

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ BAKALAURO STUDIJŲ PROGRAMA

Kodo skirstymo į paketus šablonų tyrimas

Analysis of code packaging patterns

Bakalauro baigiamasis darbas

Atliko:	Martyna Ubartaitė
Darbo vadovas:	Gediminas Rimša
Darbo recenzentas:	doc. dr. Audronė Lupeikienė

Vilnius – 2024

Santrauka

Glaustai aprašomas darbo turinys: pristatoma nagrinėta problema ir padarytos išvados. Santraukos apimtis ne didesnė nei 0,5 puslapio. Santraukų gale nurodomi darbo raktiniai žodžiai. Automatiškai naudojamos lietuviškos kabutės: „tekstas“.

Raktiniai žodžiai: raktinis žodis 1, raktinis žodis 2, raktinis žodis 3, raktinis žodis 4, raktinis žodis 5

Summary

Santrauka anglų kalba. Santraukos apimtis ne didesnė nei 0,5 puslapio. Automatiškai naudojamos angliškos kabutės: “tekstas”.

Keywords: keyword 1, keyword 2, keyword 3, keyword 4, keyword 5

Turinys

ĮVADAS	5
1. KOMPIUTERINĖS SISTEMOS VERTINIMAS	7
1.1. Kompiuterinės sistemos kokybė.....	7
1.2. Kodo skirstymo į paketus šablonų vertinimas	7
1.2.1. Bendro sąryšio principas.....	8
1.2.2. Aciklinių priklausomybių principas	8
1.2.3. Stabilių priklausomybių principas.....	8
1.2.4. Stabilių abstrakcijų principas	9
1.3. Paketų kokybės matai	9
2. GALIMI KODO SKIRSTYMO Į PAKETUS ŠABLONAI	11
2.1. Sistemų analizės procesas	11
2.2. Problemos	11
2.2.1. Pagalbinių, daugkartinio naudojimo klasių skirstymas	11
2.2.2. Didelis klasių skaičius pakete	14
2.2.3. Daug skirtingų sąsajų implementacijų	16
2.2.4. Esybių pokyčių ir versijavimo valdymas	18
3. ĮRANKIAI ŠABLONŲ ANALIZEI IR ĮVERTINIMUI	20
3.1. Reikalavimai įrankiams.....	20
3.2. Reikalavimai bendrinės sistemos analizės įrankiui	20
3.3. Reikalavimai įrankiui paketo kokybei skaičiuoti.....	20
3.4. Įrankių įgyvendinimas.....	21
4. ESAMŲ SISTEMŲ PERTVARKYMAS	22
4.1. Sistemų pasirinkimas	22
4.2. <i>crawler4j</i> sistema	22
4.3. <i>Crawler</i> sistema	28
REZULTATAI	30
IŠVADOS	31
ŠALTINIAI	32

Įvadas

Gera suprojektuota kompiuterinė sistema yra vienas iš kritinių sėkmingo verslo elementų. Tam, jog informacinės sistemas naudojantis verslas išlaikytų stabilų augimą, yra būtina sukurti sistemą, kuri sumažintų atotrūkį tarp organizacijos tikslų ir jų įgyvendinimo galimybių. Mąstant apie programinio kodo projektavimą, kodo paketų kūrimas, klasių priskyrimas jiems ir paketų hierarchijos sudarymas paprastai nėra pagrindinis prioritetas, tačiau tai parodo praleistą galimybę padaryti sistemos struktūrą labiau patikima [Sho19], suprantama [Eli10] ir lengviau palaikoma. Modernios informacinės sistemos yra didžiulės, programinis kodas yra padalintas į daugybę failų, kurie išskaidyti per skirtingo gylio direktorijas, todėl apgalvotai išskirstytas programinis kodas daro daug didesnę įtaką bendram sistemos našumui, suprantamumui bei palaikomumui, nei gali atrodyti iš pirmo žvilgsnio. Sistemos paketų struktūros analizė norint įvertinti programinės įrangos kokybę tampa vis svarbesne tema dėl augančio failų ir paketų skaičiaus [Eli10].

Diskusijose, kaip reikėtų skirstyti programinį kodą, paprastai akcentuojami du metodai – pagal *techninį sluoksnį*, kur kiekvienam funkcionalumui arba kompiuterinės sistemos sluoksniui yra sukuriamas paketas, arba pagal *dalykinės srities esybes*, kur vienos esybės kodas, dalykinės srities esybės funkcionalumas skirtingose programiniuose sluoksniuose yra patalpintas viename pake-
te [Jan21]. Dažniausiai, prioritetas teikiamas kodo skirstymui pagal dalykinės srities esybes [Eva03] dėl lengvesnio skaitomumo bei palaikomumo [Sad21]. Tačiau šis metodas negali būti vienareikšmiškai pritaikomas kiekvienoje situacijoje ir, dažniausiai praktikoje jo nėra griežtai laikomasi – klasės būna išskaidytos remiantis papildomomis taisyklėmis, siekiant išspręsti sistemos planavimo metu kylančias problemas. Norint išsiaiškinti, kaip programinis kodas gali būti skirstomas efektyviausiai, tam jog jo struktūra darytų teigiamą įtaką sistemai, reikalinga atlikti skirstymo į paketus šablonų analizę – išsiaiškinti galimus šablonus, naudojamus kylančių projektavimo problemų sprendimui, turėti aiškius šablonų apibrėžimus su jų privalumais bei trūkumais. Šiame darbe minint *šabloną kodo skirstymui į paketus* turima omenyje taisyklių arba metodų rinkinį, nurodantį, kaip grupuoti klases į paketus, užtikrinant nuoseklų stilių bei sprendžiant iškilusią problemą.

