aplikacijų sistemų analizės kriterijų prioritetai	44
3.2.5. Socialinių tinklų ir forumų analizės kriterijų prioritetai.....	44
3.3. Apklauskos rezultatų apibendrinimas.....	45
3.4. Technologinis realizacijos sprendimas.....	46
3.5. Sekų analizės sprendimo realizacijos modelis.....	47
4. Sekų analizės realizacija ir testavimas	49
4.1. Sprendimo realizacijos ir veikimo aprašas	49
4.1.1. Duomenų tvarkymo procedūra	49
4.1.2. Puslapių subendrinimas iki tipų.....	50

4.1.3. Praleisto laiko skaičiavimas	51
4.1.4. Dažnų sekų radimas.....	52
4.1.5. Įėjimo ir išėjimo taškų radimas	52
4.1.6. Ataskaitos ruošimas.....	52
4.2. Testavimo modelis, duomenys, rezultatai	56
5. Eksperimentinis sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai tyrimas	58
5.1. Eksperimento planas.....	58
5.2. Sekų analizės taikymo skirtingų tipų informacinėms sistemoms gautos ataskaitos.....	58
5.2.1. Elektroninės komercijos IS realių sekų vizualizavimo ir nubyrėjimų ataskaita.....	58
5.2.2. Tinklaraščio IS dažnų sekų taisyklių ataskaita.....	60
5.2.3. Sąskaitų ir sutarčių administravimo IS praleisto laiko ir klasterizavimo ataskaita	61
5.3. Sprendimo taikymo rekomendacijos	63
Išvados	65
Literatūros sąrašas	66
1 PRIEDAS	69
2 PRIEDAS	72

Lentelių sąrašas

1.1 lentelė. Esamų sprendimų palyginimas.....	26
2.1 lentelė. Sekų duomenų rinkinys	33
2.2 lentelė. dažnų įvykių sekų rinkiniai, kai dažnumas $min_supp=0,5$	34
2.3 lentelė. Dažnų sekų taisyklių rinkiniai, kai patikimumas $min_conf=0,5$	34
2.4 lentelė. Analizės varianto rezultato galimas vizualizavimas.....	39
3.1 lentelė. Sekų analizės savybių prioretizavimo rezultatai elektroninės komercijos sistemoms ...	42
3.2 lentelė. Sekų analizės savybių prioretizavimo rezultatai naujienų portalų ir interneto tinklaraščių sistemoms	43
3.3 lentelė. Sekų analizės savybių prioretizavimo rezultatai reprezentacinės veiklos svetainėms ...	43
3.4 lentelė. Sekų analizės savybių prioretizavimo rezultatai interneto aplikacijų sistemoms	44
3.5 lentelė. Sekų analizės savybių prioretizavimo rezultatai socialinių tinklų ir forumų informacinėms sistemoms	45
3.6 lentelė. Sekų analizės prioretizavimo rezultatai skirtingoms informacinėms svetainėms	45
4.1 lentelė. Pavyzdinė sesijų įvykių žurnalo duomenų struktūra	49
4.2 lentelė. Pavyzdinė aplankytų puslapių įvykių žurnalo duomenų struktūra.....	50
4.3 lentelė. Pavyzdinė aplankytų tipų ir praleisto laiko duomenų struktūra	51
4.4 lentelė. Pavyzdinė praleisto laiko duomenų struktūra pagal apimties procentą.....	51
4.5 lentelė. Pavyzdinė skaitytų straipsnių duomenų struktūra	52
4.6 lentelė. Pavyzdinė įėjimo išėjimo duomenų struktūra	52
4.7 lentelė. Funkcinio testavimo scenarijai	56
4.8 lentelė. Nefunkcinio testavimo scenarijai	57

Paveikslų sąrašas

1.1 pav.	Žiniatinklio gavybos kategorijos ir struktūra [7].....	15
1.2 pav.	Sankio diagramos pavyzdys „Google Analytics“ įrankyje	20
1.3 pav.	Sekų dažnio grafikas sugeneruotas TraMineR paketu	21
1.4 pav.	Piltuvėlio diagrama rankiniu būdu generuota „Google Analytics“ platformoje	22
1.5 pav.	Piltuvėlio diagrama sugeneruota komerciniame įrankyje „Hotjar“	22
1.6 pav.	Akordo diagrama sugeneruota R įrankiu	23
1.7 pav.	Tinklo diagrama sugeneruota R įrankiu	23
2.1 pav.	Sekų analizės tyrimo panaudojimo atvejų modelis	29
2.2 pav.	PA „Paruošti duomenis ataskaitai“ veiklos diagrama	30
2.3 pav.	PA „Atrinkti tinkamus duomenis“ veiklos diagrama	31
2.4 pav.	PA „Subendrinti nuorodas iki puslapių tipų“ veiklos diagrama	32
2.5 pav.	PA „Rasti dažnas sekas ar taisykles“ veiklos diagrama	35
2.6 pav.	PA „Klasterizuoti rastas sekas į grupes“ veiklos diagrama.....	36
2.7 pav.	PA „Apskaičiuoti trukmę praleistą puslapiuose“ veiklos diagrama.....	37
2.8 pav.	PA „Rasti įėjimo/ išėjimo taškus ir nubyrijimus“ veiklos diagrama	37
2.9 pav.	Įvykių žurnalo duomenų modelis (žalia spalva) ir papildymas esybėmis (geltona spalva)	38
2.10 pav.	Bendras informacinių sistemų naudotojų sekų analizės metodikos procesas	40
3.1 pav.	SQL serveryje vykdomos analitikos modelis [42]	46
3.2 pav.	Analizės rezultatų panaudojimas Power BI įrankyje	47
3.3 pav.	Diegimo diagrama	47
4.1 pav.	Vizualizacijoms ir nubyrijimams reikalinga duomenų struktūra, skirta elektroninės komercijos informacinėms sistemoms, veiklos analitikos įrankyje "Power BI"	53
4.2 pav.	Praleisto laiko skaičiavimui ir klasterizavimui reikalinga duomenų struktūra, skirta interneto aplikacijų ir socialinių tinklų, forumų informacinėms sistemoms, veiklos analitikos įrankyje „Power BI“	54
4.3 pav.	Nubyrijimų ir taisyklių radimui reikalinga duomenų struktūra, skirta reprezentacinės veiklos informacinėms sistemoms, veiklos analitikos įrankyje „Power BI“	55
4.4 pav.	Dažnų sekų ir taisyklių radimui reikalinga duomenų struktūra, skirta interneto tinklaraščių ir naujienų portalų informacinėms sistemoms, veiklos analitikos įrankyje „Power BI“	56
5.1 pav.	Realių sekų vizualizavimas per sankio diagrama.....	59
5.2 pav.	Nubyrijimų analizės vizualizavimas.....	60
5.3 pav.	Gautų dažnų sekų ir taisyklių rezultatų vizualizavimas	61
5.4 pav.	Praleisto laiko analizės vizualizavimas	62
5.5 pav.	Praleisto laiko analizės rezultatai suskirstyti į klasterius	63

Santrumpų ir terminų sąrašas

Terminai:

Žiniatinklio gavyba (angl. *Web Mining*) – duomenų gavybos metodų pritaikymas žiniatinklio sistemoms.

WUM (angl. *Web Usage Mining*) – metodologija apimanti duomenų gavybą iš internete naršymo metu sukauptų duomenų ir jų analizę.

HTML (angl. *Hypertext Markup Language*) - tai kompiuterinė žymėjimo kalba, naudojama pateikti turinį internete.

Hipersaitas (angl. *Hyperlink*) – hiperteksto nuoroda, turinti matomą dalį ir po ja paslėptą komponentą – nurodomo objekto adresą, pagal kurį naršyklė arba kuri nors kita hipertekstu operuojanti programa randa ir parodo nuorodos rodomą dokumentą.

BDAR – Bendrasis duomenų apsaugos reglamentas.

IS – Informacinė sistema

Scenarijus (angl. *Script*) – programa, sudaryta iš interpretavimui skirtų komandų.

RegExp (angl. *Regular Expressions*) – reguliarios išraiškos, konkrečios simbolių sekos, plačiai arba siaurai atitinkančios duomenų šablonus.

Paspaudimų srautas (angl. *clickstream*) – įrašas, kuriame yra duomenys apie svetainės vartotojo paspaudimus kompiuterio ekrane per pelę ar jutiklinę dalį. Šios rūšies informacija suteikia vartotojo veiklos pėdsaką su informacija kokios nuorodos buvo paspaustos.

AHP (angl. *Analytic Hierarchy Process*) – yra sprendimų priėmimo metodas, kuris apima vertinimo kriterijų hierarchijos sudarymą, santykinų svorių priskyrimą šioms kriterijoms, alternatyvų palyginimą kiekvienam kriterijui bei visų alternatyvų apibendrintų reitingų sudarymą.

SPADE (angl. *Sequential Pattern Discovery using equivalence classes algorithm*) – dažnų sekų gavybos algoritmas

SQL (angl. *Structured Query Language*) – struktūrizuota užklausų kalba, naudojama komunikuojant su duomenų baze.

Įvadas

Šis magistro baigiamasis darbas priklauso informacinių sistemų inžinerijos studijų programai.

Darbo problematika ir aktualumas

Pasaulyje vis labiau populiarėjant informacinių sistemų naudojimui, darosi svarbu suprasti naudotojų elgseną, ar naudotojai patenkinti svetainės galimybėmis, ieškoma kaip atrasti kuo geresnes pritraukimo į svetainę priemones. Svarbu, ne tik informacijos pateikimas, bet ir tai, kaip ta informacija pasiekama. Siekiant tai sužinoti, reikia analizuoti naudotojų elgseną. Šiai analizei galima panaudoti duomenis, sukauptus įvykių žurnale. Vis daugiau įmonių susiduria su problema, jog kaupiamų duomenų kiekiai didėja, auga jų įvairovė ir naudotojų skaičius. Darosi vis sudėtingiau tvarkyti, analizuoti ir vizualizuoti duomenis, rengti pagal juos ataskaitas, atlikti prognozes siekiant priimti verslo sprendimus. Taip pat nėra aišku kokius įrankius ar metodus pritaikyti, siekiant tinkamai išgauti norimą informaciją pagal informacinės sistemos tipą.

Šiame darbe problemos analizei naudojama realios informacinės sistemos, kurios įvykių žurnale kaupia duomenis apie naudotojų naršymo sesijos metu atliktus veiksmus. Surinkti duomenys yra tik panaudojami statistikai ar analizuojamos pavienės sesijos, bet bendrai įvykių sekos nėra analizuojamos.

Tyrimas atliekamas remiantis duomenų analizės kriterijų ir vizualizavimo metodų parinkimu. Tyrimo objektas – sekų analizė ir vizualizavimas iš įvykių žurnale kaupiamų naudotojo elgsenos duomenų.

Šio darbo tyrimo sritis – sukauptų naudotojų elgsenos duomenų analizė, duomenų taikymas, vizualizavimo metodai ir įrankiai.

Numatomas problemos sprendimas – klientų atliktų naršymo sekų analizė ir rezultatų vizualizavimas veiklos analitikos įrankiu.

Darbo tikslas ir uždaviniai

Šio tyrimo tikslas – palengvinti elgsenos analizę naudotojų patyrimo ekspertams skirtingų tipų informacinėse sistemose, apibrėžiant sekų analizės procesą ir realizuojant jam skirtą įrankį.

Atlikta analizė iš įvykių žurnalo duomenų leistų nustatyti kuriuos scenarijus reiktų patobulinti pagerinant naudotojų patyrimą, padėtų nustatyti, kurie puslapiai yra labiausiai lankomi ir kokią laiko trukmę čia praleidžia naudotojai. Taip pat rezultatai leistų identifikuoti problemas ar rasti scenarijus, kurie turi būti optimizuoti.

Norint pasiekti tikslą, iškeliami uždaviniai:

1. Išanalizuoti esamas sekų analizės tyrimo priemones, metodus, įrankius;
2. Išanalizuoti vizualizacijos būdus tinkamus analizės rezultatams pateikti;
3. Apžvelgti sekų analizės taikymo procesą ir žingsnius, kuriuos reikia atlikti;
4. Sudaryti ekspertinę apklausą, kuri padės išskirti analizės kriterijus pagal informacinės sistemos tipą;
5. Realizuoti prototipinę sistemą skirtingoms informacinėms sistemoms parenkant analizės kriterijus ir skirtingas vizualizacijas pagal informacinės sistemos tipą.
6. Pateikti rekomendacijas sekų analizės taikymui ir apibendrinti rezultatus.

Darbo rezultatai ir jų svarba

Darbe aprašoma sekų analizės taikymo įvykių žurnalo duomenims svarba. Išskiriami duomenų analizės kriterijai ir vizualizavimo metodai, kurių taikymui sukurta metodika. Tinkamam analizės kriterijų parinkimui pagal informacinės sistemos tipą, atlikta ekspertų apklausa, kurios rezultatai padėjo nustatyti prioritetus. Darbe buvo kuriamas sprendimas pagal sukurta metodiką skirtingų tipų informacinių sistemų duomenims. Sukurtas rezultatas bus pateikiamas naudotojų patyrimo ekspertams kaip įrankis padedantis atlikti išvalgas.

Darbo struktūra

Darbas suskirstytas į penkis skyrius. Pirmajame skyriuje aprašoma sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai svarba, aptariami analizės metodai ir vizualizavimo galimybės. Antras skyrius skirtas kuriamo sprendimo reikalavimams apibendrinti ir sekų analizės taikymo metodikos sudarymui. Trečiajame skyriuje aprašoma ekspertų apklausos sudarymas ir prioritetų išskyrimas pagal gautus rezultatus. Taip pat aprašomas technologinis realizacijos sprendimo planas. Ketvirtajame skyriuje aprašoma sekų analizės posistemio sprendimo realizacija ir testavimas. Penktame skyriuje pateikiamos eksperimentiškai kurtos atskaitos skirtingų tipų informacinėms sistemoms, kuriose išpildomi skirtingi analizės variantai. Taip pat, pateikiamos sprendimo taikymo rekomendacijos. Paskutiniame skyriuje pateikiami rezultatų apibendrinimai ir išvados.

1. Sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai sprendimų įvertinimas

Analizės dalyje pateikta sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai galimi sprendimai ir jų vertinimas.

1.1. Analizės tikslas

Analizės metu siekiama geriau suprasti kaip yra pritaikoma sekų analizė, kaip galima analizuoti kaupiamus naudotojų naršymo duomenis iš informacinių sistemų įvykių žurnalo įrašų, kokius metodus ar įrankius taikyti. Siekiama atrasti analizės būdus, kurie padėtų išaiškinti naudotojų patirtį.

Analizės tikslui pasiekti išskelti uždaviniai:

1. Aptarti kas yra naudotojų patirties analitika, jos svarba ir analizavimo būdai;
2. Išanalizuoti kas yra žiniatinklio gavyba ir kokie duomenys gali būti pritaikomi tyrimui;
3. Išanalizuoti žiniatinklio gavybos svarbą, tikslą ir pritaikymą skirtingoms informacinėms sistemoms;
4. Aptarti sekų analizės apibrėžimą, panaudojimą ir galimus analizės variantus įvykių žurnalo duomenims;
5. Aptarti galimus vizualizavimo būdus, skirtus sekoms atvaizduoti;
6. Aptarti, kokiems vartotojams, ši analizė būtų svarbi;
7. Išanalizuoti galimus įrankius sekų analizei vykdyti.

Pirmiausiai svarbu aptarti naudotojų patirties analitikos svarbą informacinėse sistemose, išanalizuoti galimus matavimo variantus informacinių sistemų duomenims. Taip pat svarbu nustatyti, kokioms informacinėms sistemoms ši analizė yra tinkama. Toliau svarbu apžvelgti kokius duomenis galima išgauti ir taikyti, kaip sekų analizė čia gali būti pritaikoma. Galiausiai, reikalinga išnagrinėti kokiais metodais ar įrankių pagalba galima taikyti skirtingus analizės variantus, kaip vizualiai apiforminti analizės rezultatus.

1.2. Tyrimo objektas, sritis ir problema

Pasaulyje vis labiau populiarėjant informacinių sistemų naudojimui, darosi svarbu suprasti naudotojų elgseną, ar naudotojai patenkinti svetainės galimybėmis, ieškoma kaip atrasti kuo geresnes pritraukimo į svetainę priemones. Svarbu, ne tik informacijos pateikimas, bet ir tai, kaip ta informacija pasiekama. Siekiant tai sužinoti, reikia analizuoti naudotojų elgseną. Šiai analizei galima panaudoti duomenis, sukauptus įvykių žurnale. Vis daugiau įmonių susiduria su problema, jog kaupiamų duomenų kiekiai didėja, auga jų įvairovė ir naudotojų skaičius. Darosi vis sudėtingiau tvarkyti, analizuoti ir vizualizuoti duomenis, rengti pagal juos ataskaitas, atlikti prognozes siekiant priimti verslo sprendimus. Taip pat nėra aišku kokius įrankius ar metodus pritaikyti, siekiant tinkamai išgauti norimą informaciją pagal informacinės sistemos tipą.

Šiame darbe problemos analizei naudojama realios informacinės sistemos, kurios įvykių žurnale kaupia duomenis apie naudotojų naršymo sesijos metu atliktus veiksmus. Surinkti duomenys yra tik panaudojami statistikai ar analizuojamos pavienės sesijos, bet bendrai įvykių sekos nėra analizuojamos.

Tyrimas atliekamas remiantis duomenų analizės kriterijų ir vizualizavimo metodų parinkimu. Tyrimo objektas – sekų analizė ir vizualizavimas iš įvykių žurnale kaupiamų naudotojo elgsenos duomenų.

Šio darbo tyrimo sritis – sukauptų naudotojų elgsenos duomenų analizė, duomenų taikymas, vizualizavimo metodai ir įrankiai.

1.3. Naudotojų patirties analitika

Turint bet kokios srities verslą, vienas svarbiausių principų, kurių reikia įsisavinti ir stengtis nuolat taikyti tai – vartotojo patirties gerinimas. Neatsižvelgiant ar tai populiarus elektroninė parduotuvė, ar komunikacija socialiniuose tinkluose ar reprezentacinis tinklapis, tai, kaip vartotojas jaučiasi verslą pristatančioje platformoje tiesiogiai lemia jo požiūrį į patį verslą. Žinoma talpinamas turinys yra svarbiausias aspektas, tačiau reikia atsižvelgti ar naudotojams ta informacija yra lengvai ir sklandžiai pasiekama.

1.3.1. Naudotojų patirties analizavimo svarba

Įmonėms siekiant būti pelningoms ir išlaikyti lojalius vartotojus, svarbu gilintis ne tik į pavienius sandorius ar pardavimus, o į ilgalaikių santykių su vartotojais kūrimą, bei poveikį jų elgsenai darančių veiksnių nustatymą. Teigiama naudotojų patirtis padeda kurti naudotojų lojalumą [1]. Tyrimai rodo, kad įmonės, turinčios gerą naudotojų patirtį padidina pardavimus, pritraukia daugiau naujų klientų, sumažina klaidų skaičių ir išlaiko didesnę dalį esamų klientų [2]. Nėra vieningos nuomonės ar būdo įvertinti ar naudotojo patirtis yra teigiama. Galima rasti ne vieną patarimą ar gaires kokį dizainą pasirinkti, kaip dėlioti navigacijos elementus ar būdus išvengti naudotojų pasimetimo. Deja, nėra vieno patarimo ar metodo, kuris būtų tinkamas skirtingiems naudotojams ar skirtingoms sistemoms ir tinklalapiams. Net pritaikius skirtingus patarimus ar atsižvelgus į gaires, reikia rasti būdą pamatuoti ar šie metodai veikia tiriamoje sistemoje. Taip pat kaip ir patarimų, matavimo būdų patarimų yra apstu. Nėra vieno mato ar būdo pamatuoti, tačiau skirtinguose šaltiniuose pabrėžiama, jog reikia stengtis matuoti ne pavienius puslapius ar pavienius matavimus, o atsižvelgti į visumą [3]. Stengtis analizuoti visą naudotojo naršymo kelią, rinkti duomenis ir sieti juos su naudotojų profiliais.

1.3.2. Informacinių sistemų naudotojų patirties analizavimo būdai

Informacinių technologijų plėtra padidino naudotojų patogumą ir paskatino informacinių sistemų naudojimo poreikį. Nustatyta, jog tam tikri veiksniai lemia naudotojų patirtį virtualioje aplinkoje. Tai – naršymo paprastumas, platformos greitis ir funkcionalumas, turinys ir patikimumas [4]. Siekiant sužinoti ar naudotojas yra patenkintas sistemos galimybėmis, dažnai pasitelkiamos išsamios apklausos ar atsižvelgiama į išsakytas naudotojų pastabas. Taip pat įvedamos skirtingos metrikos leidžiančios įvertinti naudotojų patirtį per sukauptus naršymo duomenis [5]. Išskiriamos keturios kategorijos, grupuojančios naudotojo patirtį:

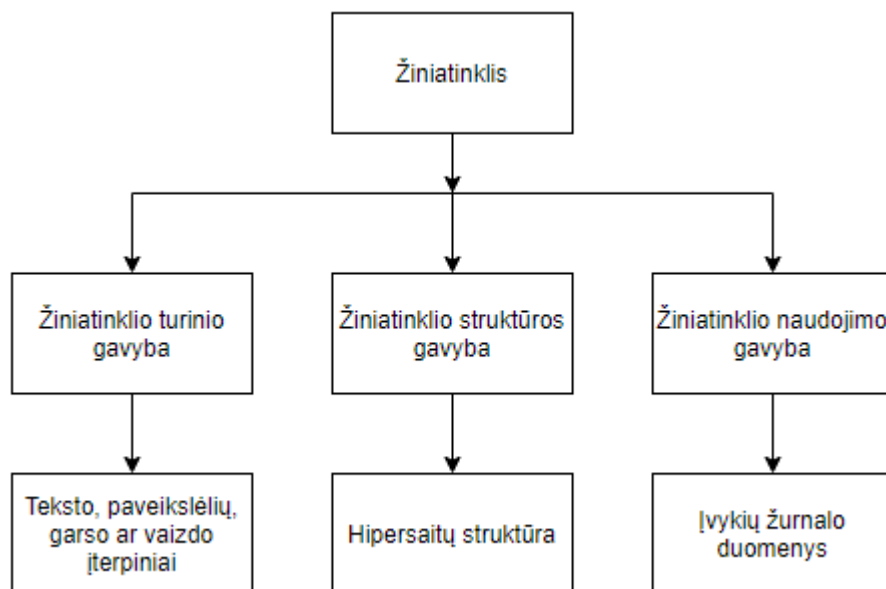
- **Kiekybiškai įvertinami matai**, tokie kaip atmetimo rodiklis (angl. *bounce rate*), paspaudimų rodiklis (angl. *click through rate*), užduoties įvykdymo sėkmė (angl. *task/goal success*), klaidų rodiklis (angl. *error rate*), užduoties įvykdymo greitis (angl. *task completion time*), nubyrėjimo rodiklis (angl. *drop-off rate*), pasimetimo rodiklis (angl. *lostness factor*), vidutinis sesijos ilgumas.
- **Naudotojų palikti įvertinimai**, kurių rezultatai apžvelgiami per Likerto skalę, sentimentų analizę, NPS matavimą (angl. *Net Promoter Score*), sistemos patogumo skalę (angl. *system usability scale*) ar kitokias apklausas.
- **Kategoriškai įvertinami matai** tokie kaip A/B testai, paspaudimų žemėlapiai (angl. *heatmaps*), paspaudimų srauto sekos (angl. *clickstream sequences*), geografinės vietos ar prietaiso tipo nustatymas.

- **Su emocijomis susiję matavimai**, kurie vertinami atsižvelgiant į emociją balse (angl. *audio emotion recognition*), širdies ritmo pakitimus (angl. *Heart rate variability*) ar odos laidumą (angl. *skin conductance*).

Įvardytiems kriterijams pamatuoti, reikalingi skirtingo formato duomenys. Iš karto galima numatyti, jog su emocijomis susijusiems matavimams reikalinga speciali įranga ir naudotojų sutikimas rinkti tokio tipo duomenis. Naudotojų paliktiems vertinimams reikalingi iš anksto sudaryti klausimynai, ar paliktų atsiliepimų apžvalga. Šio tipo įvertinimai negali atspindėti visų naudotojų patyrimo, o tik tam tikros imties rezultatus. Norint išgauti A/B testų rezultatus, reikalingas specialus pasiruošimas, o paspaudimų žemėlapiai, nėra paprastai sugeneruojami be tam tikros papildomos programinės įrangos. Rinkoje apstu paketų galinčių matuoti skirtingus kriterijus ir atlikti duomenų analizę, tačiau nenusižengus BDAR įstatymui, be vartotojo sutikimo papildomų paketų naudoti negalima [6]. Yra išskiriamas duomenų gavybos būdas – žiniatinklio gavyba. Šis būdas leidžia panaudoti naudotojo naršymo metu sukauptus duomenis analitikai.

1.4. Žiniatinklio gavyba

Žiniatinklio gavyba – vienas iš duomenų gavybos metodų, skirtų išgauti naudingą informaciją ir žinias iš tinklalapių duomenų. Šis metodas apima tinklapių turinio duomenų gavybą (angl. *Web Content Mining*), kuris tiria tekstus, dokumentus ar multimediją, taip pat tinklapių struktūros gavybą (angl. *Web Structure Mining*), kuri tiria hipersaito struktūrą. Dar žiniatinklio gavyba apima WUM, kuri tiria įvykių žurnalo įrašus [7]. Visos trys žiniatinklio gavybos kategorijos yra tiriamos siekiant pagerinti žiniatinklio paslaugas. Žiniatinklio gavybos skirstymas ir struktūra pateikiama 1.1 pav.



1.1 pav. Žiniatinklio gavybos kategorijos ir struktūra [7]

Kaip ir minėta, kiekviena iš šio metodo kategorijų apima skirtingus duomenų formatus. Žiniatinklio turinio gavyba išgauna ir tiria puslapio parametrus ir charakteristikas tiek struktūrizuota, tiek nestruktūrizuota forma iš HTML parašyto kodo ir jame naudojamų skirtingo tipo duomenų formatų (teksto, paveikslėlių, garso ar vaizdo įterpinių, metaduomenų ir pan.) ir HTML kalboje naudojamų parametrų. Žiniatinklio struktūros gavyba susijusi su hipersaitų analizavimu ir jų struktūros tyrimu.

Taip išgaunama informacija apie pačio puslapio struktūrą, žiūrint į hipersaito seką ir iš jo matomą navigacijos kelią. Trečioji kategorija apima informacijos gavybą, kuri kaupiama įvykių žurnale. Čia atsiduria įrašai tam tikru struktūrizuotu formatu [7], kuriuose kaupiama įvairi sesijos metu gauta informacija, tokia kaip laiko žyma, IP adresai, HTTP būsenos kodas, aplankyto puslapio nuoroda. Turint keletą įrašų, kurie buvo atlikti to paties vartotojo, atsižvelgiant į laiko žymes ir hipersaitus, galima matyti naršymo kelią. Taip iš įvykių žurnalo galima tirti naudotojo elgseną ir rasti dažnai pasikartojančius naršymo šablonus [8].

Augant elektroninei prekybai ar internetinei rinkodarai, atsiranda poreikis suprasti naudotojus. Anksčiau pakakdavo iš istorinių duomenų surinkti ir nustatyti populiariausius puslapius ir apskaičiuoti apsilankymų skaičių, gauti duomenis apie vidutiniškai praleistą laiką jame [9]. Didėjant sukauptų duomenų kiekiui ir platesnių analizės rezultatų poreikiui, analizuoti tokius duomenis darosi negana, o kur kas aktualiau yra tirti šiuos duomenis iš naudotojo perspektyvos [8]. Naudotojų naršymų duomenys tampa svarbiausiu šaltiniu. Iš istorinių duomenų bandoma pažinti naudotoją. Taip yra siūlomos tik jam aktualios rekomenduojamos prekės ar priedai prie prekių, skelbiama klientui aktuali reklama. Taip pat bandoma suprasti kaip pakeisti puslapį, siekiant padaryti jį patogesnę ir labiau patrauklų klientams [10]. Organizacijos, siekiančios pažinti savo klientus geriau, žiniatinklio gavyboje įprastai žinomus žurnalo duomenis sieja su naudotojų profiliais. Tačiau net ir susiejus naudotojų profilius su naršymo duomenimis, analizė ir jos rezultatų pateikimas vis tiek išlieka sunkia užduotimi.

Dažnai įmonės, norėdamos pažinti savo naudotojus, renkasi komercines papildomas programas [11], kurios gali rinkti, ne tik pirkimų duomenis ar laiką, praleistą naršant, bet ir pelės judėjimus (angl. *mouse flow*), paspaudimų vietas (angl. *heatmaps*). Įsigaliojus Bendrajam duomenų apsaugos reglamentui (BDAR) [6], daugelis tokių komercinių analitikos paketų negali atlikti visos analizės, nes gali rinkti duomenis tik iš vartotojų, kurie davė sutikimą rinkti trečiųjų šalių slapukus. Vartotojui nesureagavus su slapukais ar nesutikus rinkti informacijos analitikai ar rinkodaros tikslais, komerciniai paketai lieka be didelės dalies duomenų srauto [12].

Siekiant nepažeisti BDAR, bet tuo pačiu gauti viso srauto duomenis, dažnai yra neišnaudojami duomenys, sukaupti įvykių žurnaluose. Juose atsispindi įvairi sesijos metu gauta informacija [13]. Siekiant atlikti naudotojo veikimo analizę, atliekami keli svarbūs žingsniai: iš įvykių žurnalo išgaunamos kiekvienos sesijos metu lankytos portalo puslapių nuorodos, jos apdorojamos, suskirstant nuorodas pagal puslapio tipus, atrandamos pasikartojančios sekos ir atliekama jų analizė.