Šio darbo tikslas – identifikuoti ir įvertinti praktikoje naudojamus šablonus kodo skirstymui į paketus.

Tikslui pasiekti yra iškeliami šie uždaviniai:

- Išskirti gerai įgyvendinto kodo požymius
- Aprašyti skirstymo į paketus šablonus, remiantis praktikoje sutinkamais pavyzdžiais
- Pasiūlyti kriterijus, įvertinančius kodo suskirstymo šablono įtaką sistemos kokybei, remiantis rastais gerai įgyvendintos sistemos požymiais
- Pasirinkti kelias sistemas ir pertvarkyti jų failų struktūrą pagal aprašytus šablonus, įvertinant, kiek sudėtinga pasiekti kiekvieno šablono struktūrą
- Naudojant pertvarkytas sistemas, įvertinti kiekvieną šabloną kodo skirstymui į paketus pagal pasiūlytus kriterijus
- Pateikti rekomendacijas, ar aprašyti šablonai kodo skirstymui tinkami naudoti

Šio darbo metu nagrinėjami ir aprašomi gerai įgyvendinto kodo požymiai, užtikrinantys

sistemos stabilumą ir palaikomumą, remiantis Martin Kleppmann *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*, ir Robert C. Martin *Agile Software Development, Principles, Patterns, and Practices* knygomis. Ieškomi kriterijai, kuriuos naudojant galima įvertinti kodo suskirstymo įtaką sistemos kokybei, pavyzdžiui – komponentų skaičius, tiesioginės ir netiesioginės priklausomybės, paketų stabilumas [Mar02]. Tyrinėjami šablonai kodo skirstymui į paketus įvardinti Martin Sadin straipsnyje *Four Strategies for Organizing Code*. Nagrinėjamos atviro kodo sistemos, pasirenkant skirtingo tipo projektus, siekiant objektyvesnės šablonų analizės skirtingose srityse. Galimi tipai:

- Taikomoji programinė įranga, teikianti paslaugas įrangos naudotojams. Pavyzdžiui, internetinė programėlė priminimams ir darbams užsirašyti
- Techninė programinė įranga, naudojama taikomosios programinės įrangos duomenų saugojimui, siuntimui, paieškai. Pavyzdžiui, duomenų bazės, pranešimų eilės, talpyklos (angl. cache)
- Programinės įrangos įrankiai, skirti naudoti kitose sistemose supaprastinant programinį kodą, naudojant jau įgyvendintas funkcijas. Pavyzdžiui, Java programavimo kalbos Spring karkasas internetinių programėlių kūrimui

Tyrinėjamų projektų paketų struktūros pertvarkomos pagal pasirinktus skirstymo šablonus, pertvarkyti projektai įvertinti, naudojant išskirtus kriterijus, nustatant, kokią įtaką skirtingi skirstymo šablonai turi sistemos kokybei.

Likusi šio dokumento dalis yra išdėstyta taip – pirmas skyrius nagrinėja tvarkingos kompiuterinės sistemos sąvoką, kas ją sudaro, įgyvendinimo kokybę ir kaip galima ją įvertinti. Aprašyti kriterijai, kaip įvertinti paketų struktūros įtaką sistemos kokybei. Antras skyrius tyrinėja skirtingus šablonus klasėms į paketus skirstyti, jų privalumus bei trūkumus. Trečiame skyriuje aprašomi sukurti įrankiai, reikalingi sistemų analizei ir šablonų įvertinimui, minima, kaip jie įgyvendinti ir kaip jie yra naudojami. Ketvirtame skyriuje analizuojamos pasirinktos atviro kodo sistemos – bandoma nustatyti jų naudojamus šablonus, vertinama sistemų kokybė. Penktame skyriuje aprašomas procesas, kaip pasirinktos sistemos yra perdaromos, kad tiksliai laiktųsi antrame skyriuje aprašytų kodo skirstymų šablonų, įvertinama, kiek sudėtinga pasiekti kiekvieno šablono struktūrą. Nagrinėjama perdarytų sistemų kokybė pagal pirmame skyriuje aprašytus kriterijus, vertinami šablonų privalumai bei trūkumai.

1. Kompiuterinės sistemos vertinimas

1.1. Kompiuterinės sistemos kokybė

Norint išsiaiškinti, kokią įtaką sistemos kokybei daro skirtingi paketų skirstymo šablonai ir kaip objektyviai pamatuoti jų įtaką, pirmiausia reikėtų apsibrėžti, kokiais požymiais pasižymi gerai įgyvendinta kompiuterinė sistema. Martin Kleppmann savo knygoje *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems* [Kle17] išskiria šiuos pagrindinius kriterijus:

- Patikimumas, reiškiantis, kad net ir klaidų (įrangos, programinių ar žmogiškųjų) atveju, sistema veikia stabiliai ir patikimai, paslepiant tam tikras klaidas nuo vartotojo [Kle17].
- Prižiūrimumas, reiškiantis, jog skirtingų abstrakcijų pagalba sumažintas sistemos kompleksiskumas. Dėl to nesunku keisti esamą sistemos funkcionalumą bei pritaikyti naujiems verslo naudojimo atvejams. Tai supaprastina darbą inžinierių ir operacijų komandoms, dirbančioms su šia sistema, taip pat leidžia prie sistemos prisidėti naujiems žmonėms, o ne tik jos ekspertams. Tai ypač aktualu atviro kodo sistemoms [Kle17].
- Plečiamumas, reiškiantis, jog sistema turi strategijas, kaip išlaikyti gerą našumą užklausų srautui didėjant ir sistemai augant, tai atliekant su pagrįstais kompiuteriniais resursais ir priežiūros kaina [Kle17].

Yra daug skirtingų elementų, sudarančių sistemą, kuri tenkintų aukščiau paminėtus kriterijus, pavyzdžiui, pasirinktos technologijos, aukšto lygio architektūra, dokumentacija, sistemos testavimo procesai, jų kiekis ir pan. Vienas iš svarbių elementų, prisidedančių prie gerai įgyvendintos sistemos yra programinio kodo struktūra, jo skaitomumas, patikimumas. Konvencijos, kaip vadinti kodo paketus, kokias klases jiems priskirti ir kokios paketų hierarchijos laikytis sudaro svarbią programinio kodo struktūros dalį. Todėl, kuriant programinę įrangą, laikas, skirtas rasti sistemai tinkamus paketų skirstymo šablonus ir tų šablonų laikytis, atsiperka padarant programinę kodą geriau suprantamu, taip prisidedant prie bendro sistemos patikimumo, plečiamumo ir lengvesnio palaikomumo.

Straipsnyje *Investigating The Effect of Software Packaging on Modular Structure Stability* [Sho19], autoriai akcentuoja, kad gerai įgyvendintos objektinio stiliaus sistemos turėtų vystytis be didelių pakeitimų jų architektūroje. To siekiama, nes architektūriniai pakeitimai paveikia didelę sistemos dalį ir jų įgyvendinimo ir sistemos priežiūros kaštai yra žymiai didesni [Sho19]. Paketų struktūra, kuri užtikrina atsietą (angl. *decoupled*) komunikavimą tarp paketų TODO performuoti, enkapsuliuoja paketų vidinius elementus, neleidžiant pakeitimais išplisti už paketų ribų yra pagrindas tvirtai sistemos architektūrai, gebančiai efektyviai plėstis, nepristatant nebūtinų pokyčių ir sutaupant programos priežiūros kaštus.

1.2. Kodo skirstymo į paketus šablonų vertinimas

Ankstesniame skyriuje buvo nagrinėjama gerai įgyvendintos paketų struktūros įtaka gerai kompiuterinės sistemos struktūrai, tačiau lieka neatsakytas klausimas – kaip įvertinti šabloną

kodui į paketus grupuoti, kaip objektyviai užtikrinti, jog pasirinktas sprendimas yra būtent toks, kokio reikia, ir kokia jo įtaka kompiuterinei sistemai? Tvarkingas, aiškiai suprantamas kodas yra subjektyvi tema, priklausanti nuo komandos, naudojamos programavimo kalbos ar programinių įrankių bei programinės sistemos dalykinės srities. Kodo grupavimo į paketus metodai, taip, kaip ir bendros tvarkingo kodo praktikos, gali būti labai subjektyvūs ir patogūs tik metodą formavusiam asmeniui. Tam, kad būtų galima pagrįstai įvertinti skirtingus kodo skirstymo šablonus, pasiekiant kuo objektyvesnį, įtaką sistemos kokybei nusakantį rezultatą, reikėtų aprašyti kriterijus, nusakančius, ko tikimasi iš paketų struktūros.

Robert C. Martin savo knygoje *Agile Software Development, Principles, Patterns, and Practices* [Mar02] aprašo principus, padedančius teisingai grupuoti klases į paketus. Rodiklis, kiek kiekvienas paketas sistemoje laikosi nurodytų principų naudojant tam tikrą paketų skirstymo šabloną, gali būti kriterijus vertinant to šablono kokybę ir įtaką bendrai sistemos struktūrai.

1.2.1. Bendro sąryšio principas

Klasės pakete turėtų būti susietos kartu, kad turėtų tą pačią prielaidą pasikeisti. Pakeitimas, kuris paveikia paketą, paveikia visas to paketo klases ir jokių kitų paketų.

Kaip teigia vienos atsakomybės principas (angl. *Single responsibility principle*), klasė turėtų neturėti skirtingų prielaidų keistis. Šis principas taip pat teigia, kad paketas taip pat neturėtų turėti skirtingų prielaidų pasikeisti. Principas ragina suburti visas klases, kurios gali keistis dėl tų pačių prielaidų, į vieną vietą. Jei dvi klasės yra taip stipriai susietos, kad jos visada keičiasi kartu, tada jos turėtų būti tame pačiame pakete. Kai reikia išleisti pakeitimus, geriau, kad visi pakeitimai būtų viename pakete. Tai sumažina darbo krūvį, susijusį su pakeitimu išleidimu, pakartotiniu patvirtinimu ir programinės įrangos perskirstymu, be reikalo neatliekant validacijos ir neleidžiant kitų, nesusijusių modulių [Mar02].

1.2.2. Aciklinių priklausomybių principas

Paketo priklausomybės diagramoje neturi būti žiedinių ciklų.