Pagrindinis žiniatinklio naudojimo gavybos (WUM) tikslas yra ištirti naudotojų navigaciją ir naudotojų patirtį. Yra skirtingų analizės kriterijų kuriuos galima taikyti. Galima išskirti ir keletą pagrindinių priežasčių [10] kodėl atliekama ši gavyba:

- **Tinklalapio turinio personalizavimas.** Galima pritaikyti WUM techniką personalizuojant svetaines, atsižvelgiant į vartotojo profilį ir elgesį. Suasmeninimas yra svarbus kuriant rinkodaros strategijas ar automatizuotai produktų reklamai potencialiems klientams. Be to, žiniatinklio naudojimo gavyba siekiama pateikti tokį turinį, kuris bus atrandamas greičiau ir aplankomas dažniau.
- **Sistemos tobulinimas.** Galima panaudoti žiniatinklio naudojimo gavybos rezultatus interneto serverių ir žiniatinklio programų veikimui tobulinti. Žinant detalų srautą, gali būti parengtos sistemos gerinimo politikos ir strategijos, planuojamos apkrovos ir duomenų paskirstymas.

- **Saugumas.** Žiniatinklio naudojimo gavyba gali suteikti naudingų modelių įsilaužimams, bandymams įsilaužti ar sukčiavimui aptikti.
- **Svetainės projektavimo palaikymui.** Tinkamumas naudoti yra vienas iš svarbiausių klausimų projektuojant ir diegiant interneto informacines sistemas. Žiniatinklio naudojimo gavybos rezultatai gali suteikti dizaineriams informacijos apie naudotojų elgseną, kuri padeda priimti sprendimus dėl bet kokio programos turinio ir struktūros pertvarkymo. Be to, kai kurios priemonės automatiškai keičia svetainės struktūrą pagal atrastus naudojimo modelius.
- **Pralaidumui ir nubyrėjimui rasti.** Galima panaudoti žiniatinklio naudojimo gavybos rezultatus, tikri pradžios ir pabaigos puslapius, taip sužinant didžiausius įėjimo ir išėjimo puslapius. Taip pat imant iš eilės einančių hipersaitų seką tirti nubyrėjimus.

Apžvelgus sukauptus duomenis, galima bandyti analizuoti duomenis, išskiriant skirtingus kriterijus. Skirtingi kriterijai yra parenkami pagal tyrimo tikslą. Tam pačiam tikslui pasiekti gali būti reikalingas skirtingas kiekis analizės kriterijų ar matų [14]. Galima bandyti atlikti visus sekų analizei tinkamus analizės kriterijus, o vėliau gilintis ir bandyti išskirti tik reikšmingiausius rezultatus. Tokiu atveju tektų atlikti nemažą dalį duomenų paruošimo, o tam tikri rezultatai gali būti visai nepanaudojami. Žinant analizės kriterijaus parinkimo rekomendaciją pagal informacinės sistemos tipą, būtų sutaupoma analizės laiko, nei išpildant visus kriterijus.

1.5. Informacinių sistemų tipai

Pagal saityno informacinės sistemos pobūdį, žiniatinklio gavybą galima pritaikyti skirtingai. Vieno pobūdžio sistemoms išskiriamas bendras bruožas ar tikslas. Pavyzdžiui, elektroninės komercijos portalai ar interneto aplikacijų sistemos išsiskiria tuo, jog turi bent vieną tipišką ir tikriausiai mažai išsiskiriantį kelią nuo kitų tokio pat pobūdžio sistemų. Tai paskatina vartotojus naršyti ir sklandžiai atlikti pagrindinį sistemos tikslą – įsigyti prekę, paslaugą. Pagal paskirtį galima išskirti ir daugiau saityno tipų [15]. Dažniausiai sutinkami tipai:

- **Elektroninės komercijos sistemos.** Šių sistemų struktūroje yra daugybė bendrų ir klientams įprastų bruožų [16]. Pirkėjams įprasta peržiūrėti produktų katalogą, detaliau paskaityti apie produktą, pridėti ar pašalinti prekes iš krepšelio, atlikti mokėjimą. Šioms funkcijoms atlikti, įprasta turėti produktų sąrašo, krepšelio operacijų, pirkimo patvirtinimo puslapius. Kiekviena iš elektroninės komercijos platformų turi bent vieną tipišką ir tikriausiai mažai išsiskiriantį kelią, kuris paskatina vartotojus naršyti ir sklandžiai atlikti pirkimus. Dažniausiai ir yra tiriamas šis tipinis kelias, tiriamas nubyrėjimas. Įvertinus sekas pagal sukauptus visus lankymo duomenis, galima tikri sekos pradžios ir pabaigos puslapius, klientų nubyrėjimus iki apmokėjimo atlikimo. Vieno pirkėjo duomenys negali atspindėti kažko labai prasmingo. Bet jei daugelis pirkėjų išvyksta prieš eidami mokėti, tai gali reikšti, kad jiems nepatinka tai, ką jie įdėjo į savo pirkinių krepšelius arba pirkimo procesas yra pakankamai sudėtingas ir naudotojai verčiau palieka sistemą nei užbaigia žingsnius.
- **Reprezentacinis veiklos tinklapis.** Jame dažnai galima išskirti tam tikrą informaciją, kurią rasime daugumoje šio tipo svetainių. Šio tipo svetainių tikslas yra teikti informaciją ir kontaktus apie tam tikrą vykdomą veiklą. Nesvarbu ar tai verslo, ar valstybinės įstaigos tinklapis, šio tipo svetainėse siekiama pateikti informaciją kuo aiškiau. Dažnai šio tipo svetainės neišsiskiria hipersaitų gausą, tačiau naršymo šablonų ir naudotojų grupių identifikavimas gali būti pakankamai svarbus.
- **Interneto aplikacijų sistemos.** Šio tipo sistemų paskirtis yra atlikti tam tikrą funkciją išpildant tam tikrą procesą. Pakankamai sunku įvardinti kokios tos sistemos gali būti, tačiau

dažnai tai yra administracinės ar valdymo sistemos. Pavyzdžiui, banko, problemų registravimo ar dokumentų valdymo sistemos. Šios sistemos panašiai kaip ir elektroninės komercijos sistemos turi tam tikrus tipinius naršymo šablonus. Sistemose siekiama suteikti naudotojams galimybę kuo greičiau ir sklandžiau atlikti veiksmus.

- **Naujienų portalai ir tinklaraščių tinklapiai.** Šio tipo portaluose, dažniausiai nėra išskiriamas vienas bendras kelias ar tikslas, o dažniau aptinkami tam tikra tematika pateikti skirtingi straipsniai. Sunku išskirti naršymo šablonus, tačiau šio tipo sistemose galima tirti kartu aplankomų straipsnių kombinacijas neanalizuojant įvykių sekos. Taip galima išgauti turinio personalizavimo pasiūlymus pasitelkus kitų naudotojų aplankytų straipsnių kombinacijas.
- **Socialiniai tinklai ir forumai.** Šios svetainės sukurtos taip, kad priimtų vartotojų registracijas ir leistų jiems turėti profilį svetainėje. Šiems leidžiama dalytis turiniu su kitais vartotojais. Dažnai šios paskirties sistemose siekiama kuo labiau įtraukti naudotoją, jog šis praleistų kuo daugiau laiko. Šio tipo svetainėse nėra išskiriama plati hipersaitų gausa ar vienas tipinis naršymo šablonas, tačiau gali būti įdomu nustatyti naršymo šablonus ar identifikuoti grupes.

Išskyrus informacines sistemas pagal jų tipą ir bandant parinkti tinkamą analizės variantą, galima išgauti analizės taikymo rekomendacijas. Taip, būtų sutelkiama daugiau dėmesio skirtingų analizės variantų pritaikymui ir panaudojimui. Vienas analizės kriterijus gali būti visai neaktualus vieno tipo informacinėms sistemoms, kuomet kitas gali būti labai aktualus, o rezultatai sutelkti daug įžvalgų. Net ir numanius ką galima analizuoti, reikia aiškiai išskirti visus kriterijus tinkamus sekoms analizuoti. Taip pat praverstų ekspertų nuomonės įtraukimas nustatant tuos kriterijus skirtingų tipų informacinėms sistemoms.

1.6. Sekų analizės apibrėžimas ir panaudojimas

Sekų analizė apibrėžiama kaip struktūros arba elementų nagrinėjimas, pagal tai kaip tai yra išdėstyta laiko ar erdvės atžvilgiu, kartu apimant ir technikas kaip tai padaryti [17]. Esminė sekų analizės savybė, kuri išskiria šią analizę nuo kitų esamų – tai duomenų ėmimas, kaip sekos, ne kaip atskirų duomenų elementų. Seka – tai kažkokia tvarka išrašytų elementų E_i aibė $S[5]$, kur elementų indeksas i nurodo nario tvarką: $S = [E_1, E_2, \dots, E_i]$. Pagrindinis sekų analizės pritaikymo tikslas – tai rasti dažniausiai pasikartojančias sekas, stebėti besikeičiančias jų būsenas ar tarpusavyje palyginti sekas. Sekos pirmiausiai buvo analizuojamos biologijoje, siekiant suprasti molekulių funkcijas. Jau 1951 m. pradėtas naudoti sekų analizės apibrėžimas, kai Fredis Sangeris [18], tyrinėdamas insulino baltymų sekas, prisidėjo prie pirmojo sėkmingo DNR genomo sekos nustatymo. Ne tik biologijoje, bet ir daugelyje kitų skirtingų sričių atliekami veiksmai, kuriuos mes vadiname sekomis ir atliekame jų analizę [19, 17]. Socialiniuose moksluose svarbios sekos, kylančios iš procesų, lingvistikoje intensyviai analizuojamos simbolių sekos sudarant žodžius, ekonomikoje analizuojamos sekos, reikalingoms darbo vietoms atrasti tuometinėje darbo rinkoje, versle ir rinkodaroje analizuojamos sekos iš veiklos procesų įvykių žurnalo, taip ieškant kaip galima optimizuoti procesus. Taip pat versle taikoma sekų analizė, siekiant pažinti savo klientus ir pritaikyti tinkamus verslo sprendimus. Svarbu išskirti, kad nuo tyrimo tikslo priklauso ir sekos tipas – ar tai būsenų seka ar tai įvykių seka. Būsenų seka identifikuoja tam tikra tvarka įvykstantį būsenų pasikeitimą, bet visiškai neidentifikuoja laiko, praleisto kiekvienoje būsenoje. Būtent įvykių sekos identifikuoja ne tik įvykio būseną, bet ir įvykio laiką [20]. Taip iš šių sekų galime žinoti ne tik būsenų pasikeitimo tvarką, bet ir laiką, praleistą kiekvienoje iš jų.

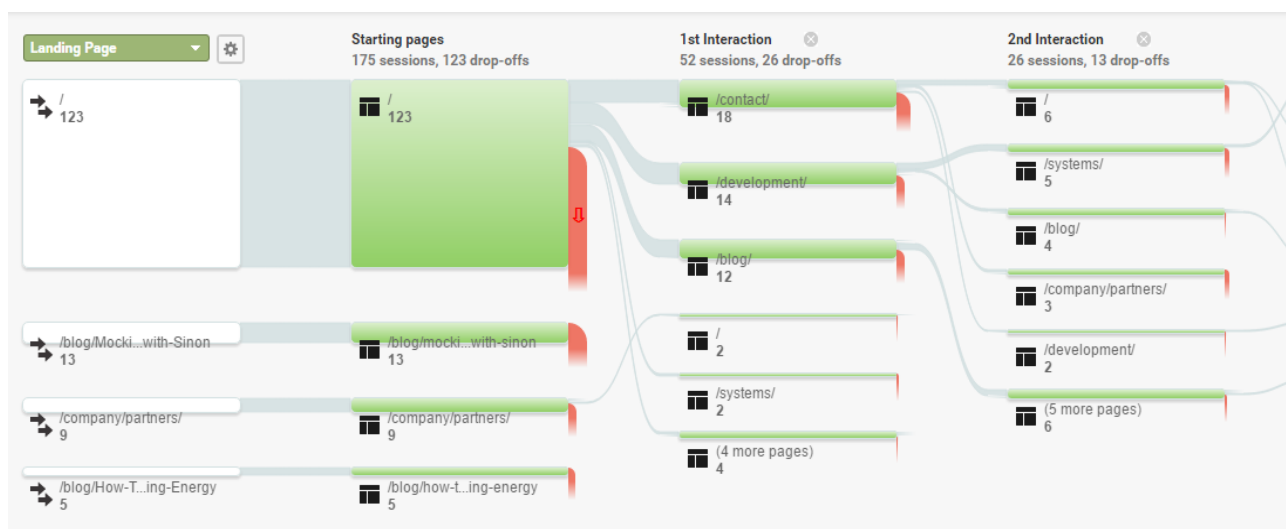
Vienas iš žiniatinklio gavybos metodų – WUM apima duomenų gavybą iš įvykio žurnalo duomenų. Šiuos duomenis transformavus į naršymo seką, galima taikyti sekų analizę. Galima išskirti, jog taikant žiniatinklio duomenų gavybą atliekamos keturios fazės [9] – duomenų gavyba, duomenų pasiruošimas, duomenų analizė, rezultatų išvedimas ir vizualizavimas. Detaliai praėjus visas fazes ir gaunamas siekiamas rezultatas. Pirmosios fazės metu yra išgaunami duomenys iš įvykių žurnalo. Toliau atliekamas duomenų pasiruošimas, kurio metu pašalinami netinkami ar nepilni duomenys. Taip pat, mažinama aplankytų skirtingų puslapių gausa šiems priskyrus tipus. Sutvarkius originalius duomenis, atliekamas duomenų transformavimas ir taip duomenys yra paruošiami ir įrašomi į struktūrizuotą formatą, tinkamą algoritmams pritaikyti. Antrosios fazės metu iš sutvarkyto duomenų failo jau galima, pritaikius algoritmą, išgauti skirtingų analizės variantų rezultatus. Vienas dažniausiai naudojamų sekų analizės variantų - tai dažnų sekų ir taisyklių radimas. Jų tikslas – atrasti dažniausiai tam tikra tvarka pasikartojančių elementų aibes ar aibių rinkinius iš visų galimų aibių duomenų rinkinių. Tiksliau tai susideda iš sub-sekų atradimo. Kaip anksčiau minėta, seką su jos elementais įvardykime kaip $S = [E_1, E_2, \dots, E_i]$. Galime įvardinti, kad S , tai sesija, E tai aplankyti puslapiai sesijos metu. Naršymo paspaudimo srauto (angl. *clickstream data*) [21, 22] seka $X = \langle x_1, x_2, \dots, x_m \rangle$ yra seka įvykių, kur $x_i \in S$. Galime išgauti informaciją, ne tik kokia tvarka įvyksta įvykiai, bet papildomai turint laiko žymę, galime ir rasti, kada tie įvykiai įvyksta ar kiek laiko tęsiasi. Įvykis sekoje gali įvykti daugiau nei kartą ir skirtingose pozicijose, todėl papildomai galima ieškoti susietumo taisyklių, atskleidžiančių tolimesnius galimus sekos elementus. Deja, neretai dažnų sekų rinkinių radimo neužtenka ir norint išgauti detalesnę informaciją, atsižvelgiama į praleistą laiką sekoje ar susiejamos sekos su papildomais kintamaisiais [23]. Galimai taikomi papildomi metodai – klasterizavimas, siekiant suskirstyti naudotojus pagal naršymo ypatybes [24], ar statistinių parametrų skaičiavimas individualioms sekoms, nubyrėjimų (angl. *funnel analysis*) tyrimai [9]. Individualių statistikų atveju dažniausiai yra skaičiuojamas vidutinis praleistas laikas tam tikrame įvykyje. Tačiau išskiriama, jog papildomai galima išgauti informacija apie sekos turbulenciją ir entropiją. Apskaičiavus šias statistikas, turbulencija parodo skaičių skirtingų sub-sekų praleistą skirtinguose įvykiuose, o entropija parodo įvykių įvairovę [25]. Paskutinioji fazė skirta gautiems analizės rezultatams susumuoti ir vizualizuoti. Šios fazės metu reikšmingiausi rezultatai pateikiami vizualiai. Taip pat, galima atlikti analizę pasitelkiant tik vizualizacijas. Netaikius jokio sekų analizės varianto, o sutvarkius duomenis ir šiuos transformavus į atitinkamą formatą, galima bandyti perteikti naudotojų naršymo kelius.

1.7. Sekų vizualizavimas

Vizualiai pateikti sekų analizės rezultatus nėra taip paprasta, kaip kad kitų analizių gautus rezultatus. Siekiant atvaizduoti realias sekas ar rastus dažnų sekų rinkinius, nubyrėjimo atvejus, reikalinga parinkti tokį vizualizavimą, kuris atsižvelgtų į visus, o ne pavienius sekos elementus. Tiriamą duomenų rinkinį gali sudaryti daugybė aibių sekų, kurių kiekvienoje nemažai skirtingų būsenų. Tiriant informacinių sistemų naudotojų naršymo duomenis egzistuoja analizės duomenų vizualizavimo problema. Išskiriamas pakankamai didelis būsenų kiekis, kuris apsunkina analizės rezultatų perteikimą. Taip pat, kiekvieną dieną sukuriamas didžiulis kiekis sesijų, kuriose aplankoma daug skirtingų puslapių, kurie sekose gali net ir pasikartoti. Toliau atskirai apžvelgiamos sankio, piltuvėlio, akordo, tinklo ir praleisto laiko apimties diagramos, jų pavyzdžiai ir taikymo galimybės.

1.7.1. Sankio diagrama

Kai kurie saityno analitikos įrankiai, pavyzdžiui „Google Analytics“ ar „WebQuilt“, gali pasiūlyti Sankio tipo diagramas, tačiau dėl didelio duomenų kiekio atvaizduoja tik daugiausiai srauto sulaukiančias sekas ir atvaizduoja tik keletą žingsnių iš šių sekų. Sankio diagrama vizualiai parodo aplankyto puslapio intensyvumą bei sekančius aplankomus puslapius. Tai įvairių išsišakojančių, pagal sekantį sekos elementą, bei susijungiančių linijų visuma. 1.2 pav. pavaizduotas Sankio diagramos pavyzdys iš „Google Analytics“ įrankio, kuriame atvaizduojamos populiariausios naudotojų sekos nuo patekimo į puslapį pradžios [26].

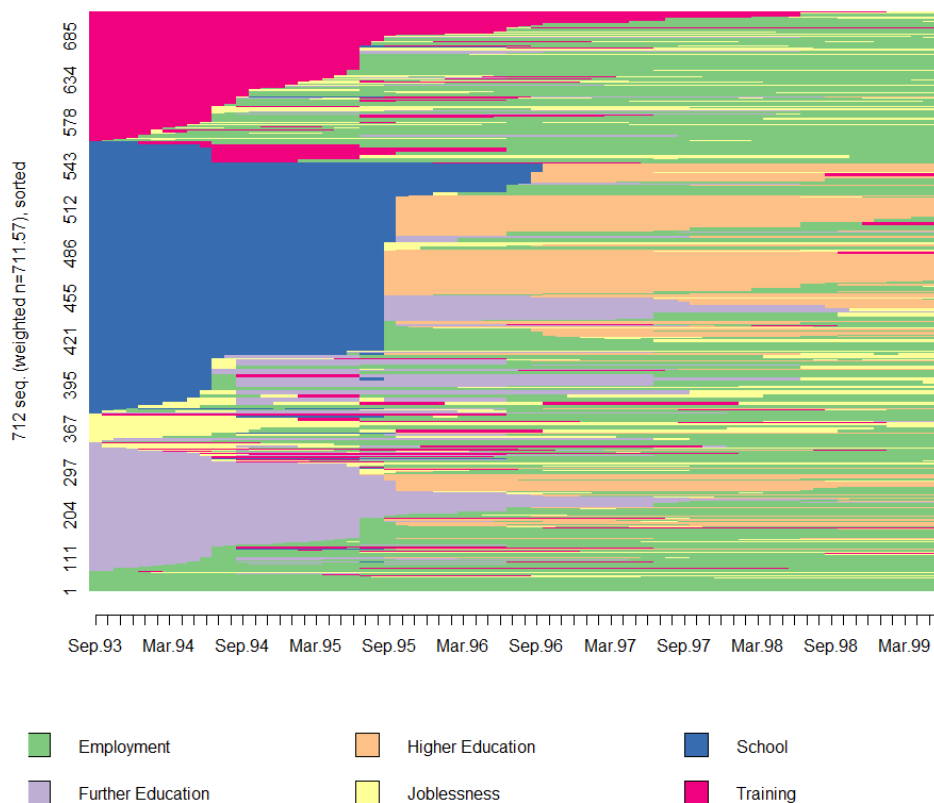


1.2 pav. Sankio diagramos pavyzdys „Google Analytics“ įrankyje

Google Analytics ar kituose įrankiuose naudojamos Sankio diagramos dažniausiai apžvelgia nedidelį skaičių sekos žingsnių, todėl nelieka galimybės pilnai įvertinti sekos pabaigos puslapių ar išgauti informaciją apie sekos ilgį. Taip pat, dažniausiai naudojamas puslapių skirstymas pagal hipersaito struktūrą tam tikrais atvejais nesuteikia galimybės lengvai išskirstyti visos sistemos hipersaitus į reikšmingus puslapių tipus.

1.7.2. Praleisto laiko dažnio ir apimties diagrama

Analizuojat sekas atsižvelgus į laiką praleistą sekoje, galima bandyti vaizduoti visas sekas viename grafike skirtingomis spalvomis išskyrus skirtingus aplankytų puslapių atvejus. Atvaizduodami visas sekas viename grafike, galime įvertinti kokioje būsenoje įvykis tęsiasi ilgiausiai ar kokių sekų gauname daugiausiai lyginant su kitomis. Tokį grafiką leidžia R paketo papildinys – TraMineR. Ši vizualizacija reikalauja suformuoti duomenų rinkinį taip, jog skirtingu laiko momentu, kiekvienas sekos elementas turėtų reikšmę. Taip y-ašyje nurodomos visos turimos sekos pagal jų id, o x-ašyje nurodoma laiko skalė. Toliau 1.3 pav. iliustruojamas vienas tokių pavyzdžių.

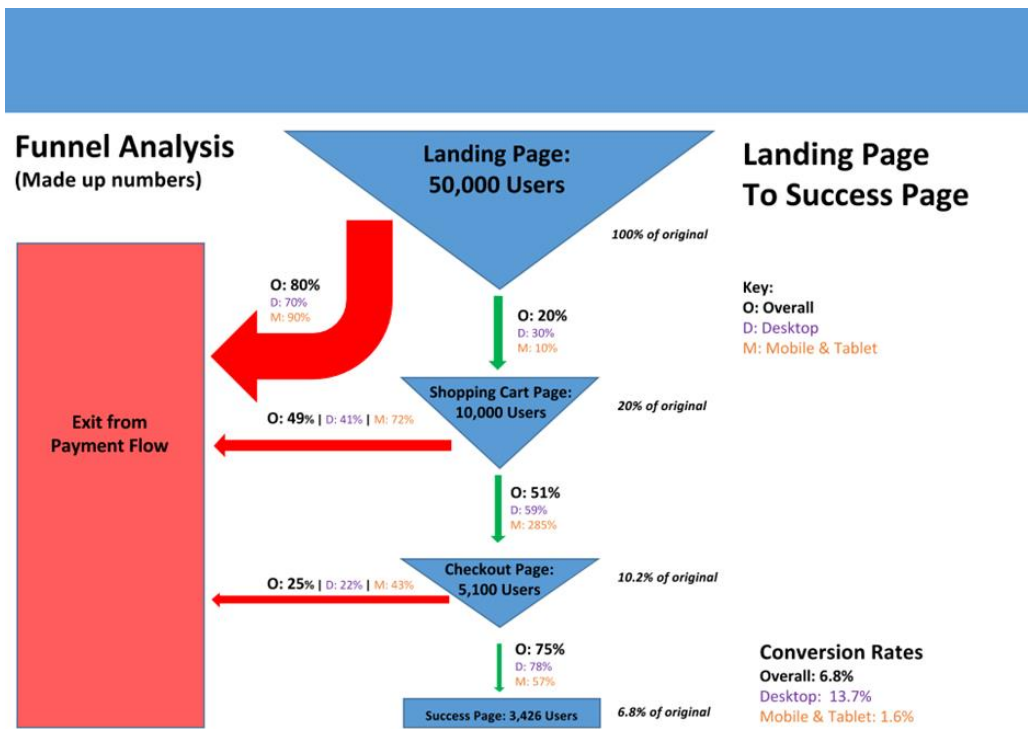


1.3 pav. Sekų dažnio grafikas sugeneruotas TraMineR paketu

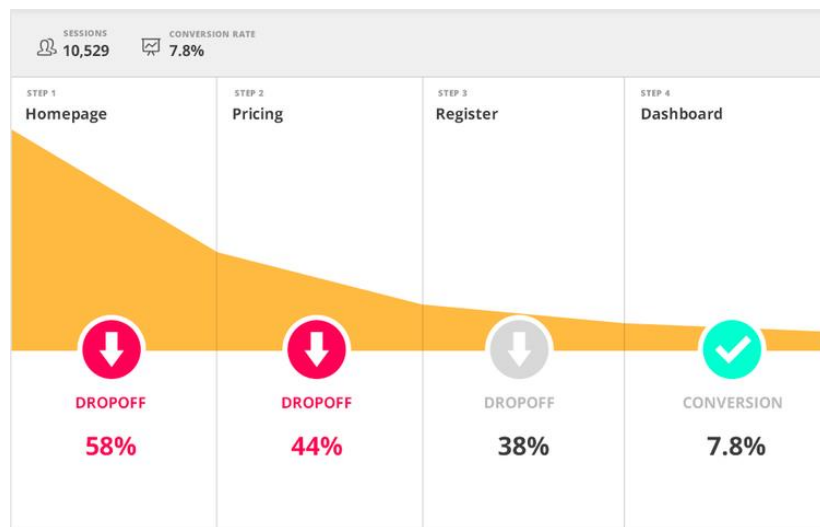
Šioje vizualizacijoje skirtingomis spalvomis parodomos galimos būsenos ir jų pasikeitimas laike. Tokiu grafiku vaizduojant naršymo sekas, gali iškilti problemų norint išgauti visoms sekoms bendrą x-ašį. Net ir turint identišką seką, gali būti praleidžiamas skirtingas laikas kiekvienoje būsenoje. Galimas sprendimas – skaičiuoti praleistą laiką procentais, taip visada išliks vienoda ašis visoms sekoms. Toks grafikas gana gerai parodo kokią laiko dalį apima tam tikras įvykis sekoje, tačiau turint labai daug įvykių ir sekų, grafikas gali tapti sunkiai suprantamas, todėl vertėtų klasterizuoti sekas ir atvaizduoti sugrupuotas sekas, kiekvienai grupei pateikiant atskirą grafiką [23].

1.7.3. Piltuvėlio diagrama

Vienas iš būdų vizualizuoti naudotojų naršymo metu atliktas sekas – tai atvaizduoti jas piltuvėlio diagrama. Šiuo grafiku galimas gana aiškus ir iš pirmo žvilgsnio lengvai suprantamas sekos vizualizavimas. Šioje diagramoje taip pat gali būti įtraukiama ir papildomų parametrų duomenys (tokie kaip prietaiso tipas ar naršyklė) ar pateikiami atvejai kai seka pasibaigia (išėjimai iš puslapio). Deja, šis būdas nepateikia bendros apžvalgos apie visas galimas sekas, o geriausiai pritaikomas vienai konkrečiai sekai peržiūrėti. Toliau 1.4 pav. ir 1.5 pav. vaizduojami piltuvėlio diagramų pavyzdžiai iš „Google Analytics“ ir „Hotjar“ komercinių analitikos įrankių paketų.



1.4 pav. Piltuvėlio diagrama rankiniu būdu generuota „Google Analytics“ platformoje

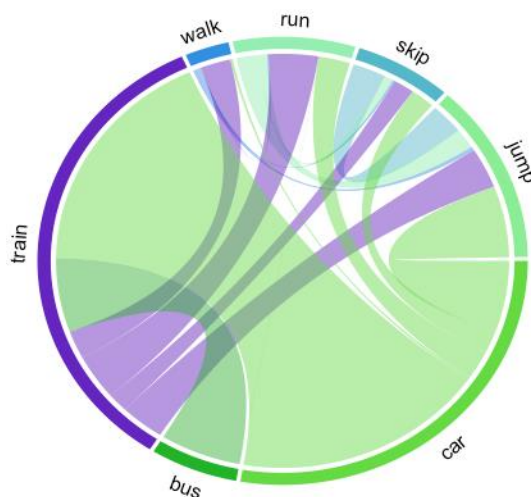


1.5 pav. Piltuvėlio diagrama sugeneruota komerciniame įrankyje „Hotjar“

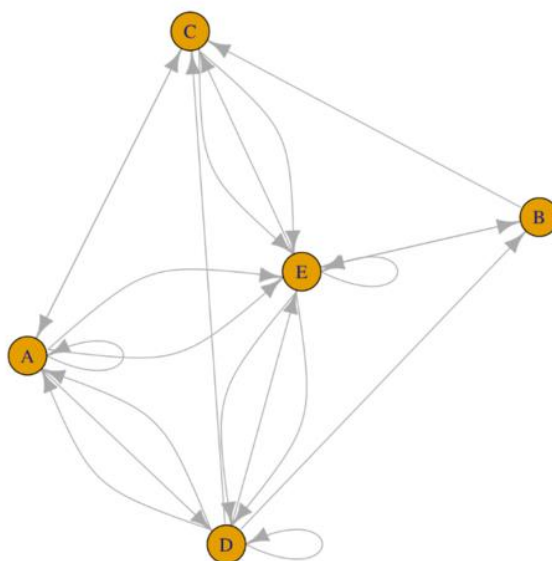
Nors piltuvėlio diagrama gana aiškiai pateikia sekos rezultatus vizualiai, gali būti sunku atvaizduoti daugiau skirtingų sekų tokia diagrama. Taip pat, ši diagrama dažniausiai nėra generuojama automatiškai, o reikalinga papildomai parinkti kiekvieną sekos įvykį. Taip vizualizavimui iš anksto reikia žinoti kiekvieną sekos elementą [27]. Daugumoje informacinių sistemų galima išskirti keletą tipinių naršymo kelių ir parenkant šių kelių žingsnius, galima matyti nubyrėjimus ir jų taškus.

1.7.4. Tinklo ir akordo diagramos

Išskiriami dar ir tinklo (*angl. Network*) ar akordo (*angl. Chord*) vaizdavimo tipai, kurie yra tinkami atvaizduoti duomenims turintiems tarpusavio ryšius pagal ryšių intensyvumą pridėdant svorius – pavyzdžiui pasikartojimų skaičių [28]. Dažniausiai šio tipo diagramos yra panaudojamos atvaizduoti dviejų elementų tarpusavio, o ne visos sekos ryšius. Toliau 1.6 pav. pateikiamas akordo diagramos pavyzdys, o 1.7 pav. tinklo diagramos pavyzdys.



1.6 pav. Akordo diagrama sugeneruota R įrankiu



1.7 pav. Tinklo diagrama sugeneruota R įrankiu

Panaudojus tinklo ar akordo tipo diagramas sekoms analizuoti, bendrą sekos kelią būtų sunku parodyti, tačiau imant sekos elementų poras, būtų galima atvaizduoti kelius į ir iš sekos elemento, jų dažnumą.

Paanalizavus galimus ir dažniausiai naudojamus vizualizavimo metodus sekų analizei, negalima nustatyti vieno geriausio metodo. Vizualizavimas priklauso tiek nuo duomenų – ar sekoje svarbus įvykių pasikeitimas laike, kokį kiekį skirtingų sekų ir jų įvykių tiriamo. Taip pat priklauso, ką norime

vizualiai atvaizduoti – individualias sekas ar visas sekas, sekų grupes. Geriausiai yra iš anksto įvertinti ir numatyti kokius analizės rezultatus norima vizualizuoti. Norinti vizualizuoti didelius sekų rinkinius, tinkama sankio diagrama, tačiau turint ir praleistą laiką galima vaizduoti praleisto laiko apimties ir dažnio diagrama. Tiriant pavienių sekų kelius, tinkama piltuvėlio diagrama, o turint nedidelį kiekį tarpusavio ryšiais susijusių elementų, juos galima vizualizuoti per tinklo ar akordo diagramas.

1.8. Sekų analizės taikymo naudotojų analizė

Sekų analizės taikymą vykdys duomenų analitikas, kuris turi atlikti duomenų paruošimą, tinkamai parinkti analizės variantus ir parametrus. Taip pat parinkti tinkamą vizualizavimo metodą analizės rezultatams pateikti. Nežinant kokius analizės variantus taikyti ar kokius vizualizavimo metodus parinkti, gauti rezultatai gali būti sunkiai interpretuojami. Analitikui pateikus sekų analizės taikymo gaires, būtų taupomas laikas analizės parinkimo identifikavimui ir vizualizacijų taikymo galimybių parinkimui.

Atlikus sekų analizę ir realizavus analitikos įrankį skirtingoms informacinėms sistemoms, galime gauti naudingos informacijos, kuri aktuali skirtingas pozicijas užimantiems asmenims. Ši informacija bus pateikiama, ataskaita, kurioje rezultatai pateikti vizualiai. Ataskaita bus lengva naudotis skirtingų profesijų asmenims. Ataskaitos rezultatai gali būti pritaikomi skirtingai, vieniems aktualu bendra statistika kokios yra pačios populiariausios sekos, kitiems, iš ko susideda seka, tretiems laikas praleistas tarp sekos elementų [3]. Žemiau įvardijami analizės rezultatų naudotojai:

Produkto vadovams (angl. *Product owner*). Jiems sekų analizės duomenys padėtų suprasti, kokio tipo naudotojai lankosi portale. Ši analizė gali padėti surasti sekas, kuomet naudotojas palieka portalą ar ieškoti ko naudotojas tikisi tame portale. Gali padėti įvertinti ar naudotojai sistemoje elgiasi taip, kaip tikimasi, pagal kuriamą funkcionalumą.

Marketingo specialistams. Šiems, specialistams, sekų analizės duomenys gali padėti rengiant marketingo strategiją, nes gali padėti atrasti kokiuose puslapiuose verčiau talpinti reklamą, parodyti vietas, kuriose sekos nutrūksta ar susideda tik iš vieno sekos elemento – ką reiškia, jog naudotojas paliko lankomą sistemą nenaršęs detaliau.

Testuotojams. Jiems sekų analizės duomenys praverstų peržiūrint kaip naudotojai naršo portale taip padedant pasiruošti testavimo atvejus, būtų galima atrinkti kritiškai svarbius kelius, kuriuos privaloma palaikyti sistemoje.

Naudotojų patyrimo specialistams (angl. *User Experience specialist*). Šie vartotojai galėtų atlikti išvalgas. Sekų analizės duomenys padėtų suprasti, kokių kelių sistemos naudotojas keliauja portale, galbūt reikia perdaryti navigaciją į paprastesnę ir lengviau suprantamą, galbūt tam tikri puslapiai yra niekada nelankomi iš vienos vietos, o visada naudotojų pasiekiami tik iš kitos vietos. Taip pat padėtų suprasti, kuriuos kelius galima optimizuoti. Atsižvelgus į laiko žymę įvertinti ar informacija pateikiama tinkamai, jei praleidžiamas toks didelis skaičius sekundžių.

Taigi, šis tyrimas padėtų duomenų analitikams paruošti ataskaitą, o gauti rezultatai suteiktų aibę skirtingų aspektų skirtingiems specialistams. Tačiau didžiausią naudą suteiks naudotojų patyrimo specialistams.

1.9. Esamų sekų analizės taikymo įrankių analizė

TraMineR – tai R kalbos paketas skirtas išgauti ir vizualizuoti sekas ar atskirai nagrinėti sekos duomenis, net jei seka turi praleisto laiko reikšmę. R ir TraMineR yra nemokami, atvirojo kodo įrankiai, todėl prieinami visiems, ir plačiai pritaikomi kartu su kitais įrankiais, kur palaikoma R integracija [23]. TraMineR yra sukurtas Ženevos universitete, Demografijos ir socialinės ekonomikos institute (IDESO) Šveicarijoje prižiūrint mokslininkų komitetui. Šis paketas skirtas biografinių išilginių socialinių mokslų duomenų analizei, tačiau galimas pritaikymas ir kitose srityse. TraMineR siūlo paprastus aprašomosios statistikos ir vizualizacijos įrankius, leidžia lengvai analizuoti sekų skirtumus, išryškina sekų dažnius. Taip pat šis paketas leidžia išgauti sub-sekas, skaičiuoti vidutinį praleistą laiką ir net turi vizualizavimo galimybes [25]. Taip pat paketas turi klasterizavimo galimybes. Taip pat vienas iš šio paketo trūkumų, tai duomenų transformavimas tam tikru specialiu formatu, siekiant išgauti, kad tiriamos sekos naudotų vienodą laiko ašį.

ArulesSequences – tai R kalbos paketas, skirtas dažnų sekų taisyklių radimui [29]. Kaip ir kiti R kalbos paketai, šis yra nemokamas ir laisvai prieinamas įrankiuose, kuriuose palaikoma R integracija. Šis paketas skirtas dažnų sekų ir taisyklių gavybai, naudojant SPADE [30] algoritimą. Naudojant šį paketą reikalingas tam tikras duomenų formatas ir kiekviena seka turi turėti savo unikalų numerį, taip pat turi būti identifikuojama elemento vieta sekoje. Nors ir būtini specialūs identifikatoriai, specialus duomenų šablonas nėra toks griežtas. Nėra būtinybės sekas skaityti horizontaliai ar vertikalčiai. Šis paketas yra skirtas būtent dažnų sekų gavybai ir taisyklių formavimui, bet neturi vizualizavimo galimybės ar detalesnių analizių galimybių. Tačiau, kaip R paketo dalis, gali būti suderinamas su kitais šios kalbos paketais.

Clickstream – tai R kalbos paketas, skirtas importuoti, analizuoti ir eksportuoti paspaudimo srauto duomenis [31]. Skirtingai nei kiti duomenų rinkiniai, naršymo sekos yra skirtingų ilgių, todėl ne visi paketai gali tinkamai nuskaityti ir apdoroti tokius duomenis. Taip pat, šis R kalbos paketas lengvai integruojamas su Markovo tinklu (*angl. Markov Chain*), kuris atskleidžia vizualizavimo ir prognozavimo galimybes. Taip pat, šiame pakete galima klasterizavimo galimybė.

Microsoft SQL Analysis Server – tai Microsoft duomenų bazių valdymo sistemos SQL Server dalis. Ši dalis turi nemažai funkcijų, galinčių atlikti analizę tiesiai iš duomenų, saugomų duomenų bazėje. Vienas iš privalumų, kad yra palaikomas sekų klasterizavimo algoritmas. Šis algoritmas galimas naudoti tik tuomet, kai turima vis įrankio – Microsoft SQL Server licencija [32]. Tai kombinuotas klasterizavimo kartu su sekų analize algoritmas. Jis atranda dažniausiai pasikartojančias sekas ir atlieka klasterizavimą pagal jų panašumą. Šis algoritmas dažniausiai naudojamas naudotojų naršymo duomenims analizuoti, analizuoti sekas iš įvykių žurnalų, apsipirkimo transakcijoms analizuoti ar prognozuoti. Kadangi šis algoritmas nėra specializuotas ir atlieka tik pagrindinę funkciją, o papildomi sprendimai ar vizualizavimai nėra galimi. Su šiuo įrankiu negalėsime tirti sekų panašumo ar skirtumų, skaičiuoti laiko praleisto kiekviename įvykyje.

EventFlow tai Maryland universiteto sukurta aplikacija, skirta tirti įvykius įvykstančius vienas po kito. Ši aplikacija labiausiai naudojama medicininiais tyrimams, o daugiausiai buvo pritaikoma atrandant gydymo būdus ir rezultatus elektroniniuose sveikatos įrašuose. Gali būti pritaikoma ir įvairesnėse srityse – transakcijų duomenims tirti, įvykių žurnalui, incidentų valdymui ar kitoms sritims, kuriose įvykių duomenys išsidėstę vienas po kito. Nuo 2018 balandžio mėnesio jau galima įsigyti šio produkto licenciją. Šis įrankis, kitaip nei kiti, geba ne tik atlikti analizę, bet ir vizualizuoti

tiek momentinius, tiek intervalo įvykių duomenis. Taip pat naudojantis šiuo įrankiu, tiesiai jame galima atlikti duomenų transformavimo procedūras, kas gali būti sunkiai suderinama kituose įrankiuose. Šio įrankio trūkumas – tai specialus duomenų šablonas. Tik teisingai užpildžius šabloną duomenimis ir importavus jį, galėsime atlikti analizę [33].

SADI tai Stata programinės įrangos paketas. Stata programinė įranga yra komercinė ir skirta statistikai ir duomenų analitikai. SADI tai tik papildinys prie esamos programinės įrangos [34]. Norint išbandyti šią programinę įrangą ir šį papildinį, reikalinga pilna paketo licencija. Panašiai kaip ir TraMineR, ši gali analizuoti sekų skirtumus ar atstumą tarp sekų, gali atrasti sub-sekas ar vaizduoti perėjimo grafikus. Skirtingai nei R papildiniuose, šiame įrankyje nėra asociacijų taisyklių radimo. Taip pat šis paketas neturi klasterizavimo galimybės, bet gali būti suderinamas su kitais Stata programinės įrangos paketais.

Įvertinus kiekvieno įrankio galimybes sudaroma ir pateikiama 1.1. lentelė su lyginamais sprendimais.

1.1 lentelė. Esamų sprendimų palyginimas

Palyginimo kriterijus	R kalbos paketai			Microsoft SQL Analysis server	EventFlow	SADI
	TraMineR	arulesSequences	Clickstream			
Galimybė vizualizuoti sekas	+	-	+	-	+	+
Galimybė analizuoti skirtingo formato sekas	+	-	+	-	+	-
Individualios išilginės sekų charakteristikos (Pavyzdžiui laikas kiekvienoj būsenoj, entropija)	+	-	-	-	+	+
Atsižvelgimas į praleistą laiko trukmę	+	-	-	-	-	+
Dažniausių sekų radimas	+	+	-	+	+	+
Asociacijų taisyklės	+	+	-	+	+	-
Klasterizavimo galimybė	+	-	+	+	-	-
Duomenų transformavimo galimybė	+	+	+	+	+	-
Specialus duomenų šablonas	+	+/-	+	-	+	-
Licencijos kaina	-	-	-	Nuo 1859\$	300\$ - 950\$	179\$/metams/ vartotojui (Stata PĮ)

Lentelėje palyginti galimi įrankiai kurie geba atlikti sekų analizę. Palygintas R kalbos paketo papildiniai – „TraMineR“, „ArulesSequences“, „Clickstream“, vienas iš duomenų bazių valdymo sistemos papildinių Microsoft SQL Server Analysis Server ir du komerciniai paketai: EventFlow ir SADI (Stata papildinys). Buvo vertinama ar įrankis turi ar neturi tam tikro funkcionalumo galimybę. Tai atvaizduojama „+“ simboliu jei įrankis tokią galimybę turi, „-“ jei įrankis tokios galimybės neturi

ir „+/-“ simboliu, jei tai veikia dalinai. Iš karto galima išskirti, jog tik R kalbos paketai nereikalauja licencijos. Vertinant savybes, galima pastebėti, kad duomenų bazių valdymo sistemos dalis SQL Analysis Server turi mažiausiai galimybių ir nėra skirtas detaliam analizei. Vienas iš svarbių vertinimo kriterijų tai ar yra vizualizavimo galimybė, ar galima analizė įtraukiant laiko skalę, ir ar galimas sekų palyginimas. Dar svarbu paanalizuoti ar įrankyje galimas asociacijų taisyklių radimas ir klasterizavimo galimybės siekiant kuo aiškiau pateikti gautus analizės rezultatus. Dar vienas punktas kurį išskirčiau, tai ar įrankyje yra galima duomenų transformavimo galimybė ir ar reikalingas specialus šablonas. Jei analizė gali būti atliekama tik į prieš tai užpildytą specifinį duomenų formatą, gali iškilti problemų analizuojant didelius duomenų kiekius. Apžvelgus įrankius, galima nustatyti, jog naudojant skirtingus R kalbos paketus, galima išpildyti daugiausiai kriterijų.

1.10. Darbo tikslas, uždaviniai ir siekiamas sprendimas

Šio tyrimo tikslas – palengvinti elgsenos analizę naudotojų patyrimo ekspertams skirtingų tipų informacinėse sistemose, apibrėžiant sekų analizės procesą ir realizuojant jam skirtą įrankį. Atlikta analizė iš įvykių žurnalo duomenų leistų nustatyti kuriuos scenarijus reiktų patobulinti pagerinant naudotojų patyrimą, padėtų nustatyti, kurie puslapiai yra labiausiai lankomi ir kokią laiko trukmę čia praleidžia naudotojai. Taip pat rezultatai leistų identifikuoti vietas kuriose naudotojai palieka sistemas ar rasti naudotojų grupes pagal naršymo šablonus ar praleistą laiką.

Norint pasiekti tikslą, iškeliami uždaviniai:

1. Išanalizuoti esamas sekų analizės tyrimo priemones, metodus, įrankius;
2. Išanalizuoti vizualizacijos būdus tinkamus analizės rezultatams pateikti;
3. Apžvelgti sekų analizės taikymo procesą ir žingsnius, kuriuos reikia atlikti;
4. Sudaryti ekspertinę apklausą, kuri padės išskirti analizės kriterijus pagal informacinės sistemos tipą;
5. Realizuoti prototipinę sistemą skirtingoms informacinėms sistemoms parenkant analizės kriterijus ir skirtingas vizualizacijas pagal informacinės sistemos tipą.
6. Pateikti rekomendacijas sekų analizės taikymui ir apibendrinti rezultatus.

Sprendimo privalumai:

- Sudaryta sekų analizės taikymo metodika identifikuoja skirtingus analizės metodus ir jų vizualizavimo galimybes. Duomenų analitikams be papildomo tyrimo pateikiami analizės atlikimo žingsniai;
- Interaktyvi ataskaita veiklos analitikos įrankyje patogiai naudotis vartotojams ir nereikalingos papildomos kompiuterinės žinios;
- Informacija duomenų modelyje automatiškai atsinaujina naujai sukauptais duomenimis;
- Sukurtas duomenų analizės modelis gali būti pritaikomas įvairaus pobūdžio informacinėms sistemoms.

Sekų analizei taikyti informacinių sistemų naudotojų patirties analitikai reikalinga unikali metodika. Šios metodikos įgyvendinimas turi apimti žingsnius nuo duomenų gavybos iki rezultatų išvedimo. Specializuota metodika turėtų išskirti gaires, nurodančias galimus analizės variantus ir jų vizualizavimo atitikmenis. Net ir sudarius metodiką, reikalinga išsiaiškinti kokius analizės kriterijus, kokiai informacinei sistemai naudoti. Tam bus atliekama naudotojų patyrimo ekspertų apklausa.

Metodikai patikrinti bus pritaikyti skirtingi sekų analizės duomenų analizės variantai naudojant R kalbą ir jos paketus „arulessequences“ ir „TraMineR“, „Clickstream“. Taikymas bus atliekamas

naudojant realių, skirtingų tipų informacinių sistemų, įvykių žurnalo duomenis. Planuojama pritaikyti ir panaudoti egzistuojantį veiklos analitikos įrankį „Power BI“ sekų analizės rezultatų vizualizacijai, sukuriant interaktyvias ataskaitas, kuriose duomenys atsinaujintų automatiškai.

1.11. Analizės išvados

Susumavus analizės metu apžvelgtus įrankius, metodus, priemones ar aspektus, gaunamos tokios išvados:

1. Naudotojų patirtis gali lemti naudotojų lojalumą informacinėse sistemose, o naršymo metu sukaupti duomenys gali būti panaudojami patirčiai nustatyti.
2. Išskiriama atskiras duomenų gavybos ir analizės būdas – žiniatinklio gavyba, kuris akcentuoja įvykių žurnalo duomenų tyrybą. Sukauptų duomenų analizė eliminuoja papildomų saityno analitikos įrankių diegimą ir taikymą.
3. Išskiriamas gausus žiniatinklio gavybos duomenų taikymas ir skirtinga informacinių sistemų aibė parodo, jog reikalingas tyrimas išskiriantis analizės variantų prioritetus pagal informacinės sistemos tipą.
4. Atskleidžiamos sekų analizės galimybės, ir išskiriami analizės variantai: dažnų sekų ir taisyklių radimas, praleisto laiko tyrimas ir individualių statistikų skaičiavimas, nubyrėjimų analizė ir klasterizavimas.
5. Apžvelgtos vizualizavimo galimybės ir nustatytos sekoms vizualizuoti diagramos – sankio, akordo, piltuvėlio, tinklo ir praleisto laiko apimties diagramos.
6. Identifikuoti vartotojai, kuriems aktualus siekiamas sprendimas. Sprendimo atlikimo procesas svarbus duomenų analitikams, o gautų analizės rezultatų vartotojai tai – naudotojų patyrimo specialistai.
7. Nustatyti R kalbos paketai: „clickstream“, „arulessequences“ ir „TraMineR“, kuriais bus išpildomi sekų analizės kriterijai ir numatytas veiklos analitikos įrankis „Power BI“ kuriame bus pateikiama rezultatų vizualizacija ir sukurta interaktyvi ataskaita, kuri automatiškai atsinaujina.

2. Sekų analizės sprendimo reikalavimų specifikacija ir projektas, formalus aprašas

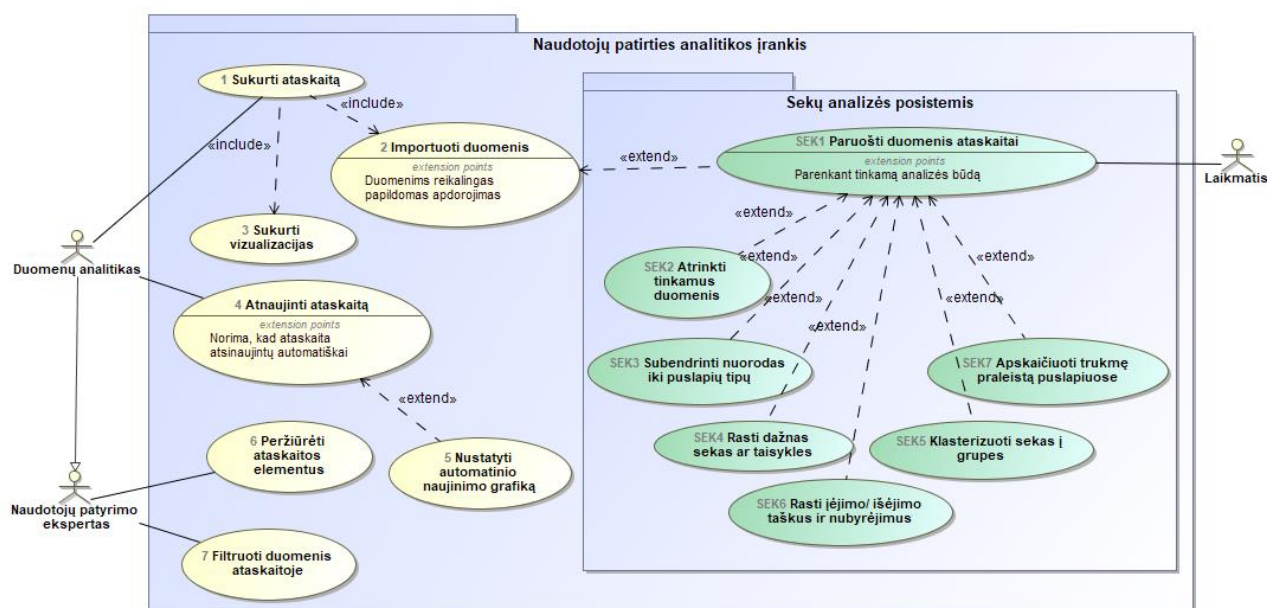
Skyriuje aprašoma sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai reikalavimų modelis.

2.1. Sekų analizės posistemio reikalavimų specifikacija

Išskiriami pagrindiniai funkciniai reikalavimai. Reikalavimai išskiriami tiek sekų analizės posistemiiui, tiek analitikos įrankiui kuriame galutinis vartotojas matys rezultatus. Toliau dokumente bus detalizuojami tik sekų analizės posistemiiui keliami reikalavimai, nes veiklos analitikos įrankio veikimas nėra autoriaus indėlis. Siekiant aprašyti procesą ir nurodyti žingsnius, kuriuos reikia atlikti nuo duomenų importavimo iki rezultatų išvedimo, kuriama metodika atskleis ir tam tikra veiklos analitikos įrankio galimybes.

Pateikiamas panaudojimo atvejų modelis, kuris atspindi visus keliamus sistemos reikalavimus. Sekų analizės taikymui naudotojų patirties analitikai reikalingas funkcionalumas toliau yra detalizuojamas veiklos diagramomis.

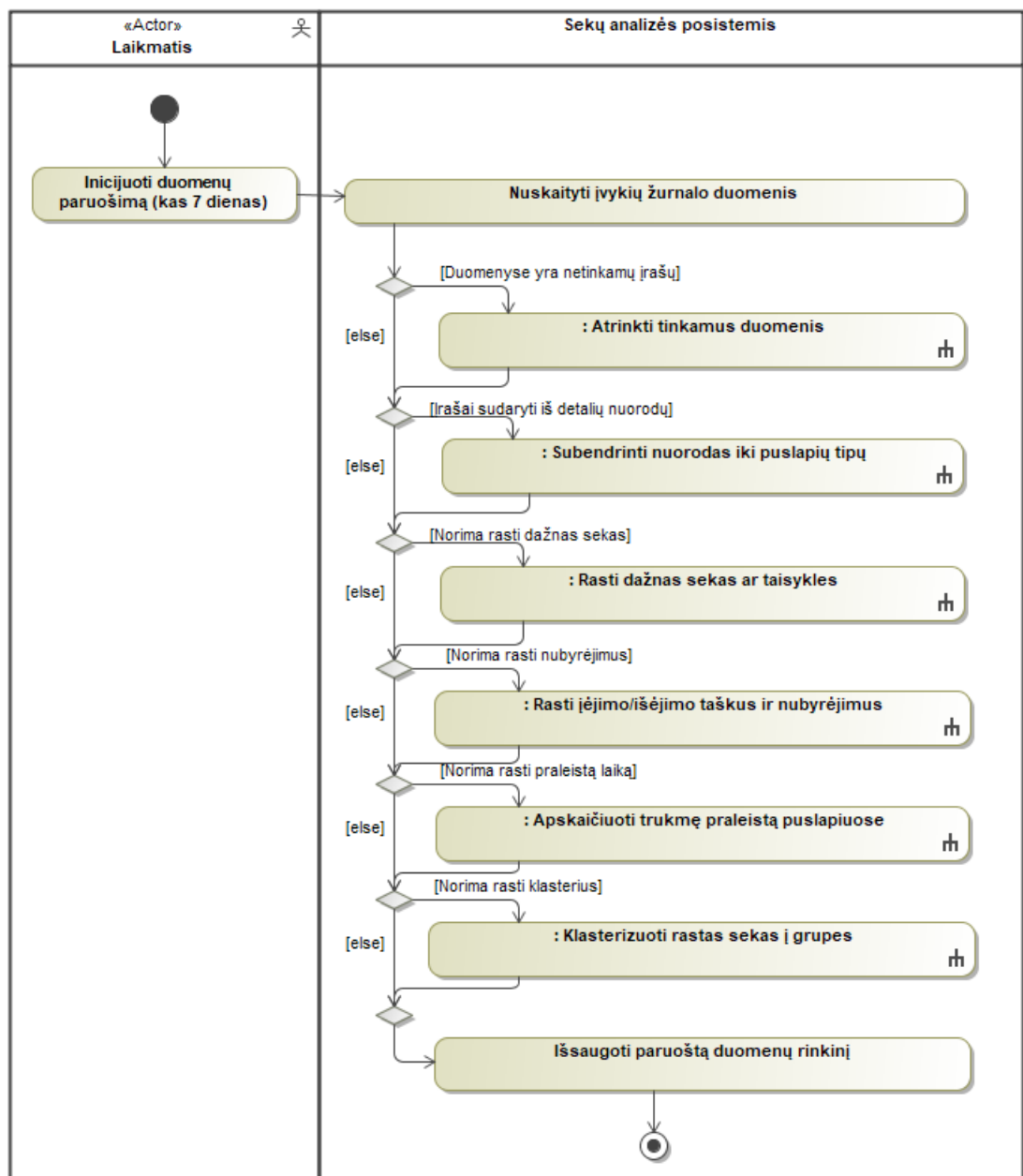
Pagrindiniai aktoriai – duomenų analitikas ir naudotojų patyrimo ekspertas, o funkcijos kurias jie gali atlikti yra pateikiami 2.1 pav. Yra išskiriamas ir vienas antrinis aktorius – Laikmatis. Naudotojų patyrimo ekspertas gali peržiūrėti skirtingus ataskaitos elementus, filtruoti duomenis ir analizuoti juos įvairiais pjūviais. Duomenų analitikas taip pat gali atlikti šias funkcijas, todėl galima įvardyti, jog šioms funkcijoms atlikti paveldi naudotojų patyrimo eksperto rolę. Duomenų analitikas dar gali sukurti ataskaitą importuojant duomenis ir sukuriant vizualizacijas joje. Taip pat, atnaujinti ataskaitą ar nustatyti ataskaitos atnaujinimą. Aktorius laikmatis inicijuoja duomenų paruošimą, pagal nustatytą grafiką. Galimi duomenų paruošimo būdai – duomenų atrinkimas, nuorodų subendrinimas, dažnų sekų ar taisyklių radimas, įėjimo/išėjimo ir nubyrijimo taškų radimas, klasterizavimas ir praleisto laiko skaičiavimas.



2.1 pav. Sekų analizės tyrimo panaudojimo atvejų modelis

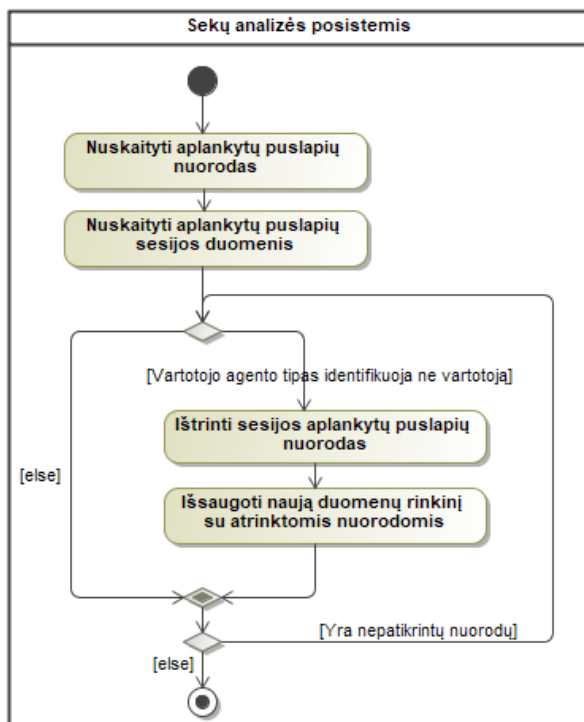
Likusioje skyriaus dalyje pateikiamos visų sekų analizės posistemio panaudojimo atvejų veiklos diagramos. Jose parodoma, kokius veiksmus atlieka sistema siekiant įgyvendinti atitinkamą analizės kriterijų ir paruošti duomenis. Taip pat, detaliai aprašomi naudojami algoritmai naudojami duomenų paruošimo procedūrose.