Priklausomybių ciklai sukuria neatidėliotinų problemų. Žiedinės priklausomybės gali sukelti domino efektą, kai nedidelis, lokalus vieno modulio pokytis išplinta į kitus modulius, dėl to, norint testuoti vieną nedidelį modulį, reikia iš naujo sukompiliuoti didžiulę sistemos dalį. Taip pat žiedinės priklausomybės lemia programos ir kompiliavimo klaidas, kadangi pasidaro labai sunku sudaryti tvarką, kaip kompiliuoti paketus. Žiedinės priklausomybės taip pat gali sukelti begalinę rekursiją, kuri sukelia nemalonių problemų tokioms kalboms kaip Java, kurios skaito savo deklaracijas iš sukompiliuotų dvejetainių failų [Mar02].

1.2.3. Stabilių priklausomybių principas

Paketų priklausomybės turėtų laikytis stabilumo krypties. Sistema negali būti visiškai statiška. Norint jai plėstis būtinas tam tikras nepastovumas. Kai kurie paketai yra sukurti taip, kad būtų nepastovūs, tikimasi, kad sistemai plečiantis, jie keisis. Nuo paketo, kuris, manoma, yra nepastovus, neturėtų priklausyti sunkiai pakeičiami paketai, priešingu atveju nepastovų paketą taip pat bus

sunku pakeisti. Gali susiklostyti situacija, kad modulis, sukurtas taip, kad jį būtų lengva pakeisti, kartais tampa sunkiai keičiamu, pridėjus netinkamą priklausomybę nuo jo [Mar02].

1.2.4. Stabilių abstrakcijų principas

Paketai turi būti tiek abstraktūs, kiek ir stabilūs. Šis principas nustato ryšį tarp stabilumo ir abstraktumo. Stabilūs paketai turėtų būti abstraktūs, todėl ir lengvai praplečiami. Tai pat šis principas teigia, kad nestabilus paketas turi būti konkretus, nes jo nestabilumas leidžia lengvai pakeisti jo turinio kodą. Taigi, jei paketas yra stabilus, jį taip pat turėtų sudaryti abstrakčios klasės, užtikrinant jo išplečiamumą. Stabilūs paketai, kurie yra lengvai išplečiami, yra lankstūs pakeitimams, nedarant didelės įtakos sitemos struktūrai [Mar02].

1.3. Paketų kokybės matai

Beveik visi autoriaus aprašyti principai turi aiškiai apibrėžtus matus, kuriais galima patikrinti, kiek paketas laikosi šių principų. Būtent šie matai bus naudojami įvertinti paketus analizuojamuose šablonuose norint patikrinti šablonų poveikį sistemai:

- **Klasių skaičius** (N) – klasių skaičiaus matas paketui nurodo, kiek klasių (konkrečių ir abstrakčių) yra pakete. Šis matas matuoja paketo dydį.
- **Aferentinės jungtys (angl. *Afferent Couplings*)** (Af) – aferentinių jungčių matas nurodo skaičių paketų, kurie priklauso nuo klasių, esančių pasirinktame pakete. Šis matas matuoja ateinančias priklausomybes.
- **Eferentinės jungtys (angl. *Efferent Couplings*)** (Ef) – eferentinių jungčių matas nurodo skaičių kitų paketų, nuo kurių priklauso klasės pasirinktame pakete. Šis matas matuoja išeinančias priklausomybes.
- **Stabilumas** (S) – stabilumo matas nurodo santykį tarp eferentinių jungčių ir visų jungčių (aferentinės + eferentinės) pakete. Šis matas matuoja paketo atsparumą pokyčiams, kuris buvo akcentuojamas stabilų priklausomybių principu:

$$Stabilumas(S) = \frac{Jungtys_{eferentines}}{Jungtys_{eferentines} + Jungtys_{aferentines}} \quad (1)$$

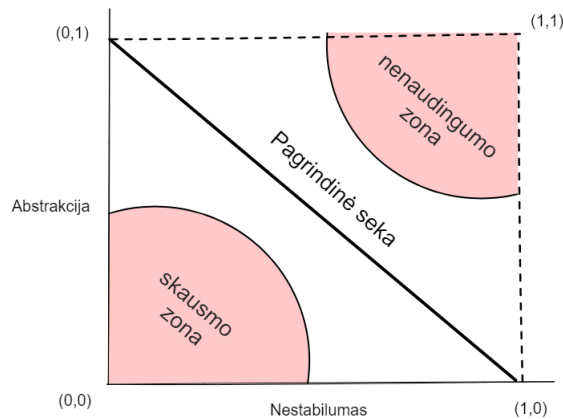
Reikšmės režiai – nuo nulio iki vieno, kur nulis nurodo visiškai stabilų paketą, o vienetą – visiškai nestabilų.

- **Abstrakcija** (A) – paketo abstrakcijos matas nurodo santykį tarp abstrakčių klasių (arba sąsajų (angl. *interface*)) pakete ir bendro klasių skaičiaus:

$$Abstrakcija(A) = \frac{N_{abstrakcios}}{N_{visos}} \quad (2)$$

Šis matas nurodo paketo abstraktumą. Abstrakcijos reikšmė gali būti tarp nulio ir vieno. Nulis reiškia, kad paketas neturi jokių abstrakčių klasių, o vienetą nurodo, kad pakete yra tik abstrakčios klasės.