Veiklos procesas pateiktas 2.2 pav. parodo, kaip sekų analizės posistemis atlieka duomenų paruošimo procesą. Laikmatis inicijuos duomenų paruošimą kartą per savaitę. Galima nustatyti duomenų paruošimą ir kas dieną, tačiau naršymo sekos taip dažnai nesikeičia, todėl pakaks ir retesnio duomenų naujinimo. Pirmiausiai, bus nuskaitomi įvykių žurnalo duomenys. Įvertinus duomenų kokybę atliekami duomenų atrinkimo ir tipų suteikimo žingsniai. Pagal norimą analizuoti kriterijų atliekami norimi scenarijai. Įvykdžius reikiamus scenarijus, duomenys išsaugojami naujose lentelėse.



2.2 pav. PA „Paruošti duomenis ataskaitai“ veiklos diagrama

Veiklos diagrama pateikta 2.3 pav. parodo, kaip sistema vykdo scenarijų ir atlieka duomenų atrinkimą. Pagrindinis duomenų atrinkimo tikslas yra išrinkti tik realių naudotojų sesijų duomenis. Jeigu tiriamoje sistemoje yra sesijų, kurios yra sugeneruotos automatinio testų ar interneto botų, šias sesijas reikia eliminuoti. Tokie įrašai gali iškreipti analizės rezultatus. Kadangi įvykių žurnale galima rasti naršyklės vartotojo agento tipą (angl. *User Agent*) ir IP adresą, iš šių duomenų galima atpažinti netinkamas sesijas. Populiarių paieškos variklių botai būtent vartotojo agento įrašė nurodo savo kilmę. Kiti interneto botai ne visada nurodo kilmę, tačiau juos eliminuoti galima agento tipe radus „*HeadlessChrome*“ įrašą. Žinant automatinio testų ar kitų, ne naudotojų generuotų vartotojo agento tipą, šalinti duomenis pagal tokį patį principą. Taip šalinimas vyksta tikrinant kiekvienos nuorodos sesiją ir išsaugomas naujas duomenų rinkinys su atrinktais duomenimis.



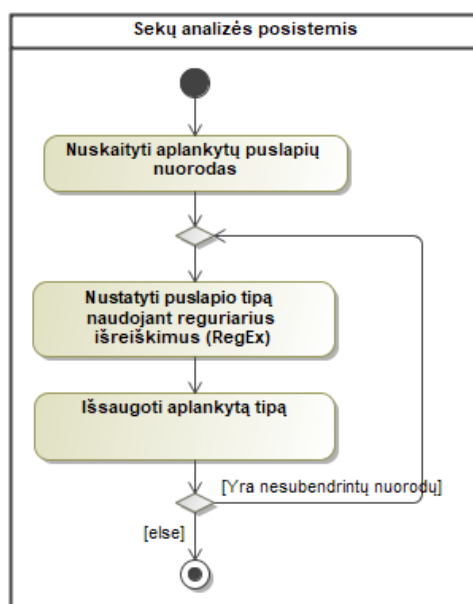
2.3 pav. PA „Atrinkti tinkamus duomenis“ veiklos diagrama

Toliau 2.4 pav. pateikiama veiklos diagrama, kaip sistema vykdo scenarijų ir atlieka puslapių bendrinimą iki puslapių tipų. Pagrindinis puslapių subendrinimo tikslas, tai dažnų sekų radimui naudoti puslapių tipus, o ne pačių aplankytų puslapių nuorodas. Dažnas sekas galima rasti ir iš originalių duomenų, tačiau gauti rezultatai gali apsunkinti interpretavimą, neatskleistų tikrųjų naršymo šablonų ar atrasti begalę panašių ir kartu skirtingų sekų. Puslapių tipų identifikavimui bus naudojamos reguliarios išraiškos – RegExp [35]. Puslapių tipai yra nustatomi individualiai pagal svetainės struktūrą ir turi būti nurodoma taip, kad būtų aiškus tipas, bet daugiau nedetalizuojama. Pavyzdžiui:

- [/eshop/product/hp-elitebook-845-g7/?prodid=1565375](#) → Produkto kortelė
- [/eshop/products/monitoriai-projektoriai/monitoriai/](#) → Produktų sąrašas

Taip, nenurodoma kokio tipo ar gamintojo produktai parenkami, tačiau tik parodomas tipas. Suteikus tipus, ieškant dažnų sekų bus išgaunami daug geresni ir aiškesni rezultatai. Suteikiami tipai yra

kiekvienai aplankyti nuorodai, ir procedūra tęsiasi tol, kol subendrinamos visos aplankytos nuorodos.



2.4 pav. PA „Subendrinti nuorodas iki puslapių tipų“ veiklos diagrama

Toliau 2.5 pav. pateikiama veiklos diagrama, kaip sistema vykdo scenarijų ir randa dažnas sekas ar taisykles. Pagrindinis dažnų sekų radimo tikslas – atrasti dažniausiai tam tikra tvarka pasikartojančių elementų aibes ar aibių rinkinius iš visų galimų aibių duomenų rinkinių. Tiksliau tai susideda iš sub-sekų atradimo. Kaip analizės dalyje minėta, seką su jos elementais įvardykime kaip $S = [E_1, E_2, \dots, E_i]$. Galime įvardinti, kad S , tai sesija, E tai aplankyti puslapiai sesijos metu. Naršymo paspaudimo srauto seka $X = \langle x_1, x_2, \dots, x_m \rangle$ yra seka įvykių, kur $x_i \in S$. Įvykis sekoje gali įvykti daugiau nei kartą ir skirtingose pozicijose. Skirtinga pozicija nurodo įvykių eiliškumą, o įvykių skaičius paspaudimo sraute žymi paspaudimo srauto ilgį. Taip paspaudimo srautas, kurio ilgis yra k , vadinamas k -paspaudimų srautu. Rinkinys X vadinamas dažnu (angl. *support*), jei jo pasikartojimas yra nemažesnis už nurodytą minimalų dažnumą min_supp . Ieškant dažnų rinkinių sekoje, reikia numatyti ar pakanka rasti, tik sub-sekas ir jų dažnius ar norima ir išgauti taisykles [36]. Dažniems rinkiniams rasti pakanka apskaičiuoti dažnį.

$$supp(X) = \frac{|\{X \subseteq T\}|}{|T|}; \quad (1)$$

Čia X dažnumas, pasitaikantis transakcijose T , apskaičiuojamas kaip proporcija transakcijų duomenyse, kuriuose yra rinkinys X . Norint atrasti taisyklę dažnam rinkiniui skaičiuojamas ir patikimumas (angl. *confidence*).

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}; \quad (2)$$

Patikimumas žymimas $conf(X \Rightarrow Y)$ yra sąlyginė tikimybė, kad transakcijoje yra rinkinys Y su sąlyga, kad rinkinyje taip pat yra rinkinys X . Nustatant taisykles apibrėžiamas minimalus dažnumas min_supp ir minimalus patikimumas min_conf [12]. Taisyklės aktualumui naudojamas svarbos matas (angl. *lift*). Šiuo matu nurodoma kaip toli nuo nepriklausomumo yra X ir Y .

$$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)}; \quad (3)$$

Net ir įvertinus ar pakanka tik dažnų rinkinių gavybos ar papildomai norima rasti taisykles, reikia atsižvelgti į turimus duomenis ir tai ar svarbus elementų išdėstymas sekoje. Jeigu yra siekiama rasti atvejus sekoje nežiūrint į elementų tvarką sekoje, pakanka naudoti Apriori algoritmą. Pavyzdžiui, norime žinoti kokius straipsnius dažniausiai skaito vienoje sesijoje, tačiau nėra svarbu skaitymo tvarka. Jeigu yra svarbi elementų išdėstymo tvarka sekoje, galima naudoti SPADE algoritmą ar TraMineR paketo algoritmą. Šiam naudoti būtina turėti praleisto laiko žymę.

Apriori algoritmas populiariausiai naudojamas pirkimų krepšelio analizėje [36]. Šis algoritmas dažniausiai yra naudojamas išgaunant sub-rinkinius neatsižvelgiant į tiriamų elementų rinkinyje išdėstymą. Taikant šį algoritmą, būtų galima nustatyti kokius puslapius sesijoje aplanko kartu, tačiau nebūtų galima identifikuoti šių įvykių vietos sekoje. Pavyzdžiui, tiriant aplankytų puslapių kombinacijas galima nustatyti dažniausiai kartu aplankomų puslapių rinkinius, tačiau norint tirti kurie puslapiai vedą į pirkimo atvejį, šio algoritmo rezultatai tokio rezultato neatskleidžia.

TraMineR pakete taip pat galima dažnų sekų ir taisyklių gavyba [37, 25]. Pirmiausiai atliekama būsenų sekų transformacija į įvykių sekas. Tuomet yra atliekama sub-sekų paieška išlaikant įvykių eiliškumą. Pavyzdžiui, (a;b) – (a) yra (a;b) – (a;b;c) – (a) sub-seka, nes yra išlaikoma įvykių tvarka. Sub-seka laikoma dažna, jei atitinka minimalų dažnį, kuriam turi priklausyti. Siekiant išgauti susietumo taisykles, nurodomas minimalus patikimumo matas ir gauname taisyklę. Pavyzdžiui, gavus (a;b;c) => (a;b) patikimumą 0,4057 reikštų, kad seka, kurioje yra (a;b;c) 40,57% kartų sekoje bus ir (a;b).

SPADE algoritmas naudoja vertikalų dažnų sekų saugojimą, todėl sumažėja pradinės duomenų bazės skenavimų skaičius. Algoritmas pirmojo skenavimo metu nustato vieno elemento posekius (pvz. „A“ elementas rodomas dažniau nei „D“). Tuomet stebimi dviejų elementų ilgių sekos įvertinant laiko seką (pvz. B -> A reikalauja, kad B būtų įvykęs anksčiau už A). Tuomet remiantis dažniausiai pasitaikančiais dviejų ilgių išėjimais pereiname prie trijų elementų sekų. Tai tęsiasi tol. Kol pasiekiamo vartotojo nustatytą maksimalų ilgį arba tol, kol pasiekiamo ilgį, per kurį nebegalime rasti dažnų išėjimų [30].

Toliau, 2.1 lentelėje pateikimas paprastas 10 sekų pavyzdys. Duomenų rinkinį sudaro trys stulpeliai. Vienne – sekos ID, kitame – įvykio ID, o trečiame – patys elementai.

2.1 lentelė. Sekų duomenų rinkinys

sequenceID	eventID	items
1	10	C, D
1	15	A, B, C
1	20	A, B, F
1	25	A, C, D, F
2	15	A, B, F
2	20	E
3	10	A, B, F
4	10	D, G, H
4	20	B, F

4	25	A, G, H
---	----	---------

Naudojant SPADE algoritmą, išgaunami dažni sekų rinkiniai. Lentelėje 2.2 pateikiami dažnų sekų rinkinių dalis, kai dažnumas $min_supp=0,5$.

2.2 lentelė. dažnų įvykių sekų rinkiniai, kai dažnumas $min_supp=0,5$

sequence	Support
<{A}>	1,00
<{D}>	0,50
<{D}, {F}>	0,50
<{D}, {B}>	0,50
<{B}, {A}>	0,50
<{F}, {A}>	0,50
<{D}, {B}, {A}>	0,50

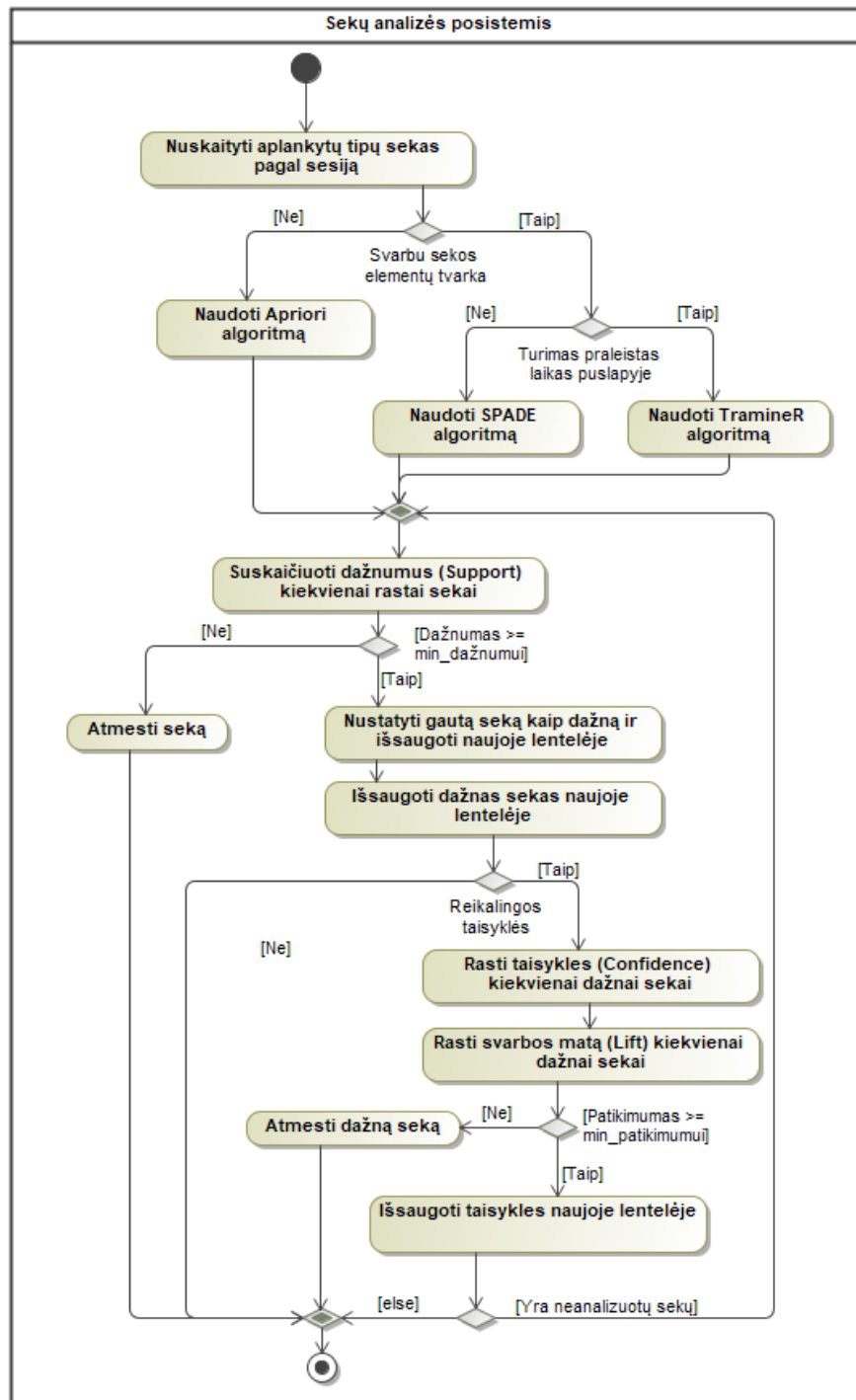
Išgavus dažnas sekas galima palyginti kaip dažnai gali įvykti viena seka lyginant su kita, rasti sekų panašumus ir skirtumus [38]. SPADE algoritmas gali ne tik rasti dažnus sekų rinkinius, bet ir rasti taisykles, prognozuojančias koks gali būti elementas sekoje, jei turime dalį sekos aibės. Lentelėje 2.3 pateikiami taisyklių rinkinių dalis, kai patikimumas $min_conf=0,5$

2.3 lentelė. Dažnų sekų taisyklių rinkiniai, kai patikimumas $min_conf=0,5$

rule	support	confidence
<{D}> => <{F}>	0,5	1,0
<{B}> => <{A}>	0,5	0,5
<{D}> => <{B}>	0,5	1,0
<{D}, {F}> => <{A}>	0,5	1,0
<{D}, {B}> => <{A}>	0,5	1,0

Taisyklių rezultato interpretavimas gali būti išreiškiamas taip peržvelgiant pirmąją eilutę: jei sekoje įvyko įvykis D, tai tikrai vėliau įvyks ir įvykis F. Peržiūrint antrąją eilutę, kur patikimumo matas nėra lygus 1, negalime griežtai teigti, kad įvykus įvykiui sekoje, įvyks įvykis A. Įvykis A gali įvykti po įvykio B, tačiau su mažesne tikimybe [30].

Parinkus tinkamą algoritmą, dažnų sekų rinkinių ar taisyklių rezultatai išsaugomi atskirose lentelėse. Išgauti rezultatai gali būti analizuojami tiesiog peržiūrėjus taisyklių lentelę ar bandomas vizualizacijos taikymas analizės rezultatams.

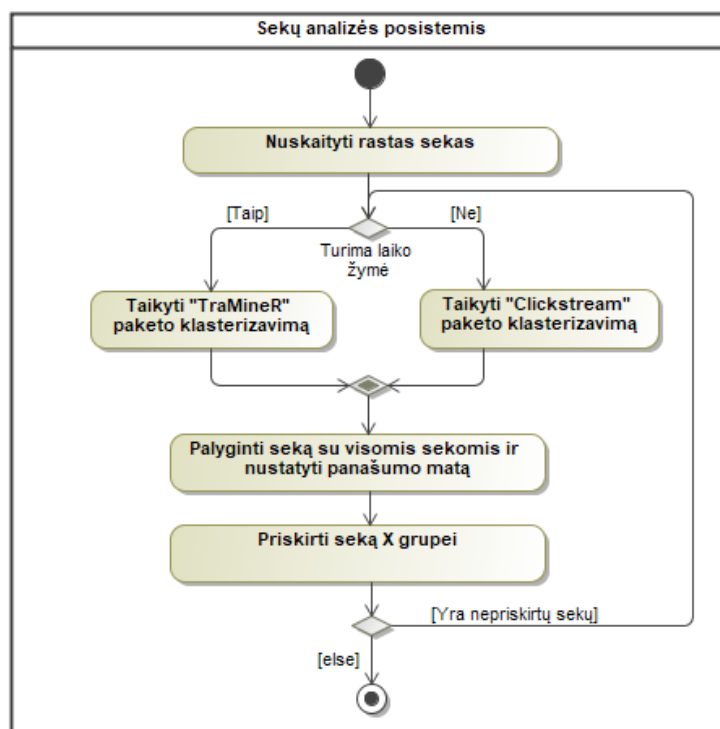


2.5 pav. PA „Rasti dažnas sekas ar taisykles“ veiklos diagrama

Norint išskirti naudotojus į grupes pagal naršymo kelius arba žinant, jog tiriamas portalas susidaro iš daugybės skirtingų tipų puslapių ir gali būti išgaunamas didelis sekų kiekis, siekiama lengviau interpretuoti sekas ir jas suskirstyti į grupes. Tokiu būdu būtų galima nustatyti klientų segmentus, rasti lankytojus, turinčius tokius pačius tikslus besilankant portaluose. Tai padaryti galima pritaikius klasterizavimo metodus. Klasterizavimo tikslas yra išskirti homogeniškas grupes, kur grupėje priklausantys elementai yra panašūs vieni į kitus, bet skiriasi nuo elementų, kurie yra kitose grupėse.

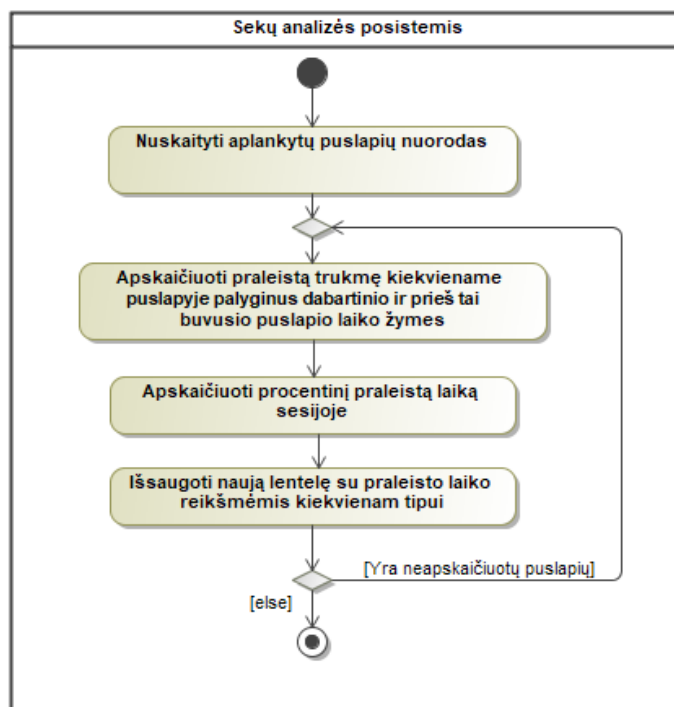
Yra hierarchiniai ir nehierarchiniai klasterizavimo metodai [24]. Hierarchinis metodas naudojamas norint suburti panašius objektus į mažesnes grupes, o šias sujungti į didesnes objektų grupes.

Nehierarchiniai metodai taikomi, kai yra iš anksto žinomas klasterių skaičius. Turint aplankyto puslapio trukmę, galimas hierarchinis klasterizavimo metodas naudojant „TraMineR“ paketą [25]. Išskiriant klasterius, lyginamas praleistas laikas sekoje ir sekos elementų pasikeitimai. Turint įrašus be praleisto laiko trukmės, o tik paspaudimo srauto seką, galimas k-vidurkių klasterizavimas naudojant „Clickstream“ paketą [24]. Tiriant aplankytų puslapių sekas, iš anksto numatyti klasterių skaičiaus negalime, todėl reikės duomenų analitiko įsikišimo numatyti tinkamą klasterių skaičių. Toliau 2.6 pav. pateikiama diagrama, kaip sistema vykdo scenarijų ir randa klasterius išgautoms sekoms.



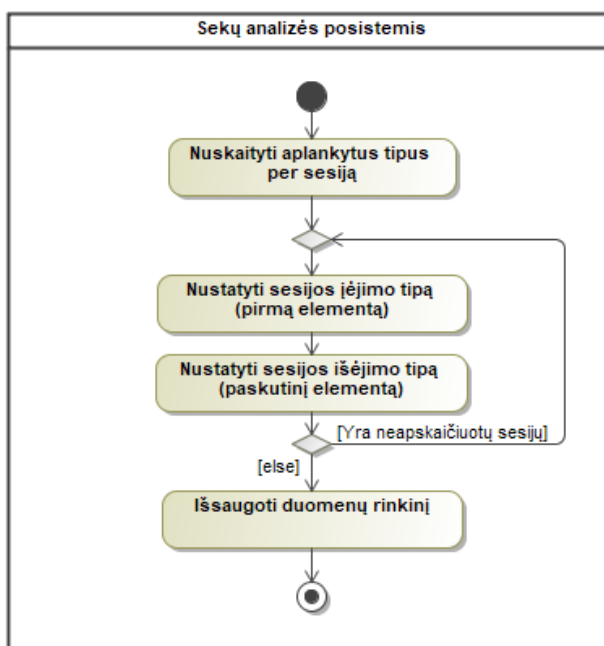
2.6 pav. PA „Klasterizuoti rastas sekas į grupes“ veiklos diagrama

Tiriant sekas galima atsižvelgti ne tik į tai, ką naudotojas aplanko, bet ir tai kiek laiko praleidžia, papildomai apskaičiuoti su praleistu laiku susijusius matavimus ar pagal praleistą laiką skirstyti sekas į grupes. Yra galimi labai skirtingi laiko praleidimo intervalai, rezultatų vizualizavimui bus paverčiamas laikas į procentus. Toliau 2.7 pav. pateikiama veiklos diagrama, kaip sistema vykdo scenarijų ir apskaičiuoja praleistą laiką puslapyje ir apskaičiuoja procentinę praleisto laiko dalį sesijoje pagal šį tipą.



2.7 pav. PA „Apskaičiuoti trukmę praleistą puslapiuose“ veiklos diagrama

Tiriant ar apžvelgiant sekų visumą ar randant tik dažniausių sekų rinkinius, neatskleidžiama, kaip naudotojai pradeda ar baigia sekas, kokie lūžio taškai galimi vizualizacijose. Šiems aspektams tirti, galima pasitelkti skirtingas vizualizacijas, tačiau reikia tinkamai paruošti duomenų rinkinį. Toliau 2.8 pav. vaizduojama, kaip yra paruošiamas duomenų rinkinys norint vėliau rasti nubyrėjimus ir kaip randami įėjimo ir išėjimo taškai.

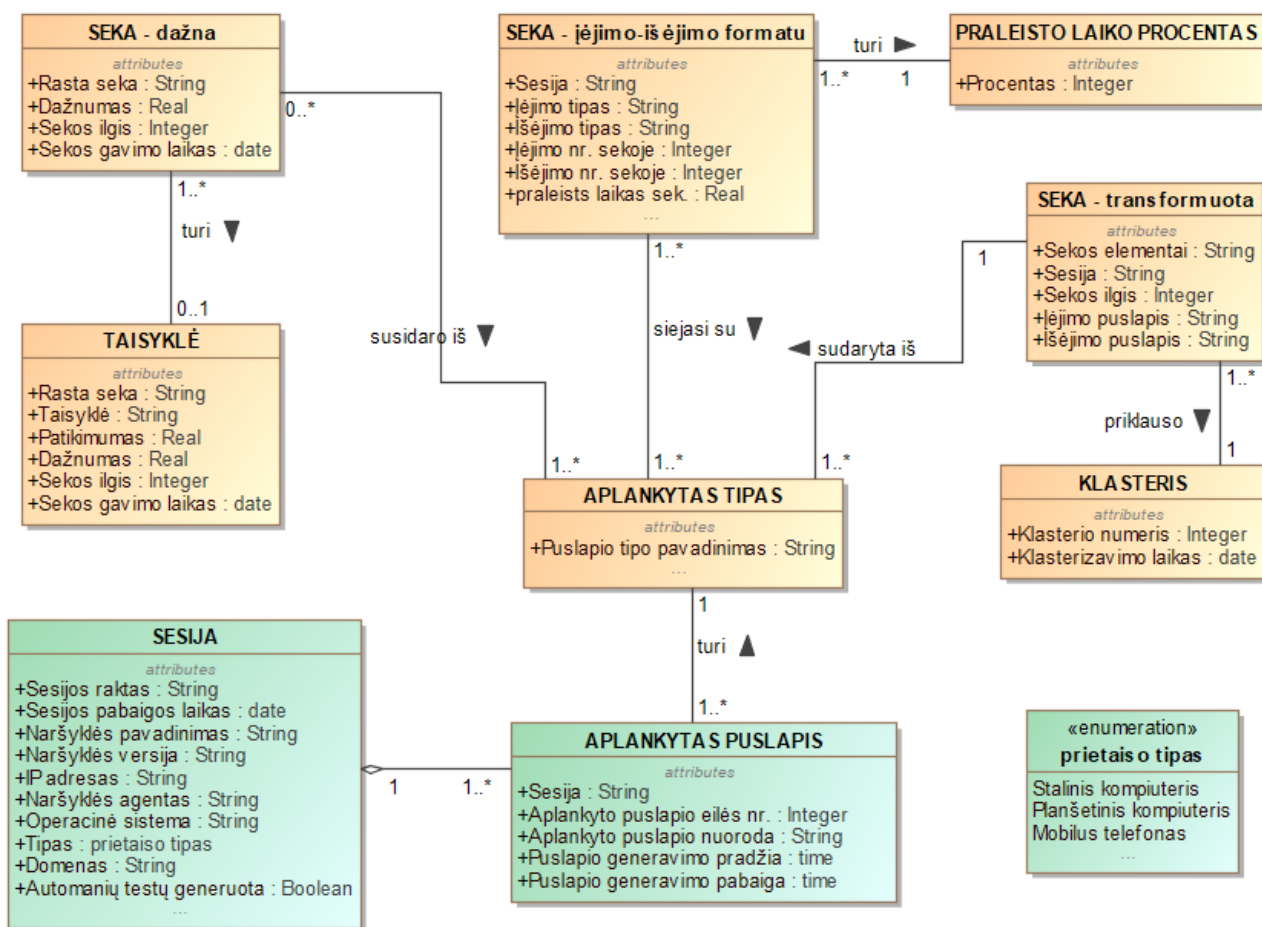


2.8 pav. PA „Rasti įėjimo/ išėjimo taškus ir nubyrėjimus“ veiklos diagrama

Išskyrus pagrindinius reikalavimus, numatyta kokius variantus gali atlikti sekų analizės posistemis ir kaip kokie sistemos žingsniai atliekami kiekviename varianto žingsnyje. Net ir įvertinus sistemos žingsnius, neatskleidžiamos ataskaitos kūrimo gairės, bei galimų analizės rezultatų vizualizavimo atitikmenys. Nuspręsta sudaryti procesą, leidžiantį nustatyti detalius žingsnius sekų analizės posistemiu išpildyti ir ataskaitai sukurti.

2.2. Sekų analizės posistemio dalykinės srities modelis

Analizuojamos sekų analizės esybės pateikiamos 2.9 pav. Žalia spalva išskiriamos esybės - aplankytas puslapis ir sesija. Jos atspindi realiai turimų ir saugomų įvykių žurnalo duomenų modelį. Kita spalva pavaizduotos esybės papildytų esamą duomenų modelį pritaikius sekų analizę.