- **Atstumas nuo pagrindinės sekos (D)** - Pagrindinė seka yra sąryšis tarp nestabilumo ir abstrakcijos. Pagrindinė seka - idealizuota linija ($\text{Abstrakcija} + \text{Nestabilumas} = 1$) kurią galima vaizduoti kaip kreivę, su abstrakcijos dydžiu y ašyje ir nestabilumu x ašyje. Paketas tiesiai ant pagrindinės sekos yra optimaliai subalansuotas atsižvelgiant į jo abstraktumą ir stabilumą. Idealūs paketai yra arba visiškai abstraktūs ir stabilūs ($\text{Stabilumas}=0$, $\text{Abstrakcija}=1$) arba visiškai konkretūs ir nestabilūs ($\text{Nestabilumas}=1$, $\text{Abstrakcija}=0$).



1 pav. Pagrindinės sekos kreivė

Atstumą nuo pagrindinės sekos galima apskaičiuoti kaip:

$$\text{Atstumas} = |\text{Abstrakcija} + \text{Nestabilumas} - 1| \quad (3)$$

Šis matas yra paketo abstraktumo ir stabilumo pusiausvyros rodiklis. Šio mato diapazonas yra nuo nulio iki vieno, kur nulis reiškia paketą, sutampantį su pagrindine seka, o vienas - paketą, maksimaliai nutolusį nuo pagrindinės sekos.

- **Žiedinės priklausomybės (C)** - žiedinių priklausomybių matas skaičiuoja atvejus, kur pasirinkto paketo išeinančios priklausomybės taip pat yra paketo ateinančios priklausomybės (tiesiogiai arba netiesiogiai). Šis matas - aciklinių priklausomybių rodiklis, minėtas aciklinių priklausomybių principu.

Šių matų patikimumas buvo ivertintas atvejo analizėje *Exploring the Relationships between Design Metrics and Package Understandability: A Case Study*, kurioje autoriai tyrinėjo sąryšį tarp minėtų matų ir vidutinių pastangų, reikalingų suprasti objektinio dizaino paketą. Tyrimas atliktas naudojant aštuoniolika paketų, paimtų iš dviejų atviro kodo programinės įrangos sistemų. Paskaičiuoti pastangas, reikalingas paketui suprasti, buvo pasitelktos trys skirtingos komandos, turinčios po tris, panašią patirtį turinčius programuotojus. Jų buvo paprašyta pilnai suprasti paketų funkcionalumą ir nuo vieno iki dešimt įvertinti pastangas, reikalingas suprasti kiekvieną paketą. Rezultatai gauti iš šio tyrimo rodo statistškai reikšmingą koreliaciją tarp daugumos matų ir paketų suprantamumo [Eli10].

2. Galimi kodo skirstymo į paketus šablonai

Skirstymo pagal dalykinės srities esybes metodas negali būti vienareikšmiškai pritaikomas kiekvienoje situacijoje. Tai lemia, kad praktikoje jo nėra griežtai laikomasi – klasės būna išskaidytos remiantis papildomomis taisyklėmis, siekiant išspręsti sistemos planavimo metu kylančias problemas. Norint išskirti šias papildomas taisykles, šablonus, buvo nagrinėjamos atviro kodo sistemos, stebimi nukrypimai nuo bendresnio kodo skirstymo būdo ir identifikuojami klausimai arba problemos, kurias jais buvo bandoma išspręsti.

2.1. Sistemų analizės procesas

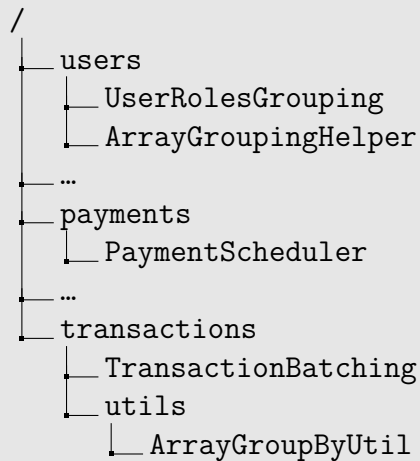
Praktikoje naudojamų šablonų paieška buvo atliekama stebint atviro kodo sistemų struktūrą ir bandant nustatyti, kodėl pasirenkama nukrypti nuo pagrindinio skirstymo būdo. Nagrinėtos įvairių tipų sistemos – tiek taikomoji, tiek techninė programinė įranga. Pastebėta, kad nukrypimai nuo pagrindinio skirstymo būdo dažnai pasikartodavo bandant išspręsti tas pačias problemas, tačiau jų sprendimo šablonai gali būti skirtingi.

2.2. Problemos

Analizuotose sistemose buvo rastos šios problemos, būdingos sistemų dizainui, kartu su jas sprendžiančiais šablonais ir pavyzdžiais, kokiose sistemose jie yra.

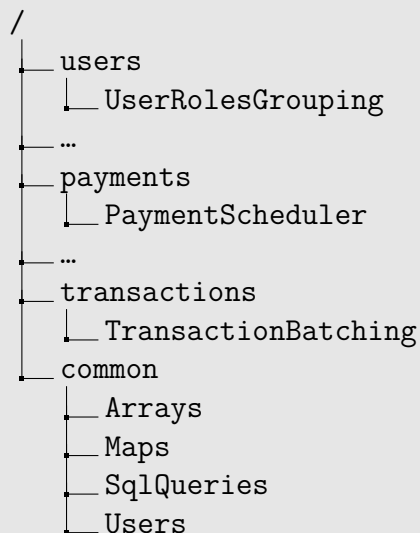
2.2.1. Pagalbinių, daugkartinio naudojimo klasių skirstymas

Pagalbinės ir daugkartinio naudojimo klasės dažniausiai negali būti patogiai grupuojamos skirstant pagal dalykinės srities esybes – pagalbinės klasės gali naudoti kelios esybės, todėl neaišku, prie kurios jas reikėtų priskirti. Didžiausia problema, susijusi su pagalbinių klasių, kurios turėtų išspręsti dažnai sistemoje sutinkamas problemas – inžinieriai, dirbantys prie sistemos nežino apie jų egzistavimą, todėl jų nenaudoja, tai veda prie didesnio kodo pasikartojimo arba kelių skirtingų to paties pagalbinių funkcionalumo įgyvendinimų.