2.9 pav. Įvykių žurnalo duomenų modelis (žalia spalva) ir papildymas esybėmis (geltona spalva)

Turimi sukaupti duomenys apie sesiją ir aplankytus puslapius sesijos metu. Sesija ir susidaro iš aplankytų puslapių, kurių turi būti bent vienas, o konkretus aplankomas puslapis priklauso vienai konkrečiai sesijai. Kiekvienas aplankytas puslapis turi turėti tipą. Sekos yra sudarytos iš aplankytų puslapių tipų. Bus reikalingos skirtingų tipų sekos – tai realios ir gautos pritaikius algoritmą, kurios dar gali turėti ir taisykles. Realios sekos naudojamos nubyrėjimams, praleisto laiko analizei ar detaliam vizualizavimui. Papildomai realios sekos yra transformuojamos. Seka-transformuota yra vienos sesijos įrašai sudaryti iš tipų, ir papildomai saugoma informacija apie sekos ilgį, įėjimo iš išėjimo puslapius. Išgaunant klasterius, galimas realių sekų skirstymas pagal sesijas, taip kiekviena seka-transformuota įgautų klasterio reikšmę. Kitas reikalingas formatas – transformuota seka į seką

įėjimo-išėjimo formatu. Ši esybė turi informaciją apie praleistą laiką puslapyje. Papildomai reikalinga sekai įėjimo-išėjimo formatu apskaičiuoti praleisto laiko procentą.

2.3. Formalus sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai sprendimo aprašas

Siekiant atlikti duomenų analizę, beveik visada galima rasti informaciją apie pagrindinių keturių žingsnių įvykdymą. Juos galima rasti ir CRISP-DM [39] ir SEM [40] metodologijose. Išskiriami pagrindiniai žingsniai:

1. Duomenų gavyba;
2. Duomenų paruošimas;
3. Duomenų analizė;
4. Rezultatų išvedimas (vizualizacijos).

Detalūs žingsniai dažnai detalizuojami tik tam tikro atvejo analizėse, ir sudėtinga rasti tikslius žingsnius, kuriuos reikia atlikti ar kokią analizę taikyti, kaip rezultatus išvesti. Taikant sekų analizę galima iš anksto išskirti analizės kriterijus, kuriuos galima atlikti naudojant iš eilės einančių duomenų įrašus. Galimi analizės variantai:

- Įėjimo/išėjimų ir nubyrežimų analizė;
- Duomenų skirstymas į klasterius;
- Dažnų sekų ir taisyklių radimas;
- Dažnų elementų ir taisyklių radimas;
- Praleisto laiko analizė.

Apžvelgus vizualizavimo metodus išskiriama skirtingi būdai, kurie leidžia vaizduoti rezultatus atsižvelgdami į seką. Vizualizavimo būdai:

- Piltuvėlio diagrama;
- Sankio diagrama;
- Akordo diagrama;
- Tinklo diagrama;
- Taisyklių lentelė;
- Praleisto laiko dažnio ir apimties diagrama.

Žinant analizės būdus ir vizualizavimo galimybes, galima papildyti duomenų analizės ir rezultatų išvedimo žingsnius įvardijant analizės kriterijus ir susiejant kiekvieną su galimu rezultatų išvedimo būdu. Lentelėje 2.4 vaizduojami analizės kriterijų ir galimo vizualizavimo atitikmenys.

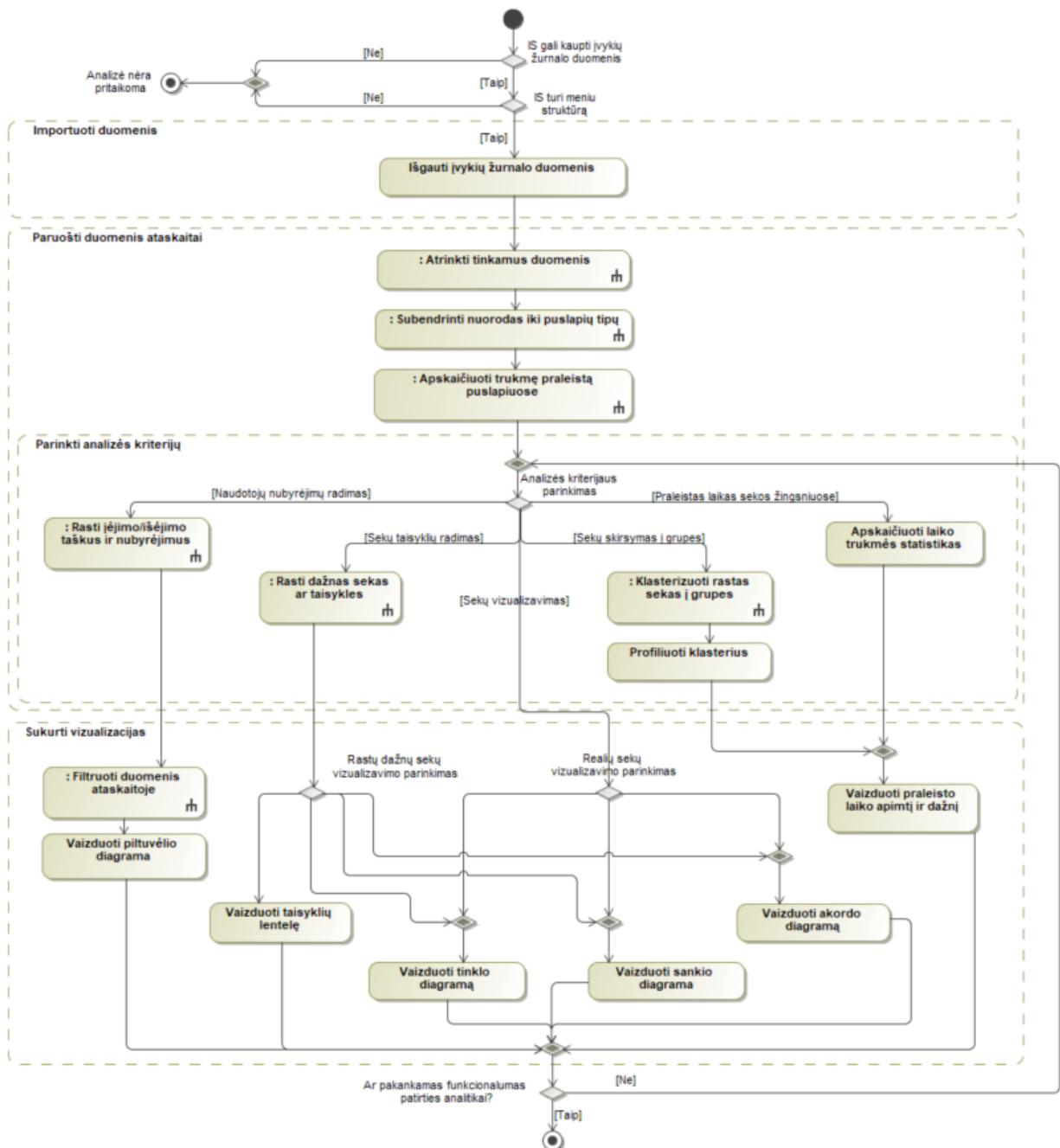
2.4 lentelė. Analizės varianto rezultato galimas vizualizavimas

Analizės variantas	Galimas vizualizavimas
Įėjimo/ išėjimų ir nubyrežimų analizė	- Piltuvėlio diagrama
Duomenų skirstymas į klasterius	- Praleisto laiko apimties ir dažnio diagramos išskirstant pagal klasterių skaičių
Praleisto laiko analizė	- Praleisto laiko ir apimties diagrama
Dažnų elementų ir taisyklių radimas	- Tinklo diagrama; - Taisyklių lentelė;
Dažnų sekų ir taisyklių radimas	- Sankio diagrama; - Taisyklių lentelė;

	<ul style="list-style-type: none"> - Akordo diagrama; - Tinklo diagrama
Realių sekų vizualizavimas	<ul style="list-style-type: none"> - Sankio diagrama; - Akordo diagrama; - Tinklo diagrama

Taip pat, žinant kokiems duomenims analizė turi būti atliekama, įvertinami ir pridedami duomenų paruošimo žingsniai. Šiuo atveju analizė bus atliekama įvykių žurnalo duomenims, todėl pridedami žingsniai naudotojų sesijoms atrinkti ir puslapio tipui iš hipersaito išgauti.

Įvertinus turimas žinias apie duomenų paruošimą, analizės taikymą ir rezultatų išvedimą, sudaroma sekų analizės taikymo proceso metodikos diagrama. Paveikslėlyje 2.10 perteikiami reikalingi panaudojimo atvejai procesui išpildyti.



2.10 pav. Bendras informacinių sistemų naudotojų sekų analizės metodikos procesas

Galima įvardyti, jog sudarytas procesas nurodo duomenų analitikui kaip sukurti ataskaitą. Metodikoje naudojami ir jau prieš tai detalizuoti panaudojimo atvejai, ir papildomi, kuriuos reikia atlikti viso proceso metu.

Sudarytas metodikos procesas tinkamas, tik IS sistemoms, turinčioms įvykių žurnalą ir meniu struktūrą (analizė nepritaikoma vieno puslapio aplikacijoms). Atlikus įvykių žurnalo duomenų gavybą, pereinama prie duomenų paruošimo. Šiame žingsnyje yra išfiltruojamos tik vartotojų sesijos atmetant interneto botų ar automatinių testų sesijas ir subendrinamos nuorodos iki tipų. Taip turima naudotojų naršymo duomenų seka. Parenkamas galimas skirtingas duomenų analizės taikymas pagal tai, ką norima tirti arba yra pereinama prie realių sekų vizualizacijų. Kiekvienas analizės variantas geriausiai perteikia rezultatus pagal tam tikro tipo diagramas. Sudaryta metodika leidžia lengvai numatyti, ką galima tirti ir kaip atvaizduoti tyrimo rezultatus.

Turint sudarytą metodiką, galima be papildomo analizės taikymo galimybių tyrimo žinoti, kokie yra galimi duomenų analizės variantai ir vizualizavimo galimybės. Deja, nėra aišku kada kokį analizės kriterijų taikyti pagal informacinės sistemos tipą. Vienoms sistemoms gali būti neaktualūs tam tikros analizės rezultatai, kai kitoms sistemoms tai gali atrodyti labai svarbu. Analizės kriterijų taikymo svarbai pagal informacinės sistemos tipą bus pasitelkiamas ekspertinis įvertinimas sudarius apklausą.

2.4. Reikalavimų analizės apibendrinimas

Atliekant sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai tyrimą buvo išanalizuoti reikalavimai ir dalykinės srities esybės. Nustatyta kaip sekų analizės posistemis gali būti integruojamas įvykių žurnalo duomenims ir rezultatai išvedami per veiklos analitikos įrankį. Taip pat išskirtos esamos esybės ir tos, kurios papildytų dalykinę sritį. Sudarytas sekų analizės taikymo metodikos procesas, kuriame nurodomi galimi analizės kriterijai ir vizualizacijos galimybės. Nustatyta, kad reikalingas ekspertinis vertinimas siekiant išskirti kokius analizės kriterijus taikyti skirtingoms informacinėms sistemoms.

3. Sekų analizės posistemio eksperimentinės realizacijos projektas

Šiame skyriuje aprašomas analizės kriterijų nustatymas, reikiamai realizacijai išpildyti pasitelkus ekspertų apklausą ir planuojamas technologinis sprendimas.

3.1. Apklausos sudarymas

Eksperimentui atlikti ir analizės kriterijų parinkimui pagal informacinės sistemos tipą gauti planuojama naudoti analitinės hierarchijos procesą (AHP) [41]. Hierarchija, tai keletas lygių sistema kurių susideda iš daugumos elementų, arba faktorių. Pagrindinis hierarchijos klausimas, kaip stipriai atskiri žemiausio lygio faktoriai įtakoja pagrindinį tikslą [42]. Kriterijų svoriai atspindi ekspertų vertintojų nuomonę apie kriterijų svarbą lyginant su kitais kriterijais. Analitinis hierarchijos metodas pateikiamas kiekvienam ekspertui. Metodo pagrindą sudaro porinio palyginimo matrica. Joje pateikiama kiekvieno elemento palyginimas su kitu elementu ir prašoma parinkti vieno kiekvienos poros elementų svarbą. Taip pat įvedamas matas reiškiantis abiejų kriterijų lygybę.

Buvo sudaryta apklausa, kurioje naudotojų patyrimo ekspertai lygino analizės kriterijus tarpusavyje skirtingų informacinių sistemų tyrimui. Pirmiausiai respondentai turėjo pasirinkti informacinės sistemos tipą, vėliau skalėje nuo 1 iki 9 palikti įvertinimą. Čia 1 reiškia, kad svarbesnis yra kriterijus esantis kairėje. 5 reiškia, jog abu kriterijai vienodai svarbūs, o 9 reiškia, jog svarbesnis kriterijus esantis dešinėje. Ekspertams buvo pristatyti galimi kiekvienos analizės rezultatai ir tikslas, o tada pateikiama apklausa. Naudotos apklausos forma pateikta 1 priede.

3.2. Ekspertų apklausos rezultatai

Apklausoje dalyvavo 13 ekspertų, dirbančių su skirtingo tipo informacinėmis sistemomis. Dauguma jų – tai naudotojų patyrimo ekspertai, tačiau dar vertinimą atliko produkto vadovai ir verslo analitikai. Apklausos metu surinkta 15 įvertinimų, nes pora ekspertų gerai išmanė skirtingų informacinių sistemų tipus. Gauti rezultatai suvedami į AHP skaičiuoklę [43] ir taip nustatomi kiekvieno respondento išskirta vieta ir vietos užimamas procentas. Skirtingų ekspertų gauti įvertinimai bus suvidurkinami ir taip išvedami prioritetai kiekvienos planuojamos savybės įgyvendinimui, kiekvienam informacinės sistemos tipui.

3.2.1. Elektroninės komercijos sistemų analizės kriterijų prioritetai

Sudaryta anketa buvo užpildyta 4 skirtingų ekspertų, kurie analizuoja naudotojų patyrimą elektroninės komercijos platformose. Gauti rezultatai kiekvieno eksperto rezultatus išskirstė vietomis ir parodė kaip procentaliai kiekvienos savybės buvimas yra svarbesnis už kitus. Toliau 3.1 lentelėje pateikiami apklausos ir apibendrinti rezultatai.

3.1 lentelė. Sekų analizės savybių prioretizavimo rezultatai elektroninės komercijos sistemoms

	ekspertas 1		ekspertas 2		ekspertas 3		ekspertas 4		Apibendrinti rezultatai	
	%	vieta	%	vieta	%	vieta	%	vieta	%	vieta
1. Praleisto laiko analizė	6,30%	5	5,90%	5	4%	5	6,70%	5	5,73%	5
2. Dažnų sekų kombinacijos	9,40%	4	14,80%	4	10,60%	3	20,40%	2	13,80%	3

3. Dažnų sekų taisyklių radimas	15,80%	3	16,90%	3	7,40%	4	11,80%	4	12,98%	4
4. Nubyrėjimų analizė	20,50%	2	18,10%	2	25,80%	2	37,50%	1	25,48%	2
5. Vizualizavimas	45,30%	1	40,30%	1	49,20%	1	17,90%	3	38,18%	1
6. Klasterizavimas	2,60%	6	4%	6	2,90%	6	5,90%	6	3,85%	6

Kiekvieno respondento rezultatas buvo pakankamai panašus. Kiekvienas iš apklausos dalyvių aukščiausiai prioretizavo vizualizavimą. Tai parodo, kad galutiniam analizės vartotojui svarbiausia yra, kaip rezultatai pateikiami vizualiai. Antroje vietoje įvertinta nubyrėjimų analizė. Šios analizės rezultatai padėtų atskleisti kelias kuriuos vertėtų optimizuoti ar identifikuotą problemą. Trečioje ir ketvirtoje vietoje išskiriami dažnų sekų rinkinių ir taisyklių radimas. Šie padėtų nustatyti pačias populiariausias sekas ir jų elementus. Penktoje vietoje išskiriama praleisto laiko puslapiuose tyrimas. Mažiausiai svarbiu yra išskiriamas klasterizavimas.

3.2.2. Naujienų portalų ir interneto tinklaraščių analizės kriterijų prioritetai

Sudaryta anketa buvo užpildyta 3 skirtingų ekspertų. Gauti rezultatai kiekvieno eksperto rezultatus išskirstė vietomis ir parodė kaip procentaliai kiekvienos savybės buvimas yra svarbesnis už kitus. Toliau 3.2 lentelėje pateikiami apklausos ir apibendrinti rezultatai.

3.2 lentelė. Sekų analizės savybių prioretizavimo rezultatai naujienų portalų ir interneto tinklaraščių sistemoms

	ekspertas 1		ekspertas 2		ekspertas 3		Apibendrinti rezultatai	
	%	vieta	%	vieta	%	vieta	%	vieta
1. Praleisto laiko analizė	5,70%	6	29,40%	1	7,40%	4	14,17%	3
2. Dažnų sekų kombinacijos	14,80%	3	8,90%	5	25,80%	2	16,50%	2
3. Dažnų sekų taisyklių radimas	38,90%	1	9,10%	4	49,20%	1	32,40%	1
4. Nubyrėjimų analizė	8,50%	5	25,10%	2	4%	5	12,53%	5
5. Vizualizavimas	20,40%	2	8,80%	6	10,60%	3	13,27%	4
6. Klasterizavimas	11,70%	4	19%	3	2,90%	6	11,07%	6

Respondentų rezultatai buvo pakankamai skirtingi, tačiau pavyko išskirti prioritetus. Aukščiausiai prioretizuota dažnų sekų ir taisyklių radimas. Tai parodo, kad galutiniam analizės vartotojui svarbiausia kokius puslapius aplanko kartu, kokias aplankytų puslapių taisykles galima generuoti. Mažiausiai svarbiais kriterijais yra išskiriamas klasterizavimas ir nubyrėjimų analizė.

3.2.3. Reprerentacinis veiklos tinklapių analizės kriterijų prioritetai

Sudaryta anketa buvo užpildyta 3 skirtingų ekspertų. Gauti rezultatai kiekvieno eksperto rezultatus išskirstė vietomis ir parodė kaip procentaliai kiekvienos savybės buvimas yra svarbesnis už kitus. Toliau 3.3 lentelėje pateikiami apklausos ir apibendrinti rezultatai.

3.3 lentelė. Sekų analizės savybių prioretizavimo rezultatai reprerentacinės veiklos svetainėms

	ekspertas 1		ekspertas 2		ekspertas 3		Apibendrinti rezultatai	
	%	vieta	%	vieta	%	vieta	%	vieta
1. Praleisto laiko analizė	13,10%	6	29,40%	1	18,10%	2	20,20%	3
2. Dažnų sekų kombinacijos	10,60%	3	8,90%	5	4%	6	7,83%	6
3. Dažnų sekų taisyklių radimas	14,30%	1	9,10%	4	40,30%	1	21,23%	2
4. Nubyrėjimų analizė	27,40%	5	25,10%	2	16,90%	3	23,13%	1
5. Vizualizavimas	17,50%	2	8,80%	6	5,90%	5	10,73%	5
6. Klasterizavimas	17,10%	4	19%	3	14,80%	4	16,83%	4

Respondentų rezultatai buvo pakankamai skirtingi, tačiau pavyko išskirti prioritetus. Aukščiausiai prioretizuota dažnų sekų taisyklių radimas ir nubyrėjimų analizė. Tai parodo, kad galutiniam analizės vartotojui svarbiausia kokius puslapius dažniausiai aplanko kartu, kada naudotojai palieka puslapį. Mažiausiai svarbiais kriterijais yra išskiriamas vizualizavimas ir sekų kombinacijų radimas.

3.2.4. Interneto aplikacijų sistemų analizės kriterijų prioritetai

Sudaryta anketa buvo užpildyta 3 skirtingų ekspertų, kurie nuolat dirba su interneto aplikacijų sistemomis. Du iš apklaustųjų analizuoja naudotojų elgesį sąskaitų ir sutarčių administravimo informacinėje sistemoje. Gauti rezultatai kiekvieno eksperto rezultatus išskirstė vietomis ir parodė kaip procentaliai kiekvienos savybės buvimas yra svarbesnis už kitus. Toliau 3.4 lentelėje pateikiami apklausos ir apibendrinti rezultatai.

3.4 lentelė. Sekų analizės savybių prioretizavimo rezultatai interneto aplikacijų sistemoms

	ekspertas 1		ekspertas 2		ekspertas 3		Apibendrinti rezultatai	
	%	vieta	%	vieta	%	vieta	%	vieta
1. Praleisto laiko analizė	20,70%	2	33,80%	1	44,3%	1	27,25%	1
2. Dažnų sekų kombinacijos	12,20%	4	15,10%	4	7,4%	6	13,65%	3
3. Dažnų sekų taisyklių radimas	5,50%	6	15,30%	3	8,0%	5	10,40%	5
4. Nubyrėjimų analizė	7,50%	5	13,10%	5	15,3%	3	10,30%	6
5. Vizualizavimas	18,90%	3	5,90%	6	16,6%	2	12,40%	4
6. Klasterizavimas	35,10%	1	16,80%	2	8,30%	4	20,07%	2

Respondentų rezultatai buvo pakankamai skirtingi, tačiau pavyko išskirti prioritetus. Aukščiausiai prioretizuota praleisto laiko analizė ir klasterizavimas. Tai parodo, kad galutiniam analizės vartotojui svarbiausia kiek laiko naudotojai praleidžia sistemoje ir kaip galima skirstyti naudotojus į grupes. Mažiausiai svarbiais kriterijais yra išskiriamas vizualizavimas ir taisyklių gavimas.

3.2.5. Socialinių tinklų ir forumų analizės kriterijų prioritetai

Sudaryta anketa buvo užpildyta 2 skirtingų ekspertų, kurie analizuoja šio tipo sistemas. Gauti rezultatai kiekvieno eksperto rezultatus išskirstė vietomis ir parodė kaip procentaliai kiekvienos

savybės buvimas yra svarbesnis už kitus. Toliau 3.5 lentelėje pateikiami apklausos ir apibendrinti rezultatai.

3.5 lentelė. Sekų analizės savybių prioretizavimo rezultatai socialinių tinklų ir forumų informacinėms sistemoms

	ekspertas 1		ekspertas 2		Apibendrinti rezultatai	
	%	vieta	%	vieta	%	vieta
1. Praleisto laiko analizė	40,10%	1	8,50%	5	24,30%	2
2. Dažnų sekų kombinacijos	15,50%	2	20,40%	2	17,95%	3
3. Dažnų sekų taisyklių radimas	10,80%	5	14,80%	3	12,80%	4
4. Nubyrėjimų analizė	15,10%	3	5,70%	6	10,40%	5
5. Vizualizavimas	6,10%	6	11,70%	4	8,90%	6
6. Klasterizavimas	12,30%	4	39%	1	25,60%	1

Respondentų rezultatai buvo pakankamai skirtingi, tačiau pavyko išskirti prioritetus suvidurkinus rezultatus. Aukščiausiai prioretizuota praleisto laiko analizė ir klasterizavimas. Tai parodo, kad galutiniam analizės vartotojui svarbiausia kiek laiko naudotojai praleidžia sistemoje ir kaip galima skirstyti naudotojus į grupes. Mažiausiai svarbiais kriterijais yra išskiriamas vizualizavimas ir nubyrėjimų analizė.

3.3. Apklausos rezultatų apibendrinimas

Suvedus visų informacinių sistemų tipų apibendrintus rezultatus, sudaroma matrica, kurioje eilutėse nurodomi analizės kriterijai, stulpeliuose informacinių sistemų tipai, o susikirtime prioritetas. Vienetu išreiškiama, jog tam tikras analizės kriterijus yra pats svarbiausias atitinkamai informacinei sistemai, o šešetu nurodoma, jog atitinkamas analizės kriterijus yra mažiausiai aktualus. Toliau 3.6 lentelėje pateikiami apibendrinti rezultatai ir spalvų intensyvumu paryškunami prioritetai.

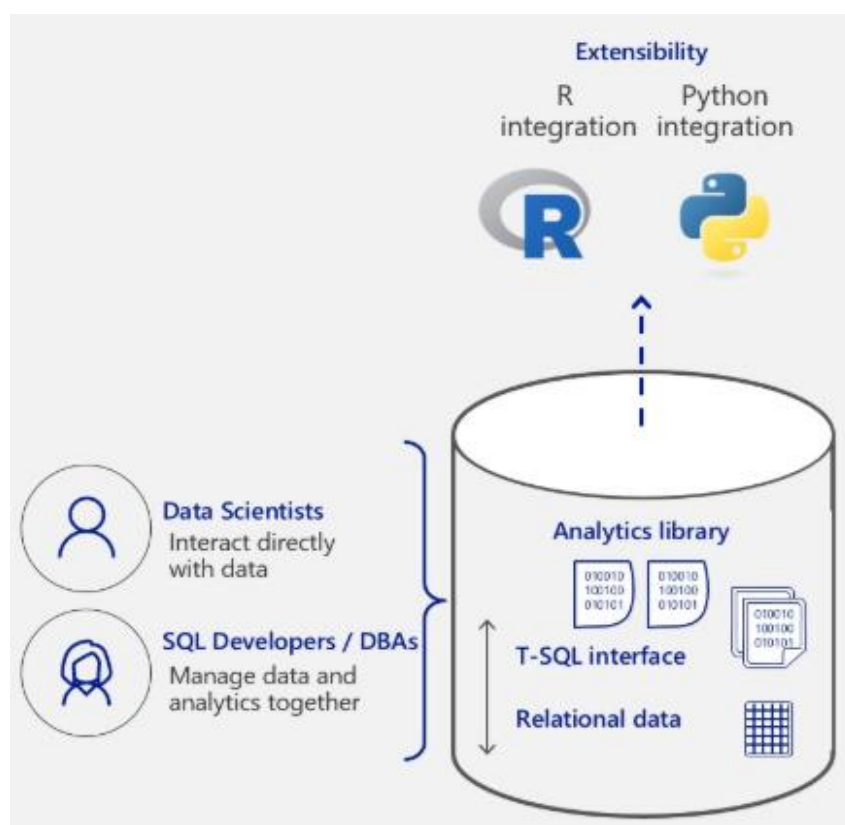
3.6 lentelė. Sekų analizės prioretizavimo rezultatai skirtingoms informacinėms svetainėms

	Elektroninės komercijos IS	Interneto aplikacijų IS	Socialinių tinklų ir forumų IS	Reprezentacinės veiklos tinklapiai	Naujų portalų ir interneto tinklaraščių IS
Praleisto laiko analizė	5	1	2	3	3
Dažnų sekų kombinacijos	3	3	3	6	2
Dažnų sekų taisyklių radimas	4	5	4	2	1
Nubyrėjimų analizė	2	6	5	1	5
Vizualizavimas	1	4	6	5	4
Klasterizavimas	6	2	1	4	6

Susumuoti analizės rezultatai parodė, jog ekspertų vertinimai skirtingai išdėstė prioritetus tarp skirtingų informacinių sistemų tipų. Planuojant sprendimo realizaciją, mažiausiai svarbūs analizės kriterijai gali būti neišpildomi iš taip išvengiama analizės rezultatų gausa. Nustatyta, jog elektroninės komercijos sistemoms rekomenduojama atlikti nubyrežimų analizę ir realių sekų vizualizavimą. Interneto aplikacijų informacinėms sistemoms ir socialinių tinklų ir forumų IS atlikti praleisto laiko tyrimą ir klasterizavimą. Reprezentacinės veiklos sistemoms atlikti nubyrežimų analizę ir dažnų sekų taisyklių radimą, o naujienų portalų ir interneto tinklaraščio IS rasti dažnų sekų kombinacijas ir dažnų sekų taisykles.

3.4. Technologinis realizacijos sprendimas

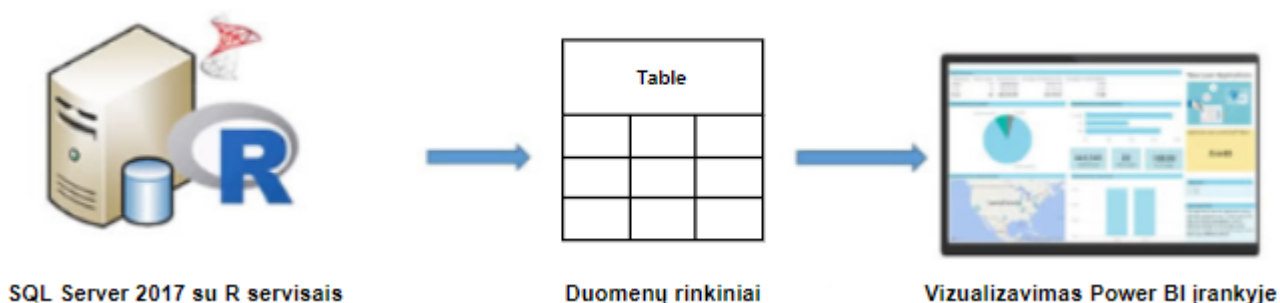
SQL Serveris nuo 2016 versijos palaiko analitikos scenarijus parašytus R, o nuo 2017 versijos kartu ir Python kalba. Tiesiai į duomenų bazę integruoti algoritmai ar mašininio mokymo modeliai supaprastina architektūrą, generuoja analizės rezultatus greičiau, saugo rezultatus vienoje vietoje kartu su originaliais duomenimis [44]. Iliustracijoje 3.1 pav. pavaizduota SQL Server 2017 analitikos galimybių modelis.