2 pav. Sistemos pavyzdys, kur labai panašią funkciją atliekančios klasės *ArrayGroupByUtil* ir *ArrayGroupingHelper* egzistuoja todėl, kad inžinierius nerado jau įgyvendintos klasės, dėl aiškios struktūros daugkartinio panaudojimo klasėms trūkumo.

Vienas iš šablonų, sprendžiančių šią problemą, galėtų būti turėti vieną paketą, skirtą visoms pagalbinėmis klasėmis, kuris yra paminėtas sistemos dokumentacijoje ir apie jo egzistavimą teoriškai žino visi komandos nariai. Pakete reiktų turėti atskiras klases kiekvienai bendrinei dalykinei sričiai, iš kurios pavadinimo programuotojas galėtų nuspresti, kad jo ieškomas funkcionalumas bus būtent toje klasėje.



3 pav. Sistemos pavyzdys, kur visas bendrinio panaudojimo kodas guli *common* pakete, pirmame sistemos paketų lygyje, todėl pagalbinės klasės yra lengvai randamos.

Jei pagalbinių klasių dydis labai išauga, galima vietoje vienos atskiros klasės vienai bendrinei sričiai sukurti vieną paketą, ir jame turėti kelias pagalbines klases, susijusias su ta dalykine sritimi. Toks papildomas skirstymo šablonas gali praplėsti skirstymą pagal dalykinę sritį – jei bendro naudojimo klasės grupuojamos pagal dengiamą sritį, toks skirstymas atitinka bendresnį. Tokiu

atveju reikia užtikrinti, kad iš klasių pavadinimo aišku, kokią smulkesnę dalykinės srities sritį padengia klasė.

```
/common
├── arrays
│   ├── ArrayFilters
│   └── ArrayComparators
├── ...
├── maps
│   ├── MapTransformations
│   └── MapJoining
├── json
│   └── JsonParser
├── ...
└── database
    ├── DatabaseConnection
    └── DatabaseQueries
```

4 pav. Sistemos pavyzdys, kur bendrinio panaudojimo kodas guli *common* pakete, po dalykinės srities subpaketais, taip sumažinant klasių dydį

Tokį pagalbinių klasių skirstymo metodą naudoja keletas repozitorijų – pavyzdžiui, užrašinės aplikacijos *Omni-Notes*¹ bei *Fire Sticker*². *Omni-Notes* atveju, *utils* pakete esančios klasės papildomai skirstomos pagal atitinkamą dalykinę sritį. Naudojant tokį šabloną, programuotojas, susiduriantis su bendrine problema, kuri, labai tikėtina, jau yra išspręsta sistemoje, turėtų aiškų procesą, kaip elgtis šioje situacijoje:

1. Atsidaryti vieną paketą, skirtą bendrinio panaudojimo kodui
2. Pakete surasti klasę, kurios pavadinimas būtų susijęs su jo problema
3. Klasės funkcijų saraše surasti jam tinkamą funkciją.
4. Jei reikalingas funkcionalumas nerastas, įgyvendinti jį pasirinktoje klasėje, padengti jį testais, bei aprašyti dokumentaciją, kaip funkcija turėtų būti naudojama.
5. Iškviesti rastą arba sukurtą funkciją iš bendrinio panaudojimo kodo paketo savo funkcionalume

Šiai problemai spręsti galimas ir kitoks būdas – turėti keletą pagalbinių klasių paketų, patalpintų po atitinkamu bendresniu dalykinės srities paketu. Tokį skirstymo į paketus šabloną naudoja transakcijoms skirta programinė įranga *Seata*³. Čia pagalbinių klasių yra *utils* pakete žemesniame sistemos lygmenyje, pavyzdžiui, po *engine* paketu.

¹<https://github.com/federicoiosue/Omni-Notes/tree/develop/omniNotes/src/main/java/it/feio/android/omninotes/utils>

²https://github.com/hackjutsu/Fire_Sticker/tree/master/app/src/main/java/com/gogocosmo/cosmoqu/fire_sticker/Utils

³<https://github.com/apache/incubator-seata/tree/2.x/saga/seata-saga-engine/src/main/java/org/apache/seata/saga/engine/utils>

```

/engine
├── strategy
├── sequence
├── ...
├── utils
│   └── ExceptionUtils
│       └── ...