3.1 pav. SQL serveryje vykdomos analitikos modelis [45]

Naudojant reliacinius duomenis, T-SQL transformavimo kalbą ir analitikos bibliotekų integravimą, duomenų apdorojimas ir analitikos taikymas yra įgyvendinamas vienoje vietoje. Žinoma, saugant įvykių žurnalo duomenis toje pačioje duomenų bazėje su pačia aplikacija, galima pakenkti pačios aplikacijos veikimui ir greičiui. Turint nuo aplikacijos atskirtą SQL serverį įvykių žurnalo duomenims, toks sprendimas tiktų ir išspręstų papildomą scenarijų diegimo klausimą.

Gavus analizės duomenis toje pačioje duomenų bazėje, juos nesunkiai galima susieti su veiklos analitikos įrankiais. Juose atlikti duomenų vizualizavimą, užtikrinti duomenų atsinaujinimą. Toks duomenų panaudojimas vaizduojamas 3.2 pav.

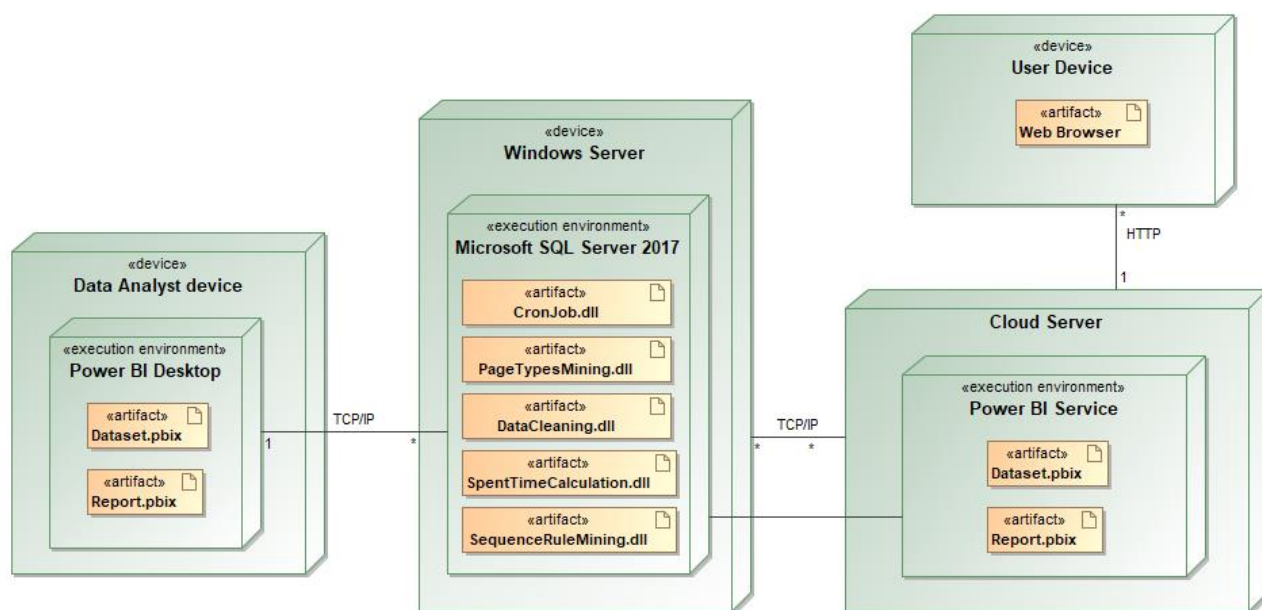


3.2 pav. Analizės rezultatų panaudojimas Power BI įrankyje

Parodomas abstraktus modelis kaip duomenys yra paruošiami SQL Server 2017 kartu naudojant R servisus ir taip išgaunant naujus duomenų rinkinius. Duomenų rinkiniai gali būti importuojami į veiklos analitikos įrankį „Power BI“ ir taip sukuriama interaktyvi ataskaita.

3.5. Sekų analizės sprendimo realizacijos modelis

Įvertinus galimą realizacijos išpildymą duomenų bazės serveryje, 3.3 pav. vaizduojama diegimo diagrama. Joje pateikiama sekų analizės posistemio diegimas į SQL serverį ir ataskaitos kūrimas „Power BI“ Desktop versijoje. Taip pat nurodomi kuriami artefaktai.



3.3 pav. Diegimo diagrama

Sprendimo įgyvendinimui naudojamas Microsoft SQL Server 2017. Planuojant diegti R kalba paruoštus scenarijus, minimaliai rekomenduojama ši arba naujesnė SQL Server versija. Duomenų bazėje integruojamos procedūros, galinčios sekų analizei reikiamus duomenis. Taip pat nustatoma

laikmačio procedūra, galinti inicijuoti naujų duomenų paruošimą atitinkamu laiku. Pavyzdžiui kas 7 dienas. Kuriant ataskaitą veiklos analitikos įrankiu „Power BI“, originaliai turimi aplankytų puslapių ir sesijų duomenys bei paruošti duomenų rinkiniai importuojami į Veiklos analitikos įrankį „Power BI“. Duomenų analitikas šio įrankio „Desktop“ versijoje sukuria ataskaitą parenkant reikšmingiausias vizualizacijas. Sukurta ataskaita yra išpublikuojama į „Power BI Service“ versiją, kuri yra debesyse. Ši versija saugo ataskaitos ir duomenų rinkinio kopijas. Taip, nustačius duomenų atnaujinimą, „Service“ versija naudotų duomenų rinkinius tiesiai iš SQL serverio. Sukurtos ataskaitos naudotojai per interneto naršyklę pasiektų „Service“ ataskaitos versiją.

Turint sekų analizės taikymo gaires, išsiaiškinus analizės variantų taikymo rekomendacijas skirtingo tipo informacinėms sistemoms ir įvertinus technologinius aspektus įvykių žurnalo duomenų saugojime, pasiruošta atlikti realizaciją. Planuojama realizacija apimtų skirtingų tipų informacines sistemas, siekiant parodyti galimus skirtingų analizės variantų rezultatus.

4. Sekų analizės realizacija ir testavimas

Šiame skyriuje aprašoma sekų analizės įrankio realizacija ir testavimas. Detalizuojami žingsniai reikalingi duomenų tvarkymui ar ataskaitos ruošimui.

4.1. Sprendimo realizacijos ir veikimo aprašas

Realizacijai reikalingi informacinių sistemų įvykių žurnalo duomenys, kurie būtų saugomi nuo aplikacijos atskirtoje Microsoft SQL duomenų bazėje. Gali būti įvykių žurnalo duomenys pasiekiami *.txt* įrašytame faile, įmonės serveryje. Tokiu atveju reikia duomenis importuoti į duomenų bazę. Numatytas sprendimo realizacijos būdas – integruoti analitikos sprendimus SQL serveryje, todėl naudojami sukaupti duomenys SQL serveryje. Toliau detaliam apžvelgiamos turimos ir kuriamos duomenų struktūros, nustatomi reikalingi duomenų modeliai pagal informacinės sistemos tipą.

4.1.1. Duomenų tvarkymo procedūra

Prieš atliekant sekų gavybą ar puslapių subendrinimą iki tipų, pirmiausiai atliekamas duomenų tvarkymo žingsnis. Šiame žingsnyje teisinga tvarka surikiuojami duomenys, panaikinami dublikatai, išfiltruojamos netinkamos sesijos. Dažniausiai skirtingos informacinės sistemos saugo įvykių žurnalo duomenis panašiu struktūrizuotu formatu. Realizacijoje tiriamų sistemų atveju, visos informacinės sistemos saugo įvykių žurnalo duomenis vienuodu formatu. Vienoje lentelėje saugomi duomenys apie sesiją, kitoje apie aplankytus puslapius. Sesijos duomenys pateikiami 4.1 lentelėje, o aplankytų puslapių 4.2 lentelėje.

4.1 lentelė. Pavyzdinė sesijų įvykių žurnalo duomenų struktūra

Session_Key	EndTime	Host	Remote Addr	User_agent	Browser	Browser version	Traffic type	OS	Automated_test
38f033a6ef818e907c9015208b87f979	2021-04-19 16:31:16	www.atea.fi	193.166.253.150	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 Chrome/89.0.4389.128 Safari/537.36	Chrome	Chrome89.0	BROWSER	Win10	False
aa57bd6b490190529ba6fa3b52f08a18	2021-04-19 16:31:19	www.atea.se	141.98.255.143	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 Chrome/75.0.3770.145 Safari/537.36	Chrome	Chrome75.0	BROWSER	Win10	False
9ecafb08857fb2a1b44328b68f67608e	2021-04-19 16:31:27	www.atea.no	198.23.235.123	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/59.0.3071.115 Safari/537.36	Chrome	Chrome59.0	BROWSER	Win10	False

Čia kiekvienas stulpelis reiškia:

- „Session_Key“, varchar. Identifikuoja unikalios sesijos id;
- „EndTime“ datetime. Sesijos pabaigos laiko žymė;
- „Host“, varchar. Sistemos domenai;
- „Remote_Addr“, varchar. Lankytojo IP adresas;
- „User_agent“, varchar. Naršyklės naudotojo agento tipas;
- „Browser“, varchar. Naudota naršyklė;
- „Browser_version“, varchar. Naudotos naršyklės versija;
- „Traffic_type“, varchar. Naudoto prietaiso tipas;
- „OS“, varchar. Operacinė sistema;
- „Automated_test“ boolean. Požymis ar automatinių testų generuota sesija.

4.2 lentelė. Pavyzdinė aplankyto puslapių įvykių žurnalo duomenų struktūra

Session	Hit	Page	Req	Rsp
38f033a6ef818e907c9015208b87f979	1	www.atea.fi/eshop/product/startech-com-8k-usb-c-to-displayport-adapter/	2021-04-19 16:20:47	2021-04-19 16:20:47
aa57bd6b490190529ba6fa3b52f08a18	1	www.atea.se/eshop/product/hpe-proliant-microserver-gen10/	2021-04-19 16:28:11	2021-04-19 16:28:11
aa57bd6b490190529ba6fa3b52f08a18	2	www.atea.se/eshop/products/	2021-04-19 16:29:51	2021-04-19 16:29:52

Čia kiekvienas stulpelis reiškia:

- „Session“, varchar. Identifikuoja unikalios sesijos id;
- „Hit“ integer. Nurodo aplankyto puslapio eilę;
- „Page“, varchar. Nurodo aplankyto hipersaito nurodą;
- „Req“ datetime. Puslapio generavimo užklauskos siuntimo laiko žymė;
- „Rsp“ datetime. Puslapio generavimo atsakymo gavimo laiko žymė.

Nuskaičius aplankytus puslapius, atliekamas rikiavimas pagal „Session“ ir „Hit“ laukus didėjančiai siekiant užtikrinti, jog sekos elementai, būtų išdėstyti teisingai. Toliau iš aplankyto puslapių hipersaitų nustatomos ir pašalinamos neprodukcinės aplinkos sesijos. Šio eksperimento atveju, radus sesijoje nors vieną hipersaitą prasidedantį „*wwwtest.atea...*“ arba „*wwwqa.atea...*“ šalinama visa sesija.

Toliau, panašiu principu kaip buvo naikinamos ne vartotojų generuotos sesijos – naikinamos ir sesijos inicijuotos automatinių testų ir interneto botų. Šio eksperimento atveju, naikinamos sesijos ir aplankyto puslapių įrašai, jei „Automated_test“ turėjo reikšmę „True“ ar „User_agent“ turėjo nors vieną iš šių įrašų („google“, „yahoo“, „bing“, „duckduckgo“, „yandex“, „HeadlessChrome“).

4.1.2. Puslapių subendrinimas iki tipų

Norint atlikti dažnų sekų gavybą, atliekama procedūra, kurios metu kiekvienas hipersaitas yra priskiriamas tipui. Puslapio tipas gali būti išgaunamas ir pagal hipersaito struktūrą, pavyzdžiui imant hipersaitą iki tam tikro lygio. Šio eksperimento metu interneto tinklaraščio aplankytiems puslapiams bus išgaunamas paskutinė hipersaito dalis nesiekiant tikslumo, o tik išgaunant straipsnių pavadinimus. Kitoms tiriamoms sistemoms, ir jų aplankytiems puslapiams, siekiant tikslumo, tipas buvo išgaunamas atlikus šablono paiešką hipersaite. Norint išgauti tipus šiuo metodu, reikia išmanyti portalo struktūrą.

Sprendimo realizacijos metu pagal šabloną suteikti tipai pateikiami 2 priede. Buvo numatyti ir išskirti 46 tipai elektroninės komercijos portalui ir 9 tipai sąskaitų administravimo sistemai. Išgautus tipus, saugant vietą, galima įterpti kaip naują stulpelį aplankyto puslapio lentelėje.

4.1.3. Praleisto laiko skaičiavimas

Atliekamas papildomas žingsnis sukuriamas naujas aplankytų puslapių duomenų rinkinys įėjimo-išėjimo formatu. Taip pat apskaičiuojamas praleistas laikas sekundėmis tarp kiekvieno gretimo sekos elemento sesijoje. Laikas skaičiuojamas imant puslapio generavimo atsakymo (Rsp) laiko žymę. Toliau išskaičiuojami praleisto laiko procentali dalis sesijoje. 4.3 lentelėje vaizduojama pavyzdinė struktūra.

4.3 lentelė. Pavyzdinė aplankytų tipų ir praleisto laiko duomenų struktūra

Session	Rsp	Source	Sourceid	Target	Targetid	Hit_end	Timediff	Percent
001a43ee73cf1d9e1439d632979d884d	2021-04-24 22:30:20	Home-page	1	Tickets	2	2021-04-24 22:30:26	6,17	36
001a43ee73cf1d9e1439d632979d884d	2021-04-24 22:30:26	Tickets	2	Contracts	3	2021-04-24 22:30:37	11,15	64
0027c32319ddf3f695e47cf820a921ca	2021-04-21 12:00:44	Home-page	1	Other	2	2021-04-21 12:00:55	11,67	100

Ši suformuota struktūra įėjimo-išėjimo formatu gali būti ne tik praleisto laiko analizei atlikti, bet ir skirtingoms vizualizacijoms kurti. Pavyzdžiui, būtent sankio, akordo ar piltuvėlio diagramoms reikalingas įėjimo-išėjimo formatas.

Norint atlikti praleisto laiko apimties ir trukmės vizualizacijas, reikia atlikti papildomą transformaciją ir kiekvienam procentui priskirti aplankyto puslapio tipą. Toks transformavimas reikalauja kiekvienam procentui sukurti atskirą stulpelį. Galima procentus stambinti ir išskirti didesniu žingsniu nei vienas procentas, tačiau šio tyrimo atveju paliekamas vieno procento žingsnis. Pavyzdinė praleisto laiko procentų duomenų struktūra pateikiama 4.4 lentelėje.

4.4 lentelė. Pavyzdinė praleisto laiko duomenų struktūra pagal apimties procentą

Session	1	2	3	...	100
001a43ee73cf1d9e1439d632979d884d	Tickets	Tickets	Tickets	...	Orders
aa57bd6b490190529ba6fa3b52f08a18	Home-page	Home-page	Other	...	Other

Tokiu formatu išsaugotą duomenų rinkinį, naudojant „TraMineR“ paketą, galima ne tik vizualizuoti praleisto laiko dažnį ir apimtį, bet ir klasterizuoti sekas. Taip pat norint rasti dažnas sekas pagal praleistą laiką, būtent šis formatas ir yra reikalingas. „Power BI“ įrankyje pasirinkus vizualizacijas iš R kodo, turint taip paruoštą duomenų rinkinį, galima panaudoti ir vizualizuoti ne vieną su laiku susijusią diagramą.

4.1.4. Dažnų sekų radimas

Turimas duomenų failas su nustatytais tipais transformuojamas į formatą kai vienos sesijos aplankyti puslapiai yra pateikiami vienoje eilutėje atskirti tam tikru simboliu, pavyzdžiui kableliu. Toks formatas reikalingas norint rasti dažnas sekas naudojant SPADE ar Apriori algoritmus. Pavyzdinė duomenų struktūra pateikiama 4.5 lentelėje.

4.5 lentelė. Pavyzdinė skaitytų straipsnių duomenų struktūra

Session	Articles
e4dcba2ccfa3a1206f8988c5bce3cd48	slik-unngar-du-slosesjokket
39db19c06f9caa55eb4979e89a74d569	leie-i-stedet-for-eie,slik-unngar-du-slosesjokket
936613c7efbe2a6dd5ff4097226847c0	ti-gode-grunner-til-a-velge-apple,slik-unngar-du-slosesjokket

Vienoje eilutėje įrašyti aplankyti puslapiai atskirti specialiu simboliu nuskaitomi specialiu formatu ir nusprendus ar tiriamame rinkinyje svarbi sekos elementų išdėstymo tvarka, parenkamas SPADE ar Apriori algoritmas. Algoritmu išgaunamos dažnos sekos nustačius minimalius dažnumo koeficientus, o sekų taisyklės pagal algoritme nustatytus minimalius dažnumo ir patikimumo koeficientus. Šio tyrimo metu, algoritme naudojami labai minimalūs koeficientai, taip leidžiant išgauti kuo daugiau dažnų sekų ir taisyklių variacijų, o galutinėje ataskaitoje vartotojui pridedami gautų parametrų intervalai ir leidžiama filtruoti pagal šiuos koeficientus.

4.1.5. Įėjimo ir išėjimo taškų radimas

Siekiant rasti nubyrėjimus, apskaičiuoti vidutinį sesijos ilgį ar apskaičiuoti galimus sesijos pradžios ar pabaigos taškus, transformuojama sesijos aplankytų puslapių seka į vieną eilutę atskiriant tam tikru simboliu, apskaičiuojamas sesijos ilgis bei sesijos pradžios ir pabaigos taškai. 4.6 lentelėje pateikiama pavyzdinė transformuotos sekos struktūra.

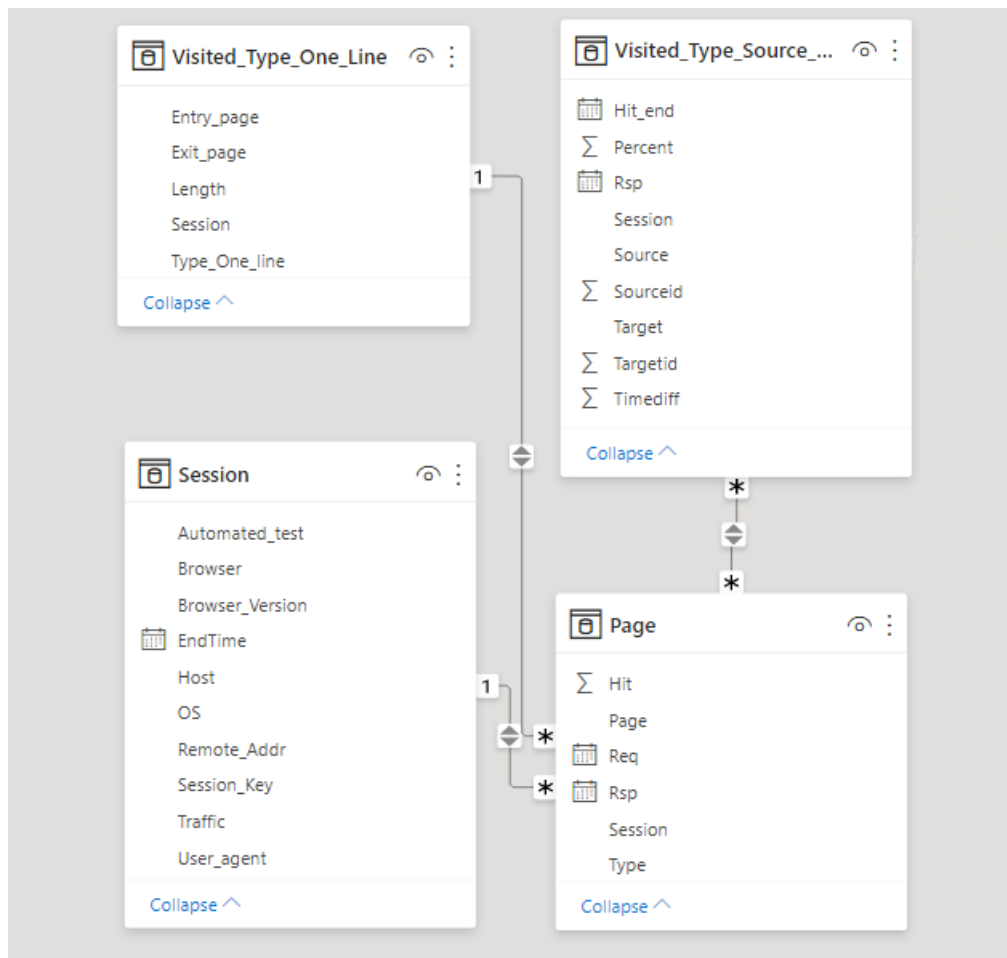
4.6 lentelė. Pavyzdinė įėjimo išėjimo duomenų struktūra

Session	One_line	Length	Entry_page	Exit_page
e4dcba2ccfa3a1206f8988c5bce3cd48	Home-page,Product	2	Home-page	Product
001a43ee73cfd9e1439d632979d884d	Home-page	1	Home-page	Home-page
0027c32319ddf3f695e47cf820a921ca	Product,Checkout-cart,Checkout-information,Checkout-receipt	4	Product	Checkout-receipt

Taip transformuota seka naudinga ir ataskaitoje išpildant filtravimo galimybę.

4.1.6. Ataskaitos ruošimas

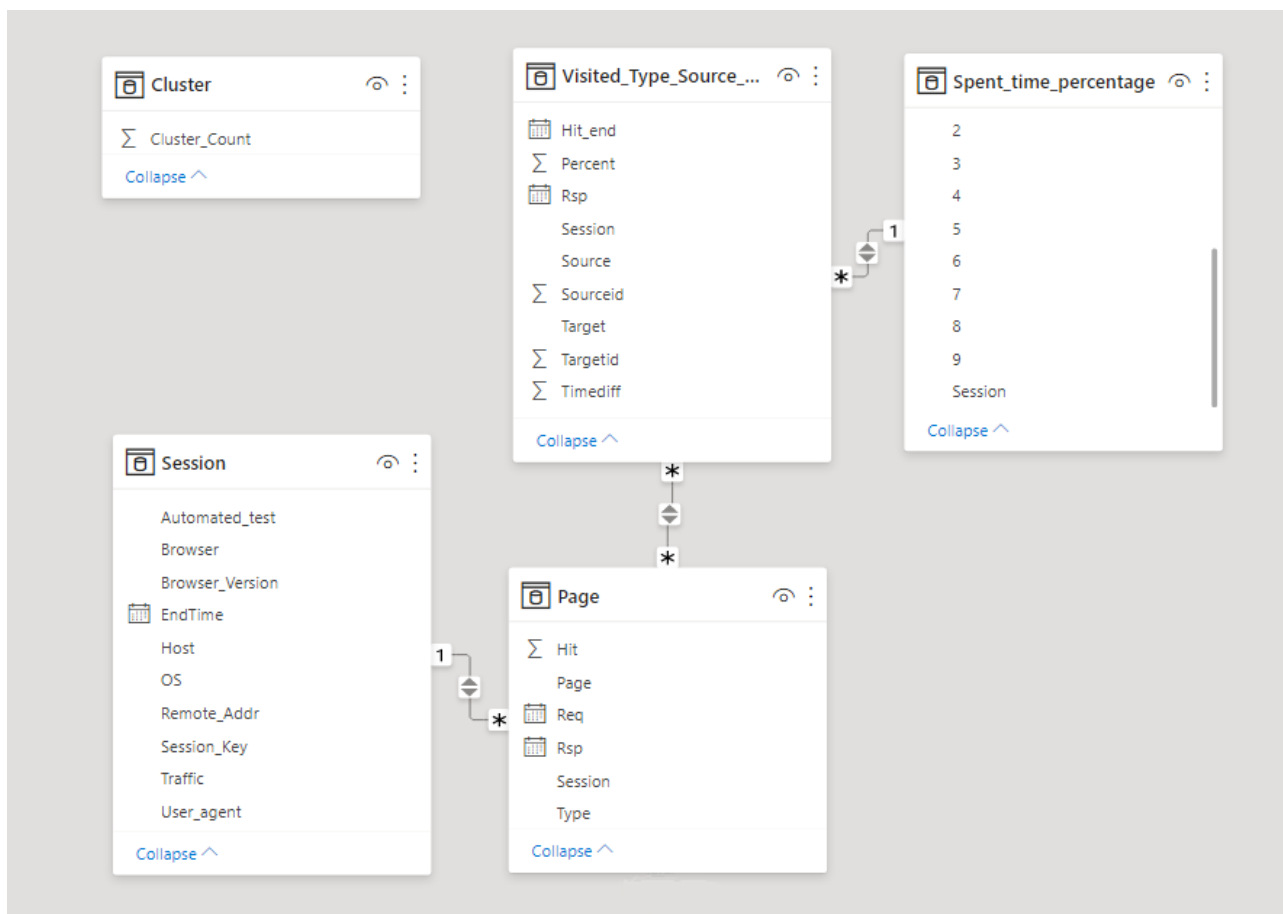
Duomenų paruošimo metu gauti duomenų rinkiniai buvo importuoti į Power BI veiklos įrankį. Čia duomenų rinkiniai sujungiami tarpusavyje ryšiais. Pateikiami keturi skirtingi duomenų modeliai. Duomenų modeliai pritaikyti kiekvieno tipo informacinei sistemai, išpildant aukščiausią vietą turinčius analizės variantus ir jų vizualizavimo galimybes. Kiekviename modelyje paliekamas ir turimų egzistuojančių sesijos ir aplankyto puslapio duomenų lentelės. Toliau 4.1 pav. pateikiamas pavyzdinis duomenų modelis, skirtas elektroninės komercijos informacinėms sistemoms.



4.1 pav. Vizualizacijoms ir nubyrejimams reikalinga duomenų struktūra, skirta elektroninės komercijos informacinėms sistemoms, veiklos analitikos įrankyje "Power BI"

Be originaliai turimų duomenų apie sesiją ir puslapį, realių sekų vizualizavimui naudojama duomenų struktūra, kurioje yra seka transformuota įėjimo-išėjimo formatu. Tokia transformacija reikalinga realių sekų vizualizacijoms. Taip pat, naudojama transformuota seka į vieną eilutę. Šioje lentelėje yra informacija apie sekos ilgį ir sesijos pradžios ir pabaigos taškus, kurie padės nubyrejimų analizėje.

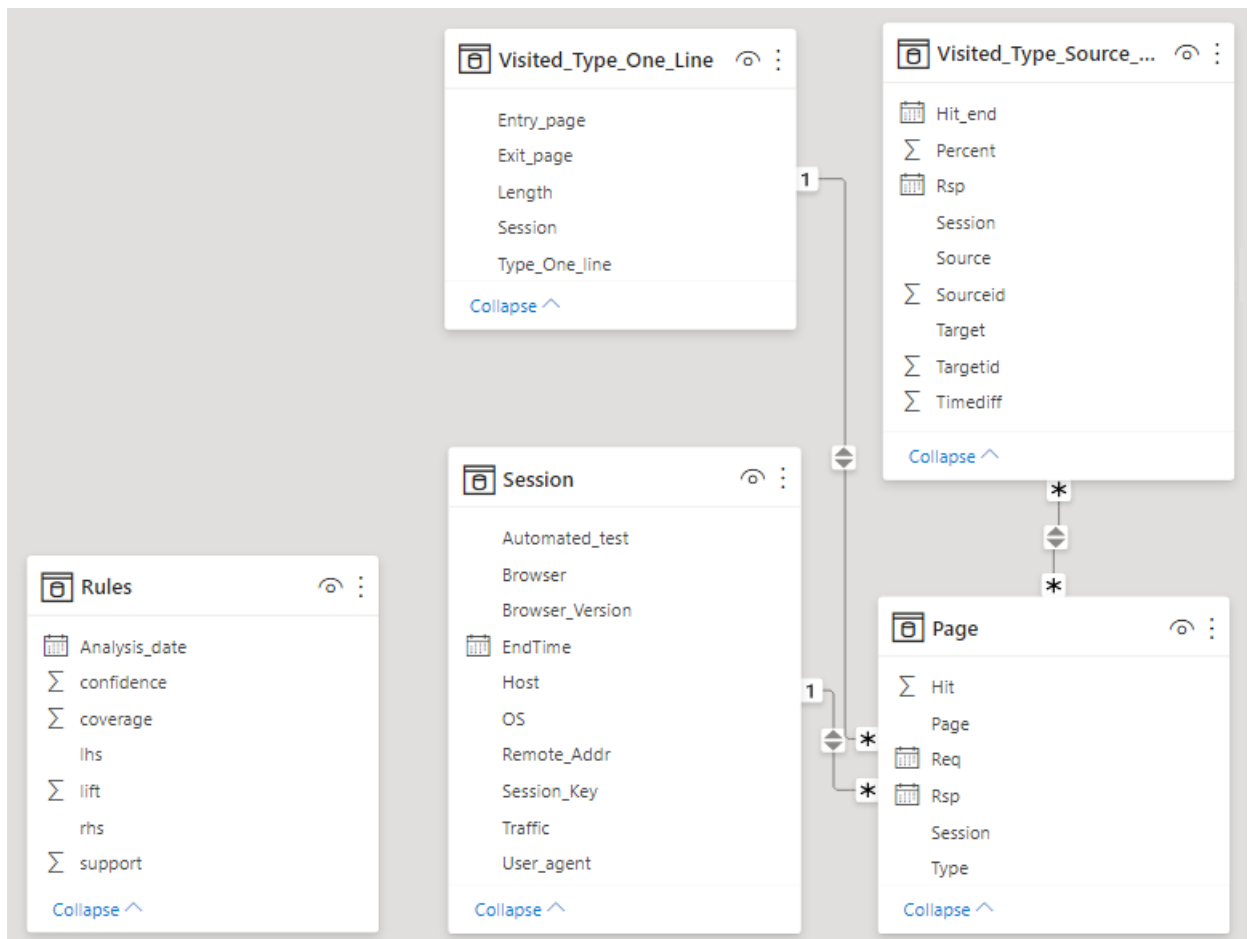
Kita naudota duomenų struktūra sudaryta interneto aplikacijų ir socialinių tinklų, forumų informacinėms sistemoms. Šioms sistemoms, pagal nustatytus prioritetus svarbus praleisto laiko tyrimas ir sekų klasterizavimas. Duomenų modelis sudarytas taip, jog būtų galima išpildyti šiuos kriterijus. Modelio pavyzdys pateikiamas 4.2 pav.



4.2 pav. Praleisto laiko skaičiavimui ir klasterizavimui reikalinga duomenų struktūra, skirta interneto aplikacijų ir socialinių tinklų, forumų informacinėms sistemoms, veiklos analitikos įrankyje „Power BI“

Be turimų duomenų apie sesiją ir puslapį, praleisto laiko analizei reikalinga transformuota seka įėjimo išėjimo formatu, o labiau praleistos trukmės procentui priskirtas tipas. Būtent naudojant šį formatą, pasirinkus vizualizavimui R kalbos parinktį, naudojant „TraMineR“ paketo komandas, vizualizuojamos naršymo sekos pagal praleistą laiką. Norint skirstyti sekas į grupes, įvedama tarpinė klasterių skaičiaus lentelė. Ši lentelė neturi jokių ryšių su kitomis turimomis lentelėmis. Šią lentelę siekiama naudoti kaip filtrą, kurio parinkimas perskirstytų sekas į tokį klasterių skaičių, koks parinktas filtre.

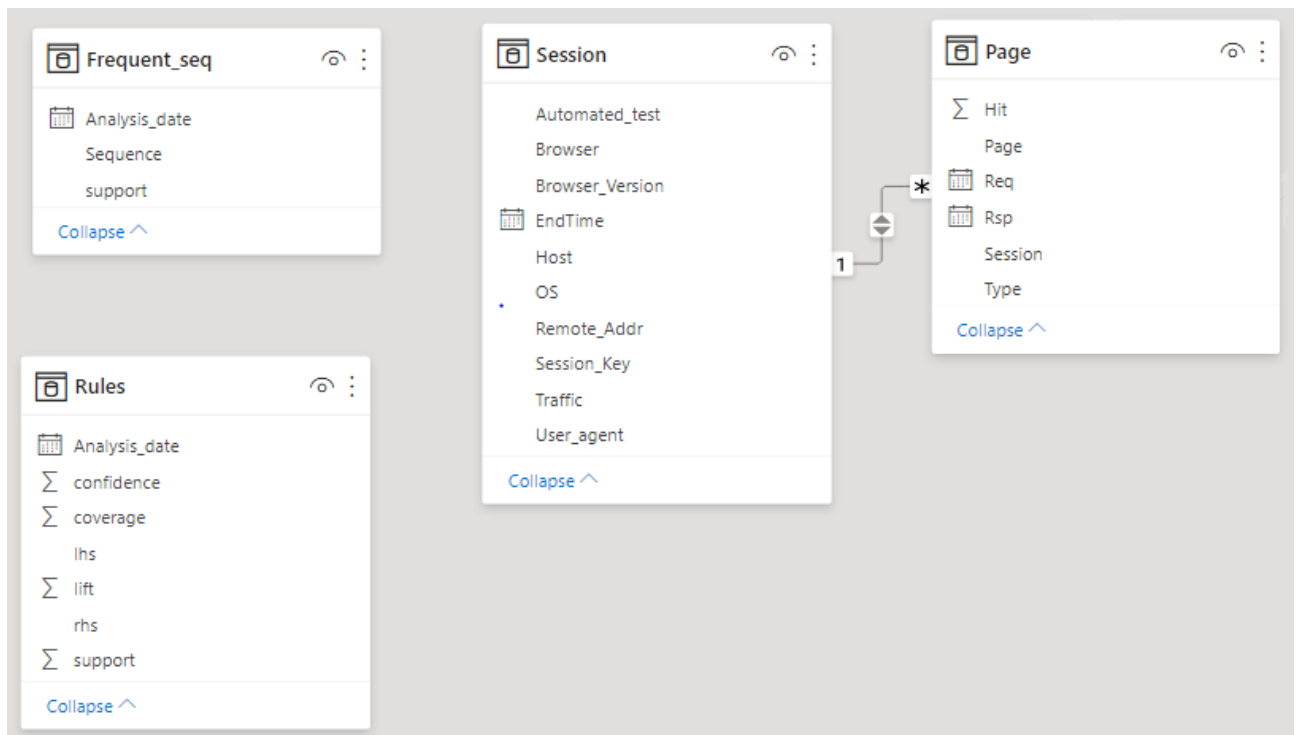
Kita galimas duomenų modelis sudarytas reprezentacinės veiklos informacinėms sistemoms. Šių sistemų duomenims rekomenduojama tirti nubyrėjimus ir rasti dažnų sekų taisykles. Gautas duomenų modelis yra labai panašus į modelį skirtą elektroninės komercijos informacinių sistemų duomenims, tik papildomai reikalinga taisyklių lentelė. Pavyzdinis modelis pateikiamas 4.3 pav.



4.3 pav. Nubyrėjimų ir taisyklių radimui reikalinga duomenų struktūra, skirta reprezentacinės veiklos informacinėms sistemoms, veiklos analitikos įrankyje „Power BI“

Be originaliai turimų duomenų apie sesiją ir puslapį, piltuvėlio diagramos vaizdavimui reikalinga transformuota seka įėjimo-išėjimo formatu. Dar, nubyrėjimams rasti, reikalinga ir transformuota seka į vieną eilutę. Šioje lentelėje yra informacija apie sekos ilgį ir sesijos pradžios ir pabaigos taškus, kurie padės nubyrėjimų analizėje. Šio tipo informacinėms sistemoms, be nubyrėjimų svarbus ir sekų taisyklių gavimas. Gautos taisyklės neturi jokio ryšio su ryšio su turimomis lentelėmis. Turimų taisyklių rezultatai gali būti tiesiog atvaizduojami taisyklių lentelė, nes tai jau yra analizės rezultatas. Siekiant didesnio aiškumo, galima taisyklių vizualizavimas tinklo diagrama.

Paskutinė naudota duomenų struktūra kurta siekiant tirti naujienų portalų ir interneto tinklaraščių informacinių sistemų duomenis. Pagal ekspertų vertinimus, buvo nustatyta, jog šio tipo sistemoms svarbiausia išgauti dažnų sekų rinkinius ir taisykles. Pavyzdinis duomenų modelis pateikiamas 4.4 pav.



4.4 pav. Dažnų sekų ir taisyklių radimui reikalinga duomenų struktūra, skirta interneto tinklaraščių ir naujienų portalų informacinėms sistemoms, veiklos analitikos įrankyje „Power BI“

Turimos lentelės apie sesiją ir puslapį. Importuoti rastų dažnų sekų ir taisyklių rezultatai nesisieja su kitomis lentelėmis. Gautus analizės rezultatus galima perteikti ir tik iš gautų dažnų sekų ir taisyklių lentelių, tačiau norint parodyti, kiek sesijų ar puslapių buvo tiriama, galima pritaikyti ir originalių lentelių duomenis.

Paruošus duomenų modelį, galima pereiti prie vizualizacijų kūrimo ir jų išdėliojimo ataskaitoje.

4.2. Testavimo modelis, duomenys, rezultatai

Sekų analizės posistemio testavimu svarbiausia užtikrinti ar duomenų tvarkymo žingsnyje teisingai paruošiami duomenys, priskiriami tipai, atliekamos teisingos transformacijos ir praleisto laiko apskaičiavimas. Taip pat svarbu užtikrinti ar veikia inicijavimas automatiniam duomenų atnaujinimui. Grafinė vartotojo sąsaja, kuria atvaizduojama atlikta ataskaita veiklos analitikos įrankiu „Power BI“ nebus testuojama pagal funkcinį testavimą. Skaitoma, kad veiklos analitikos įrankis „Power BI“ veikia tinkamai ir atskirų jo komponentų testuoti neverta, tik integraciją su kuriu sprendimu.

Toliau lentelėse 4.6 ir 4.7 pateikiami išskirti testavimo scenarijai funkciniam ir nefunkciniam reikalavimams tirti.

4.7 lentelė. Funkcinio testavimo scenarijai

Testuojamas reikalavimas	Testavimo eiga	Testavimo rezultatas
Duomenų paruošimas	Patikrinti ar po filtravimo procedūros duomenų rinkinyje nėra automatinų testų ar interneto botų sesijų	Originaliuose duomenyse surastos sesijos, kurios buvo generuotos automatinų testų ir patikrinta, kad

		šių sesijos puslapių nėra duomenų rinkinyje po filtravimo
	Patikrinti ar po filtravimo procedūros duomenų rinkinyje nėra nereprodukcinių aplinkų sesijų	Originaliuose duomenyse surastos sesijos, kurios buvo generuotos iš programuotojams ir testuotojams skirtų vidinių aplinkų, patikrinamos, duomenų rinkinyje po filtravimo
	Patikrinti ar teisingai priskiriami tipai aplankytiems puslapiams	Išvestame originalių puslapių ir priskirtų tipų sąrašė nerasta neatitikimų
Praleistos trukmės skaičiavimas	Patikrinti ar teisingai apskaičiuojama praleisto laiko trukmė ir procentali dalis tarp dviejų sekos elementų sesijoje turint kiekvieno aplankyto puslapio pradžios laiko žymę	Peržiūrėtos kelios skirtingos sesijos ir užtikrinta kad praleista laiko trukmė sekundėmis apskaičiuota teisingai tarp skirtingų sekos elementų sesijoje ir elementams tarp skirtingų sesijų
Ataskaitos atnaujinimas	Nustatyti automatinį ataskaitos atnaujinimą ir patikrinti ar realizuota ataskaita papildoma duomenimis	Nustačius automatinį atnaujinimą, po atnaujinimo užtikrinama, kad atsirado naujų sekų ataskaitoje

4.8 lentelė. Nefunkcinio testavimo scenarijai

Testuojamas reikalavimas	Testavimo eiga	Testavimo rezultatas
Analizės atlikimo greitis	Patikrinama kiek laiko trunka analizės atlikimas su skirtingu duomenų kiekiu nuo duomenų nuskaitymo iki duomenų išvedimo	Randamos didžiausią laiką užimančios procedūros, ieškoma optimesnių duomenų tvarkymo žingsnių
Savybių išpildymo tvarka	Patikrinama ar savybės gavusios aukščiausią vietą ekspertų vertinime, yra pateikiamos pirmuosiuose ataskaitos puslapiuose	Aukščiausiai įvertintoms savybėms ataskaitoje, skiriama daugiausiai vietos, rezultatai išvedami pirmuosiuose puslapiuose

Patvirtinus visus funkcinis reikalavimus ir įsitikinus, jog duomenys yra atrenkami ir transformuojami teisingai, ar reikiami laukai apskaičiuojami teisingai, galima patvirtinti turimų duomenų kokybę. Taip pat, reikia patikrinti ar ataskaitos atnaujinimas atliekamas tinkamai ir po atnaujinimo inicijavimo pasipildo nauji duomenys. Papildomai patikrinus analizės atlikimo laiką ir įsivertinus ar atlikimo laikas yra tinkamas, galima patvirtinti sekų analizės posistemio kokybę. Sudarant ataskaitą, papildomai atkreipiamas dėmesys elementų išdėstymui ir stengiamasi išdėlioti elementus taip, kad aukštesniu prioritetu įvertinti analizės variantai būtų pateikiami pirmiau.

Detaliai apžvelgus sekų analizės posistemių sutvarkytus ar sukurtus duomenų rinkinius ir duomenų modelius, bei jų ryšius pagal informacinės sistemos tipą, išpildome sekų analizės taikymo procesą iki vizualizacijų. Sudarytose gairėse žinome galimus vizualizavimo atitikmenis, todėl belieka tik sudėti rezultatus į ataskaitą. Norima patikrinti, ar sudarytos gairės, ekspertų išskirti prioritetai ir sukurta realizacija atskleidžia rezultatus, rodančius naudotojų naršymo patirtį. Šiam patikrinimui planuojamas eksperimentas iš realių, skirtingų tipų informacinių sistemų duomenų.

5. Eksperimentinis sekų analizės taikymo informacinių sistemų naudotojų patirties analitikai tyrimas

Šiame skyriuje aprašomas eksperimentinis tyrimas sudarytai metodikai tirti. Naudojant realių informacinių sistemų įvykių žurnalo duomenis, siekiama patikrinti ar naudojantis metodika galima sukurti įrankį, kuris palengvintų naudotojų patyrimo analizę.

5.1. Eksperimento planas

Eksperimentu siekiama patikrinti ar sudaryta sekų analizės taikymo metodika veikia ir ar analizės rezultatai gali suteikti įžvalgas naudotojų patyrimo ekspertams.

Sudarytai metodikai realizuoti bus naudojami realios įmonės „Atea“ B2B trijų informacinių sistemų įvykių žurnalo duomenys. Tiriamų informacinių sistemų tipai tai - elektroninės komercijos, sąskaitų ir sutarčių administravimo ir reprezentacinės įmonės svetainės su tinklaraščio dalimi įvykių žurnalo duomenys. Visos šios sistemos apima ne tik Lietuvos, bet ir Švedijos, Danijos, Norvegijos, Suomijos, Latvijos ir Estijos rinkas. Eksperimentui naudojami duomenys taip pat apims septynių skirtingų portalų įvykių žurnalų duomenis. Rezultatams iliustruoti naudojami vienos savaitės naršymo duomenys (laikotarpiu 2021.04.26-2021.05.02).

Eksperimentui atlikti reikia:

- Analizės variantų išskyrimo, kurie bus taikomi tiriamai informacinei sistemai (3.6 lentelė);
- Metodikos proceso diagramos, kuri atskleidžia ataskaitos kūrimo žingsnius (2.10 pav.);
- Naujų duomenų apskaičiavimų ir išvedimo tam tikra struktūra (4.3, 4.4, 4.5, 4.6 lentelės);
- Duomenų modelių sukūrimo pagal IS tipą (4.1, 4.2, 4.4 pav.)

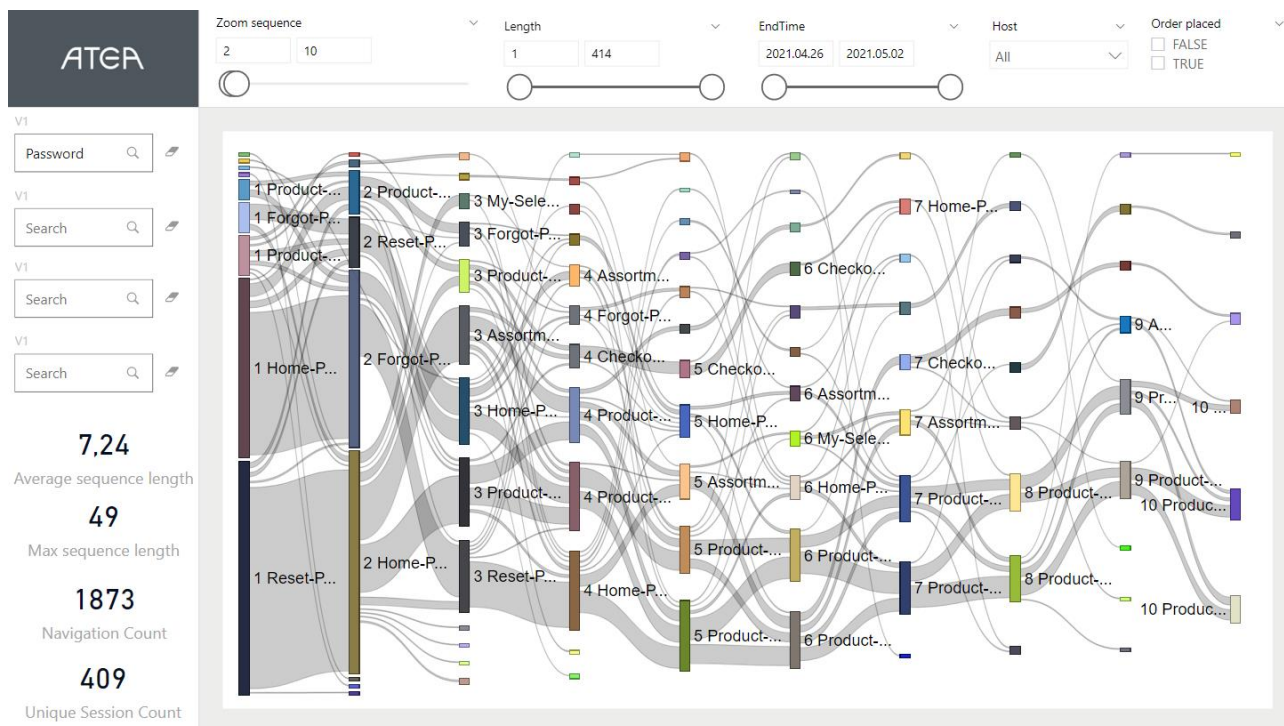
Turint paruoštus duomenis, sukurtą duomenų modelį ir žinant vizualizavimo atitikmenis, sudėliojami analizės rezultatai į ataskaitas. Siekiant suteikti lankstesnę galimybę apžvelgti rezultatus, papildomai sukuriamos filtravimo galimybės ne tik pagal analizės parametrus, bet ir pridėdant filtravimą pagal šalį ir datą. Dar papildomai suskaičiuojami keli matai, kurie parodo tiriamų sekų skaičių ar vidutinį sekos ilgį.

5.2. Sekų analizės taikymo skirtingų tipų informacinėms sistemoms gautos ataskaitos

Pagal ekspertų išsakytus analizės variantų prioritetus skirtingų informacinių sistemų tipams, imti pirmas dvi vietas užimantys analizės variantai ir pagal juos sukuriamos rezultatus apibendrinančios ataskaitos. Realizuojant elektroninės komercijos įvykių žurnalo duomenų ataskaitą, išvedamas realių sekų vizualizavimas ir nubyrežimų analizė. Realizuojant tinklaraščio informacinės sistemos įvykių žurnalo duomenų ataskaitą, išvedamos dažnų sekų taisyklės. Realizuojant sąskaitų ir sutarčių administravimo informacinės sistemos įvykių žurnalo duomenų ataskaitą, išvedamos praleisto laiko analizę atspindinčios diagramos ir šios skaidomos ir vaizduojamos pagal klasterius.

5.2.1. Elektroninės komercijos IS realių sekų vizualizavimo ir nubyrežimų ataskaita

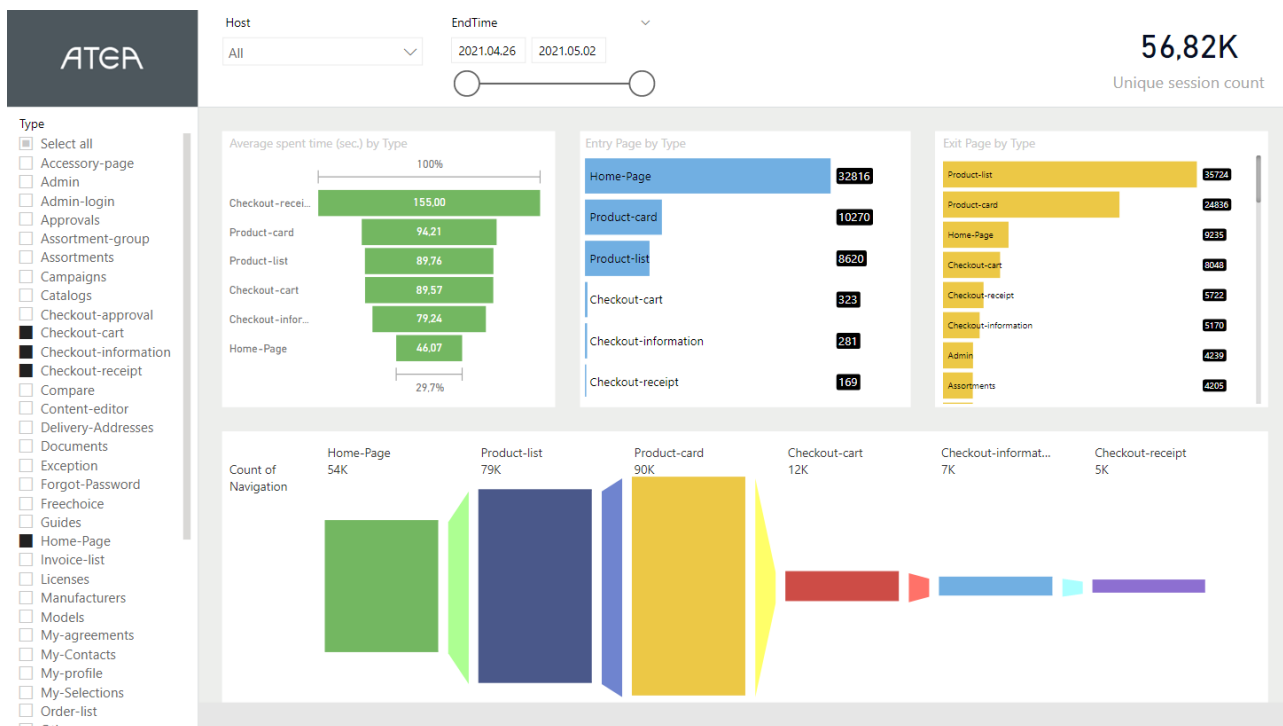
Išpildant atskaitą šio tipo informacinės sistemos duomenims, sukuriami du ataskaitos puslapiai. Pirmame, nusprendus parinkti sankio diagramą, vaizduojami realių sekų naršymo keliai pagal puslapių tipus. Sukurtos ataskaitos puslapis vaizduojamas 5.1 pav.



5.1 pav. Realių sekų vizualizavimas per sankio diagrama

Apžvelgdami šį ataskaitos puslapį, naudotojų patyrimo ekspertai gali matyti informaciją kokiuose puslapių tipuose prasideda ar baigiasi sekos, kaip jos tęsiasi ir koks yra kiekvieno tipo aplankymo dažnis. Ataskaitos vartotojui pateikiama galimybė filtruoti sekas pagal datą, šalį, sekos ilgį ar išfiltruoti tik tas sekas, kurios turi pirkimo žingsnį. Dar vartotojui leidžiama tekstu išsifiltruoti sekas, kuriose buvo aplankytas tiriamas elementas. Paveikslėlyje pateikiamas pavyzdys išfiltravus sekas pagal raktažodį „*Password*“. Prie išskirtų tipų, buvo nustatyti tipai „*Forgot-Password*“ ir „*Reset-Password*“, tad pagal naudotą raktažodį, pateikiamos sesijos turinčios vieną ar kitą tipą. Taip galima įvertinti kokia dalis naudotojų atlieka pirkimo žingsnį po slaptažodžio atstatymo žingsnio, turima informacija apie vidutinį sekos ilgį, unikalių sesijų ar naršymų skaičių.

Antrasis ataskaitos puslapis, skirtas nubyrėjimams detalizuoti, rasti kritinius puslapius ir jų dažnius, sužinoti kada naudotojas palieka sistemą. Nubyrėjimų analizei naudota piltuvėlio diagrama, kurią rekomenduojama apžvelgti išfiltravus tam tikrą seką. Ataskaitoje naudotojui bus pateikiami visi galimi aplankytų puslapių tipai, taip leidžiant pasirinkti visus tipus, kuriuos norima įtraukti į tiriamą seką. Gauta ataskaita pateikiama 5.2 pav.

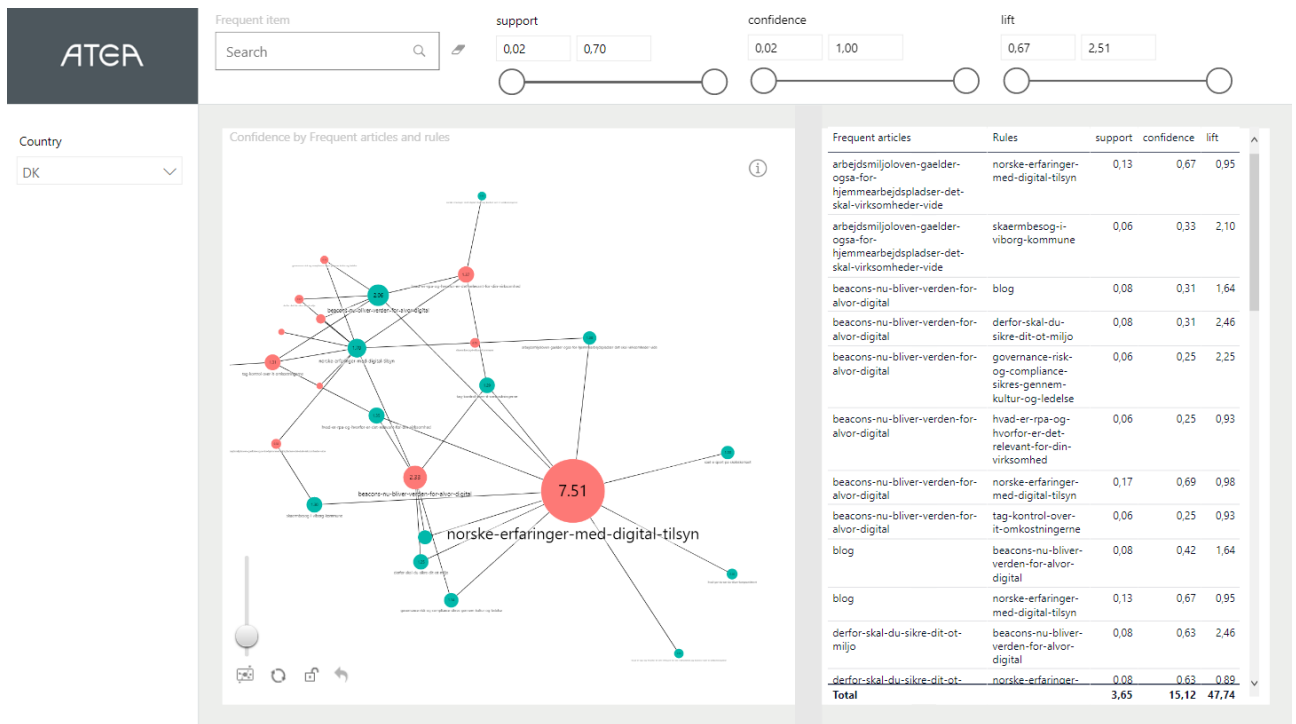


5.2 pav. Nubyrejimų analizės vizualizavimas

Pateiktoje ataskaitoje pateikiama filtravimo galimybė pagal datą ar pagal šalį, o svarbiausiai aplankytų puslapių tipai, kurių parinkimas sudedamas į piltuvėlio diagramą. Šis grafikas leidžia peržiūrėti sekas, sudarytas iš skirtingų elementų, nubyrejimų detales. Paveikslėlyje pateikiamas pavyzdys, kuriame parinkti tokie tipai: „Home-Page“, „Product-list“, „Product-card“, „Checkout-cart“, „Checkout-information“ ir „Checkout-receipt“. Atsižvelgiant į dažniausiai sutinkamas sekos elementų vietas, piltuvėlio diagramoje tipai surikiuojami atitinkama tvarka. Piltuvėlio diagramoje atspindimos visos sekos, kuriose egzistuoja nors vienas parinktas aplankytas tipas. Jei sesijos metu, tipas aplankomas kelis kartus, šis skaičius priskaičiuojamas diagramoje kartu. Iš sudarytos piltuvėlio diagramos, galima numatyti, kuriose sistemos vietose srautas išauga, o kuriose sumažėja. Pavyzdinėje sekoje tiriama tipinis elektroninės komercijos sekos kelias ir visi pirkimo žingsniai. Galime susidaryti įžvalgas apie nubyrejimų dalį kiekviename pirkimo žingsnyje iki patvirtinimo. Taip pat ataskaitoje pateikiamos lentelės iliustruojančios tipus ir skaičių tipų, kuriose prasideda ar baigiasi sekos. Pasiruošus transformuotą sekos lentelę įėjimo-išėjimo formatu, turima informacija apie praleistą laiką aplankytame tipe. Apskaičiavus vidutinį praleistą laiką, galima pateikti papildomą piltuvėlio diagramą rodančią užimamą laiką kiekviename sekos žingsnyje. Tokia sudaryta ataskaita gali padėti nustatyti ne vieną įžvalgą apie sistemos naudotojus ir jų naršymo įpročius.

5.2.2. Tinklaraščio IS dažnų sekų taisyklių ataskaita

Kita sudaryta ataskaita skirta atvaizduoti rastų dažnų sekų ir taisyklių rezultatus. Šiems rezultatams išgauti buvo naudoti interneto tinklaraščio sistemos duomenys. Pagal išskirtą prioritetą, gautos lankymo taisyklės yra svarbiausias kriterijus šio tipo informacinėms sistemoms. Apsvarsčius variantus, numatyta taikyti Apriori algoritmą. Siekiama išskirti, kokius straipsnius naudotojai skaito kartu, neatsižvelgiant kokia tvarka jie yra skaitomi. Toliau 5.3 pav. vaizduojamas sukurtos ataskaitos langas.