```

5 pav. Sistemos pavyzdys, kur bendrinio panaudojimo kodas guli *utils* pakete esančiame po dalykinės srities subpaketais

2.2.2. Didelis klasių skaičius pakete

Sistemai plečiantis, neišvengiamai didėja ir klasių skaičius. Net pasirinkus tinkamą skirstymo klasių skirstymo į paketus metodą ir palaikant šią tvarką, klasių skaičius pakete gali išaugti. Tokį pavyzdį galima matyti paieškos įrankio *Elasticsearch*⁴ *transport* pakete. Nors klasės yra sugrupuotos pagal dalykinės srities esybę, bendras jų skaičius pakete siekia beveik devyniasdešimt. Tokiame pakete naviguoti sunku, jis taip pat gali turėti daug priklausomybių. Jei tokia tendencija yra būdinga visai sistemai, sistema tampa mažiau lanksti pokyčiams, kadangi net ir paprastas pakeitimas daro įtaką reišmingai sistemos daliai, pokyčiai yra labiau linkę keisti bendrą sistemos architektūrą. Taip pat naujos sistemos versijos išleidimo (angl. *release*) procesas tampa sudėtingesnis, kadangi yra paveikiama daugiau klasių. Dažnai matoma, kad susidarius šiai situacijai, nukrypstama nuo skirstymo pagal dalykinės srities esybes, pradedamos grupuoti techninio sluoksnio klasės. Tokį skirstymą galima matyti *DBeaver* duomenų bazės įrankyje⁵ – Aukštesniame lygmenyje grupuojant pagal dalykinės srities esybes, pavyzdžiui, *MsSql*, vėliau pereinama prie skirstymo pagal techninį sluoksnį pakete *model*. Robert C. Martin bendro sąryšio principas, kuris teigia, kad visos tarpusavyje susijusios klasės turėtų būti vienam pakete, akcentuoja siekiamybę turėti gan mažus paketus, turinčius aiškiai apibrėžtą funkcionalumą, priežastį egzistuoti, taip užtikrinant ir glaudų tarpusavyje susijusių klasių sąryšį. Šis principas galėtų būti kodo skirstymo šablonas, užtikrinantis mažesnę klasių bei aferentinių jungčių skaičių paketuose. Vadovaujantis šiuo šablonu, klases reikėtų skirstyti pagal smulkesnį jų teikiamą funkcionalumą. Tokį skirstymo tipą galima stebėti *IntelliJ* įskiepyje⁶. Čia *plugin* paketo esybė yra papildomai išskirstyta pagal *IntelliJ* sistemos domeno sistemos esybes. Šį šabloną dar būtų galima patobulinti sprendžiant aferentinių jungčių skaičiaus problemą. Didelis priklausomybių nuo specifinio paketo skaičius (arba aferentinės jungtys), reiškia, kad pokyčiai tame pakete turės įtaką kelioms klasėms. Funkcionalumas kitiems paketams galėtų būti pasiekiamas per vieną minimalią sąsają (angl. *interface*), kuri atskleidžia tik konceptus (metodus arba duomenų tipus), kurie yra glaudžiai susiję su komponento teikiama paslauga, bei klase, grąžinančią minėtos

⁴<https://github.com/elastic/elasticsearch/tree/main/server/src/main/java/org/elasticsearch/transport>

⁵<https://github.com/dbeaver/dbeaver/tree/devel/plugins/org.jkiss.dbeaver.ext.mssql/src/org/jkiss/dbeaver/ext/mssql>

⁶<https://github.com/wix-incubator/source-wizard/tree/master/src/main/kotlin/plugin>

sąsajos įgyvendinimą. Tai gali būti paprasta klasė, su statine funkcija, kurios rezultatas yra ši sąsaja, arba, esant keliomis sąsajos implementacijomis, *Static factory* dizaino šablonas, kuris nusprendžia, kurį įgyvendinimą reikėtų grąžinti pagal paduotus argumentus. Visos kitos klasės naudoja *package* pasiekiamumo modifikatorių, taip kompiliatoriaus pagalba užtikrinant, kad jos nebus pasiektos iš išorės. Paketas, turintis vieną funkciją, yra naudojamas tik tų paketų, kuriems reikia būtent tos funkcijos, taip užtikrinant tik mažos sistemos dalies priklausomybę nuo vieno paketo.

Toks principas yra sutinkamas *Typescript* programavimo kalboje, kur kiekvienas modulis (šios kalbos paketo atitikmuo) turi *index.ts* failą, veikiantį kaip sąsaja, apibrėžianti, kokios modulio klasės bei funkcijos gali būti pasiekiamos už paketo ribų. Tai užtikrina glaustą kontraktą ir aiškų modulio funkcionalumą, kurį gali pasiekti kiti moduliai, bei paslepia modulio įgyvendinimo detales. Toki šabloną naudoja turinio valdymo sistema *Keystone*⁷, kur kiekviena esybė turi savo modulį, atliekantį vieną funkciją, kurios kontraktas aprašytas *index.ts* faile.

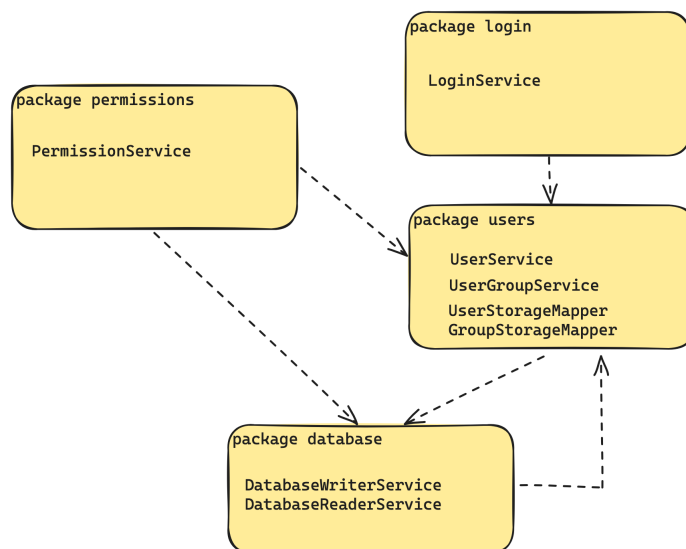
```
/types
├── image
│   ├── views
│   ├── utils
│   └── index
├── text
│   ├── views
│   └── index
├── ...
└── checkbox
    ├── views
    └── index
```