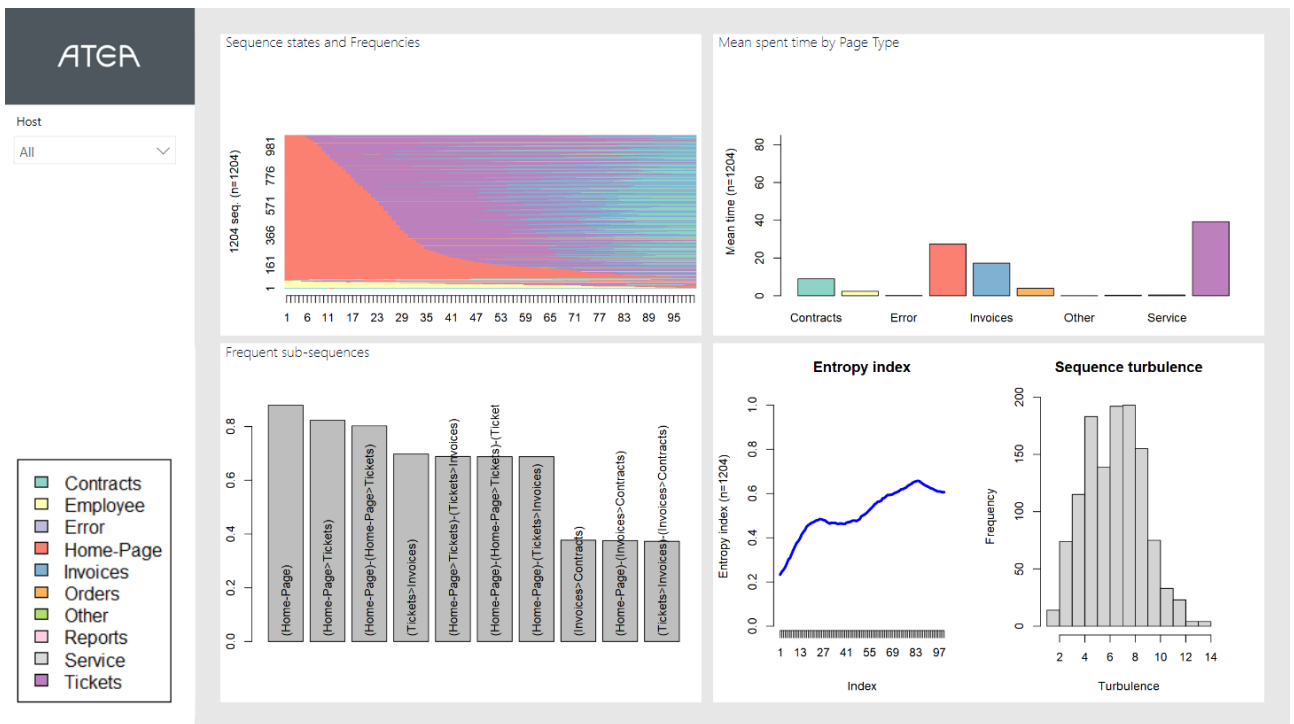


5.3 pav. Gautų dažnų sekų ir taisyklių rezultatų vizualizavimas

Pateiktame dažnų sekų ir taisyklių ataskaitos lange pagrindiniai ataskaitos rezultatai vartotojui pateikiami tinklo grafiku ir taisyklių lentele, kuriuose parodytos visos išgautos sekos. Čia vartotojui leidžiama filtruoti pagal šalį, kuriam priklauso tinklaraštis, sekos elementą, dažnio, patikimumo, svarbos parametrus. Šios sekos gali būti naudojamos rekomendaciniams patarimams. Pavyzdinėje iliustracijoje apžvelgiamos Danijos tinklaraščio sistemos aplankytos sekos. Tinklo diagramoje aiškiai parodomas sąsajos tarp skirtingų tinklaraščio straipsnių. Žalia spalva išskiriamos dažnų aplankytų straipsnių elementai, o raudona spalva rodomi galimi taisyklių elementai pagal patikimumo matą. Tokioje diagramoje aiškiai pateikiami ryšiai, jų stiprumas tarp aplankomų straipsnių. Atskirai pateikiama taisyklių lentelė, kurioje pateikiama informacija apie dažnus aplankomus straipsnius, gautą taisyklę, sekos dažnį, taisyklės patikimumą ir svarbą. Pavyzdžiui, interpretuojama vieną iš gautų taisyklių, kuri atrodo: „*blog*“ => „*norske-erfaringer-med-digital-tilsyn*““. Gautas dažnumas 0,13 parodo, jog 13% turimų transakcijų turi aplankomą „*blog*“ tipą. Patikimumas 0,67 parodo tikimybę, jog aplankius tipą „*blog*“ 67% taip pat aplankys ir „*norske-erfaringer-med-digital-tilsyn*““. Gautas svarbos matas 0,97 parodo ar gauta taisyklė yra įdomi. Žinoma, jog gavus svarbos matą arti vieneto galime numatyti, jog taisyklė nėra įdomi. Šios ataskaitos gautų rezultatų interpretavimas, naudotojų patyrimo ekspertams padėtų numatyti kartu skaitomų straipsnių atvejus.

5.2.3. Sąskaitų ir sutarčių administravimo IS praleisto laiko ir klasterizavimo ataskaita

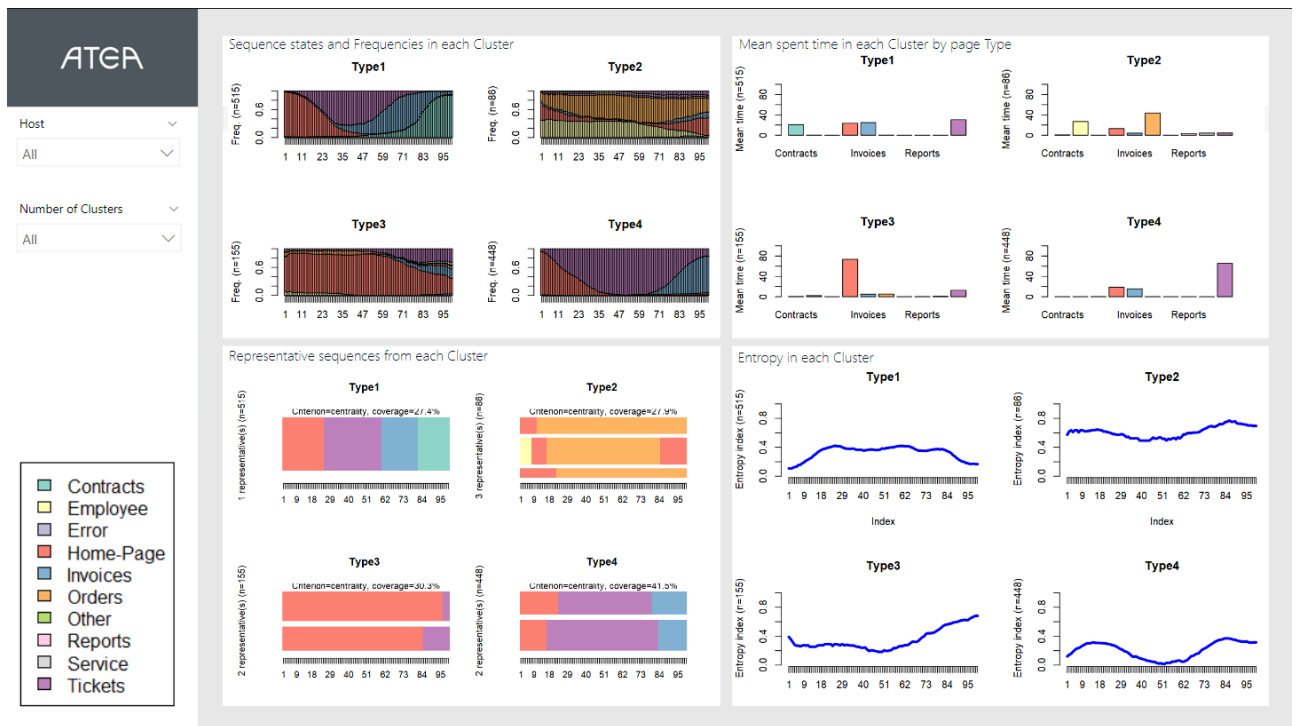
Paskutinė realizuota ataskaita sudaryta sąskaitų iš sutarčių administravimo informacinės sistemos duomenims. Ataskaita sudaryta iš dviejų puslapių. Viename puslapyje pateikiamas bendra visų sekų analizė pagal praleistą laiką puslapyje, o kitame pateikiama praleisto laiko analizė išskiriama pagal klasterius – tai yra kiekvienam klasteriui. Toliau 5.4 pav. pateikiamas pirmas ataskaitos puslapis su bendra praleisto laiko apžvalga sekose.



5.4 pav. Praleisto laiko analizės vizualizavimas

Šiame ataskaitos puslapyje pirmoje diagramoje pateikiamas visų sekų grafikas. Viso turimos 1204 sekos. Y-ašyje nurodomi sekos numeriai, o x-ašyje nurodomas praleisto laiko procentas. Skirtingomis spalvomis parodomi aplankytų puslapių tipai. Apžvelgus tik šį grafiką, galima matyti perėjimus tarp sekos elementų, praleistos trukmės dalį ar aplankytų tipų apimtį. Papildomai apskaičiuojamas ir atvaizduojamas vidutinis praleistas laikas pagal puslapio tipą. Pavyzdžiui, galima matyti, kad sekose, kuriose yra aplankomas „Home-page“, vidutiniškai praleidžiama pusė laiko. Dar atsižvelgus į praleistą laiką ir sekos išdėstymo tvarką, randamos dažnos sekos. Jos vaizduoja dažniausius elementų pasikeitimus. Pavyzdžiui, dažnas rinkinys „(Home-page)-(Tickets>Invoices)“ gali būti interpretuojamas taip – sekoje aplankius tipą „Home-page“ vėliau bus aplankoma „Tickets“ ir iš kart po jo „Invoice“ tipai. Dar, ataskaitoje vaizduojamas gautas praleisto laiko analizės entropijos indeksas ir turbulencija. Turbulencija parodo skaičių skirtingų sub-sekų praleistų skirtinguose įvykiuose, o entropija parodo įvykių įvairovę.

Sekančiame ataskaitos puslapyje vaizduojama praleisto laiko analizės padalinus sekas į klasterius. 5.5 paveikslėlyje pateikiamas pavyzdinis ataskaitos puslapis.



5.5 pav. Praleisto laiko analizės rezultatai suskirstyti į klasterius

Naudojant hierarchinį klasterizavimo metodą, pirmiausiai buvo parenkamas tinkamas klasterių skaičius. Sudarant ataskaitą išskirtas klasterių skaičius pagal tuo metu turimus duomenis. Tinkamam klasterių skaičiui nustatyti, pirmiausiai buvo išvedama dendograma, kuri leido įvertinti ir parinkti optimaliausią skaičių. Įvertinus, kad duomenys ataskaitoje atsinaujina automatiškai ir ataskaitos metu nustatytas klasterių skaičius gali netikti, ataskaitoje pateikiamas filtras su galimais klasterių skaičiais (nuo 2 iki 10), o parinkus norimą skaičių, ataskaitos elementai atnaujinami atitinkama pagal naujai parinktą skaičių. Rekomenduojama gautus klasterius profiliuoti ir šiems suteikti prasmingus pavadinimus, tačiau įvertinus galimą klasterių skaičiaus keitimą, pavadinimai nėra suteikiami. Išgavus 4 klasterius, vaizduojama kiekvieno klasterio praleisto laiko ir apimties diagrama, pagal sekos dažnį. Visos sekos padalintos taip, kad pirmam klasteriui pakliuvo 515 sekų, antram klasteriui – 86 sekos, tračiam klasteriui 155 sekos, o ketvirtam – 448 sekos. Atvaizduojami tipai ir vidutinis praleistas laikas kiekviename pagal klasterį. Taip pat, išskiriamos kiekvieną klasterį reprezentuojančios sekos. Dar, apskaičiuojama entropija kiekvieno klasterio atveju.

Sudarius praleisto laiko analizės ataskaitą, naudotojų patyrimo ekspertai gali ne tik įsivertinti, kokią laiko dalį vartotojai praleidžia atitinkame puslapyje, bet ir tai, kokie elementai yra sekoje ir koks yra jų išdėstymas sekoje. Pasitelkus klasterizavimą, galima nustatyti klasterius, taip skirstant naudotojus pagal naršymo šablonus. Tai parodo, kokia naudotojų dalis atlieka vienus žingsnius sistemoje, o kokia dalis visai kitokius.

5.3. Sprendimo taikymo rekomendacijos

Sprendimo taikymo metu nustatyta, jog norimoms vizualizacijoms išpildyti reikalingas papildomas duomenų rinkinių transformavimas ir saugojimas. Tai papildomai užima duomenų saugyklų vietą, tačiau suteikia galimybę po duomenų atnaujinimo turėti analizės rezultatus perteiktus per skirtingas vizualizacijas. Taip pat paliekama plati duomenų filtravimo galimybė, kuri leidžia filtruoti analizės

rezultatus įvairiais pjūviais, ar net parinkti analizės parametrus, pavyzdžiui klasterizavimo ar dažnų sekų rasime – dažnumo, patikimumo ar svarbos parametrus.

Eksperimentų išskirti prioritetai leido sumažinti analizės kriterijų taikymą skirtingoms informacinėms sistemoms, todėl pritaikyta tik aktualiausia analizė skirtingai informacinei sistemai. Galutinis ataskaitos vartotojas gavęs ataskaitą turi būti supažindinamas su skirtingais parametrais, kurie numatyti kaip filtrai. Atsižvelgiant į tai, jog duomenys ataskaitoje atsinaujins automatiškai, iš anksto nėra parenkami parametrai ir filtrai. Iš anksto parinkti parametrai gali netikti ateityje sukauptų duomenų rinkiniams. Kuriant ataskaitą, rekomenduojama parinkti aiškiai įskaitomas vizualizacijas, stengiantis prasmingai pavadinti ataskaitos elementus ir filtrus. Tai palengvintų naudotojų patyrimo ekspertų ataskaitos naudojimą. Sukurta ataskaita negali automatiškai pateikti įžvalgų, tai yra tik duomenų analizės įrankis, kuriuo naudojantis specialistas suprantantis sistemos dalykinę sritį gali numatyti optimizavimui skirtus veiksmus.

Išvados

1. Darbe sprendžiama sistemose aplankytų puslapių sekų duomenų tyrimo problema. Esami aplankytų puslapių įrašai saugomi įvykių žurnale ir šie duomenys organizacijose paprastai yra nepakankamai analizuojami. Naudotojų patyrimo ekspertai norintys apžvelgti naršymo kelius, gali gauti pavienių sesijų rezultatus ar atskirų puslapių aplankymo statistikas, bet ne apibendrintas išvalgas.
2. Darbe buvo išanalizuoti ir palyginti įrankiai padedantys atlikti sekų. Nuspręsta, kad atvirojo kodo R paketai „TramineR“, „arulesSequences“ ir „Clickstream“ yra tinkamesni sprendimui įgyvendinti labiau, nei komerciniai paketai, dėl lankstumo ir analizės variantų įvairovės, o analizės rezultatų vizualizavimui pasirinktas veiklos analitikos įrankis „Power BI“.
3. Sudaryta sekų analizės taikymo metodika - gairės išskyrė pagrindinius žingsnius duomenų praruošimui, galimus analizės variantus įvykių žurnalo sekoms tirti, bei kiekvienam variantui priskyrė tinkamus vizualizavimo būdus.
4. Pasitelkus ekspertų vertinimą buvo išgauti analizės taikymo prioritetai skirtingoms informacinėms sistemoms. Nustatyta, jog elektroninės komercijos sistemoms rekomenduojama atlikti nubyrežimų analizę ir realių sekų vizualizavimą. Interneto aplikacijų informacinėms sistemoms ir socialinių tinklų, forumų informacinėms sistemoms atlikti praleisto laiko tyrimą ir klasterizavimą. Reprezentacinės veiklos sistemoms atlikti nubyrežimų analizę ir dažnų sekų taisyklių radimą, o naujienų portalų ir interneto tinklaraščio IS rasti dažnų sekų kombinacijas ir dažnų sekų taisykles.
5. Pasiūlytos gairės buvo išmėgintos realių, skirtingų informacinių sistemų tipų, įvykių žurnalo duomenų analitikai. Tyrimo metu buvo realizuoti prototipiniai įrankiai šiems informacinių sistemų tipams: elektroninės komercijos, sąskaitų ir užsakymų administravimų, bei interneto tinklaraščio informacinėms sistemoms. Rezultate sekų analizę gairių dėka kiekvienam tipui pavyko susiaurinti iki esminių aspektų.
6. Sukurtos metodikos tyrimas parodė, jog realizuotas sekų analizė įrankis skirtingų tipų informacinėms sistemoms atskleidžia naudotojų naršymo įpročius ir gali pagelbėti naudotojų patyrimo specialistams, suprantantiems informacinės sistemos dalykinę sritį.

Literatūros sąrašas

1. C. Gentile, N. Spiller ir G. Noci, „How to Sustain the Customer Experience:: An Overview of Experience Components that Co-create Value With the Customer,“ *European Management Journal*, t. 25, nr. 5, pp. 395-410, 2007.
2. J. Ross, „The Business Value of User Experience,“ 2 Commerce Drive, Cranbury, NJ, 2014.
3. K. Beauchaud ir H. Kroemker, „A framework to measure User eXperience of interactive online products,“ *ACM International Conference Proceeding Series*, 2010.
4. B. Coker, „Antecedents to website satisfaction, loyalty, and word-of-mouth,“ *JISTEM - Journal of Information Systems and Technology Management*, t. 10, nr. 2, pp. 209-218, 2013.
5. P. Dixit, „This UX measurement tool is all you need!,“ UX Planet, 28 Kovo 2018. [Tinkle]. Available: <https://uxplanet.org/this-ux-measurement-tool-is-all-you-need-87f1b35896d4>. [Kreiptasi 03 Balandžio 2021].
6. Your Europe, „Privatumas internete,“ [Tinkle]. Available: https://europa.eu/youreurope/business/dealing-with-customers/data-protection/online-privacy/index_lt.htm. [Kreiptasi 14 Gruodžio 2020].
7. A. Dwivedi, „Analysis of various link algorithms in web mining using link,“ 2012.
8. P.G. ir O. Prakash, „Predicting the user behavior from weblogs by improved span classification,“ *International Journal of Pure and Applied Mathematics*, pp. 4187-4199, 2018.
9. P. Weichbroth, M. Owoc ir M. Pleszkun, „Web User Navigation Patterns Discovery from WWW Server Log Files,“ įtraukta *Federated Conference on Computer Science and Information Systems*, Poland, 2012.
10. O. R. Zaiane, „Web Usage Mining for a Better Web-Based Learning Environment,“ 2001.
11. „Web Heatmap Journal – Best Website Heatmap Tool You Should Be Using,“ Mouseflow, [Tinkle]. Available: <https://akite3d.com/web-heatmap-journal-best-website-heatmap-tool-you-should-be-using/>. [Kreiptasi 22 11 2019].
12. H. Li, L. Yu ir W. He, „The Impact of GDPR on Global Technology Development,“ *Journal of Global Information Technology Management*, t. 22, nr. 1, pp. 1-6, 2019.
13. R. Shettar, „Sequential pattern mining from web log data,“ *International journal of engineering science & advanced technology*, t. 2, nr. 2, pp. 204-208, 2012.
14. O. Müller, I. Junglas, J. v. Brocke ir S. Debortoli, „Utilizing big data analytics for information systems research: challenges, promises and guidelines,“ *European Journal of Information Systems*, nr. 25, pp. 289-302, 2016.
15. Colombo Web Solutions, „Different types of Websites,“ Colombo Web Solutions, 13 Vasario 2016. [Tinkle]. Available: <http://colombowebs.com/different-types-of-websites/>. [Kreiptasi 14 Balandžio 2021].
16. S. A. Ehikioya, „A Path Analysis Model for Effective E-Commerce Transactions,“ pp. 55-71, 06 2019.

17. A. Abbott, „Event Sequence Analysis: Methods,“ įtraukta *International Encyclopedia of the Social & Behavioral Sciences*, Illinois, USA, University of Chicago, 2001, pp. 4966-4969.
18. S. F. ir T. H., „The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates,“ *Biochemical Journal*, pp. 463-481, 1951.
19. „Sequence alignment,“ [Tinkle]. Available: https://en.wikipedia.org/wiki/Sequence_alignment. [Kreiptasi 22 lapkričio 2019].
20. G. Ritschard, A. Gabadinho, M. Studer ir N. S. Muller, „Converting between various sequence,“ įtraukta *Advances in Data Management*, Switzerland, Department of Econometrics and Laboratory of Demography, 2009, pp. 155-175.
21. A. Abbot, „A Primer on Sequence Methods,“ t. 1, nr. 4, pp. 339-431, 1990.
22. H. M. Huynh, L. T. T. Nguyen, B. Vo, Z. K. Oplatkova ir T.-P. Hong, „Mining Clickstream Patterns Using IDLists,“ įtraukta *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bary, Italy, 2019.
23. University of Geneva, „TraMineR Sequence analysis in R,“ [Tinkle]. Available: <http://traminer.unige.ch/index.shtml>. [Kreiptasi 22 lapkričio 2019].
24. G. Wang, X. Zhang, S. Tang, H. Zheng ir B. Y. Zhao, „Unsupervised Clickstream Clustering for User Behavior Analysis,“ 2016, pp. 225-236.
25. A. Gabadinho, G. Ritschard, N. S. Muller ir M. Studer, „Analyzing and Visualizing State Sequences in R with,“ *Journal of Statistical Software*, t. 4, nr. 40, pp. 1-37, 2011.
26. J. Zhao, Z. Liu, M. Dontcheva, A. Hertzmann ir A. Wilson, „MatrixWave: Visual Comparison of Event Sequence Data,“ *Conference on Human Factors in Computing Systems*, pp. 259-268, 2015.
27. R. Helms, „Visualizing the sales funnel: A simple template,“ 20 rugsėjo 2016. [Tinkle]. Available: <https://medium.com/@randall.helms/visualizing-the-sales-funnel-a-simple-template-4dfb0ff4400>. [Kreiptasi 23 gruodžio 2019].
28. „CHORD DIAGRAM,“ Data to Viz, [Tinkle]. Available: <https://www.data-to-viz.com/graph/chord.html>. [Kreiptasi 18 12 2020].
29. C. Buchta, „arulesSequences,“ R, [Tinkle]. Available: <https://www.rdocumentation.org/packages/arulesSequences/versions/0.2-25>. [Kreiptasi 26 Vasario 2020].
30. M. J. Zaki, „SPADE: An Efficient Algorithm for Mining,“ *Machine Learning*, nr. 42, pp. 34-60, 2001.
31. M. Scholz ir T. v. Kraay, „clickstream: Analyzes Clickstreams Based on Markov Chains,“ CRAN, 30 01 2020. [Tinkle]. Available: <https://cran.r-project.org/web/packages/clickstream/index.html>. [Kreiptasi 21 01 2021].
32. Microsoft, „Microsoft Sequence Clustering Algorithm,“ 05 rugpjūčio 2018. [Tinkle]. Available: <https://docs.microsoft.com/en-us/analysis-services/data-mining/microsoft-sequence-clustering-algorithm>. [Kreiptasi 22 lapkričio 2019].

33. University of Maryland, „EventFlow: Visual Analysis of Temporal Event Sequences and Advanced Strategies for Healthcare Discovery,“ [Tinkle]. Available: <https://hcil.umd.edu/eventflow/>. [Kreiptasi 12 12 2019].
34. B. Halpin, „SADI: Sequence Analysis Tools for Stata,“ *The Stata Journal*, pp. 1-25, 2017.
35. N. Bakar ir S. A. Abu, „Using regular expressions for mining data in large software repositories,“ įtraukta *2014 the 5th International Conference on Information and Communication Technology for the Muslim World*, Malaysia, 2014.
36. Q. M, „Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data,“ įtraukta *International conference and Workshop on High Dimentional Data Analysis*, Indonesia, 2021.
37. R Bloggers, „Sequence of shopping carts in-depth analysis with R – Sequence of events,“ 28 Sausio 2015. [Tinkle]. Available: <https://www.r-bloggers.com/2015/01/sequence-of-shopping-carts-in-depth-analysis-with-r-sequence-of-events/>. [Kreiptasi 17 Sausio 2021].
38. T.-P. Hong, C.-Y. Wang ir Y.-H. Tao, „A new incremental data mining algorithm using pre-large itemsets,“ *Intelligent Data Analysis · March 2001*, 2001.
39. J. Hipp, „CRISP-DM: Towards a standard process model for data mining,“ *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2000.
40. D. Dragan ir D. Topolšek, „Introduction to Structural Equation Modeling: Review, Methodology and Practical Applications,“ įtraukta *The International Conference on Logistics & Sustainable Transport 2014*, Slovenia, 2014.
41. T. L. Saaty, „The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation,“ 1980.
42. I. Vinogradova, „Neapibrėžtumo įtaka AHP metodo vertinimams,“ *Lietuvos matematikų draugijos darbai*, t. B, nr. 53, pp. 243-248, 2012.
43. K. D. Goepel, „AHP Priority Calculator,“ BPMSG, 19 Rugsėjo 2019. [Tinkle]. Available: <https://bpmsg.com/ahp/ahp-calc.php?n=6&c%5B0%5D=Spent+time+analysis&c%5B1%5D=Sequence+elements+%28combination%29&c%5B2%5D=Rule+element+%28Next+possible+element%29&c%5B3%5D=Bottleneck+analysis&c%5B4%5D=Vizualizations&c%5B5%5D=Clustering>. [Kreiptasi 12 Sausio 2021].
44. „SQL machine learning documentation,“ Microsoft, [Tinkle]. Available: <https://docs.microsoft.com/en-us/sql/machine-learning/?view=sql-server-ver15>. [Kreiptasi 20 Birželio 2020].
45. „SQL Server 2017 – Intelligent Database that runs Anywhere,“ Microsoft, [Tinkle]. Available: <https://blogs.partner.microsoft.com/mpn-canada/sql-server-2017-intelligent-database-runs-anywhere/>. [Kreiptasi 20 Birželio 2020].

5/2/2021

Survey about Application of Sequence Analysis for Information Systems' User Experience Analytics

When comparing two criteria items, consider evaluation 5 as equal.

1. Please indicate the information system type that you would like to have analyzed *

Mark only one oval.

- E-commerce system
- Business website
- News & Blog Websites
- Social media sites & Forums
- Other type application system (E.g. Banking, Web apps)

2. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Spent time analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sequence elements (combination)

3. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Spent time analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Rule element (Next possible element)

4. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Spent time analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Bottleneck analysis

5. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Spent time analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Vizualizations

6. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Spent time analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clustering

7. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Sequence elements (combination)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Rule element (Next possible element)

8. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Sequence elements (combination)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Bottleneck analysis

9. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Sequence elements (combination)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Vizualizations

10. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Sequence elements (combination)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clustering

11. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Rule element (Next possible element)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Bottleneck analysis

12. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Rule element (Next possible element)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Vizualizations

13. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Rule element (Next possible element)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clustering

14. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Bottleneck analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Vizualizations

15. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Bottleneck analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clustering

16. Rank which analysis is more important for the system that you are planning to analyze (5 means equal) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	
Vizualizations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Clustering

This content is neither created nor endorsed by Google.



Elektroninės komercijos puslapių šablonai ir tipai

Šablonas	Tipas
regex("/eshop/\\z")	Home-Page
regex("/eshop/m4n\\z")	Home-Page
regex("/eshop/m4n/\\z")	Home-Page
"eshop/products"	Product-list
"eshop/product"	Product-card
"eshop/models"	Models
"eshop/manufacturers"	Manufacturers
"eshop/admin"	Admin
"assortments"	Assortments
"eshop/assortment-group"	Assortment-group
"eshop/tag/"	Tags
"eshop/freechoice"	Freechoice
"eshop/myselection"	My-Selections
"eshop/campaigns"	Campaigns
"eshop/compare"	Compare
"eshop/packages"	Packages
"checkout/cart"	Checkout-cart
"checkout/information"	Checkout-information
"checkout/receipt"	Checkout-receipt
"checkout/approval"	Checkout-approval
"eshop/accessories/"	Accessory-page
"eshop/documents/"	Documents
"eshop/mycontacts"	My-Contacts
"/orderinformation/order_list"	Order-list
"/orderinformation/tracking"	Tracking
"/orderinformation/serial"	Serial-number
"/orderinformation/invoice"	Invoice-list
"/cache/pdf"	PDF
"eshop/licens"	Licenses
"eshop/spla"	Licenses
"eshop/my-licens"	Licenses
"eshop/user"	Users
"eshop/return"	Returns
"eshop/quote"	Quotes
"eshop/switch-customer"	Switch-Customer
"approvalslist"	Approvals
"eshop/deliveryaddresses"	Delivery-Addresses

"eshop/parked-orders"	Parked-orders
"eshop/guides-and-tools"	Guides
"web/P01.m4n"	Admin-login
"reset-password"	Reset-Password
"eshop/my-profile"	My-profile
"eshop/catalog"	Catalogs
"/orderreport/"	Reports
"ForgotPassword"	Forgot-Password
"eshop/product-configurator"	Product-configurator
regex("eshop/\\w+content")	E-Content-editor
"eshop/content"	Content-editor
"innovaeditor"	Content-editor
"/responsiveprofiles/"	Content-editor
"eshop/my-agreements"	My-agreements
"eshop/exception"	Exception

Sąskaitų administravimo puslapių šablonai ir tipai.

Šablonas	Tipas
regex("my.atea.\\w+\\z")	Home-Page
"/Orders"	Orders
"/orders"	Orders
"/Tickets"	Tickets
"/Invoices"	Invoices
"/Employee/Admin"	Employee
"/Contracts"	Contracts
"/Dashboard"	Home-Page
"/dashboard"	Home-Page
"/Reports"	Reports
"/service-request/"	Service
"/selfservices/"	Service
"/Error/"	Error