6 pav. *Keystone* sistemos dalis, kurioje matomi smulkūs moduliai su *index.ts* failuose aprašytais kontraktais

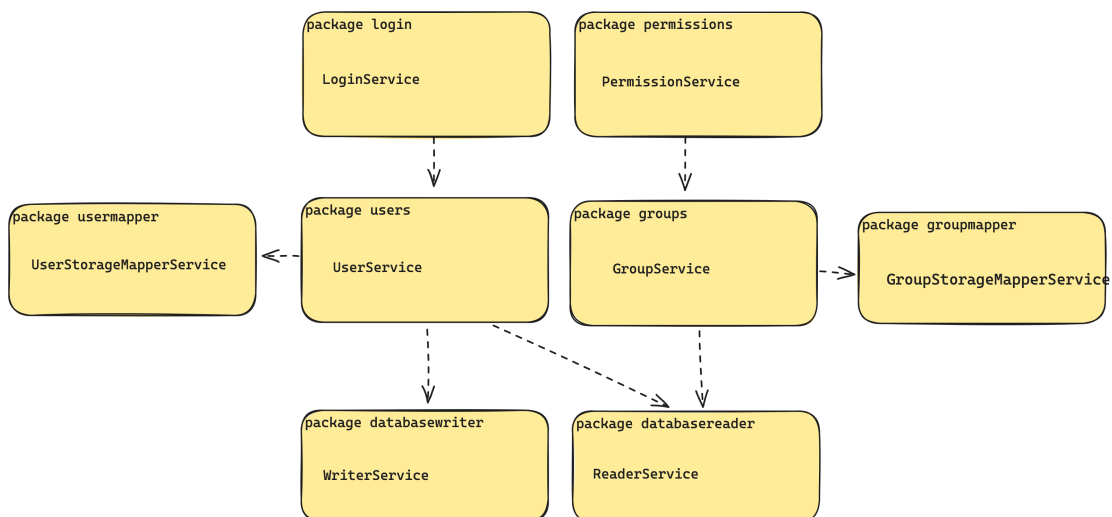
Taip pat mažas paketo funkcionalumas reiškia, kad minėtas paketas skirtas funkcionalumui įgyvendinti naudos minimalų kitų sistemos esybių skaičių, taip sumažinant ir eferentinių jungčių skaičių.

Žemiau esančiuose paveikslėliuose galima matyti, kaip išskaidant paketus, turinčius kelis funkcionalumus, yra sumažinamas paketų priklausomybių skaičius.

⁷<https://github.com/keystonejs/keystone/tree/main/packages/core/src/fields/types>



7 pav. Sistemos pavyzdys su kelias funkcijas atliekančiais paketais



8 pav. Sistemos pavyzdys su aiškia, vieną funkciją turinčiais paketais

Toks skirstymo būdas taip pat sprendžia ciklinių priklausomybių problemą – pavyzdžiui, įrankio, skirtą spręsti ciklinių priklausomybių problemą, kūrimo aprašas [Aga15] teigia, kad ciklinių priklausomybių problemą galima spręsti laikantis trijų principų – bendro panaudojimo, bendro keitimosi bei paleidimo ir pernaudojimo ekvivalentumo. Šie principai yra išvesti iš bendro sąryšio principo ir akcentuoja, kad kartu besikeičiančios klasės turėtų būti viename pakete. Norint užtikrinti šiuos reikalavimus, reikėtų turėti mažą, tiksliai apibrėžto funkcionalumo paketą, kurio visi komponentai – tarpusavyje susiję. Tokiu atveju ciklinių priklausomybių tikimybė sumažėja.

2.2.3. Daug skirtingų sąsajų implementacijų

Sistemai plečiantis galima susidurti su problema, kad išauga sąsajos implementacijų skaičius. Ši problema gali iškilti skirstant klases pagal dalykinės srities esybes, gali būti neaišku, kur turėtų būti sukuriamas naujas esybės paketas. Tokiu atveju navigacija pakete pasidaro sudėtinga, neaišku, kas kurią klasę implementuoja. Šią problemą galima spręsti sukuriant sąsajos implementacijas

tame pačiame pakete kaip ir pati sąsaja. Toks skirstymo į paketus šablonas yra sutinkamas *Java* standartinėje bibliotekoje⁸, kur visos standartinės sąsajos bei jų implementacijos yra tame pačiame pakete. Dėl tokios paketų struktūros, inžinieriai, naudojantys standartinę *Java* biblioteką, gali lengvai rasti jiems reikiamus sąsajos įgyvendinimus:

```
/util
├── Map
├── HashMap
├── LinkedHashMap
├── WeakHashMap
├── TreeMap
├── ...
├── List
├── LinkedList
└── ArrayList
```

9 pav. *Java* standartinės bibliotekos pavyzdys, kuriame sąsajų įgyvendinimai guli tame pačiame pakete, kaip ir sąsajos

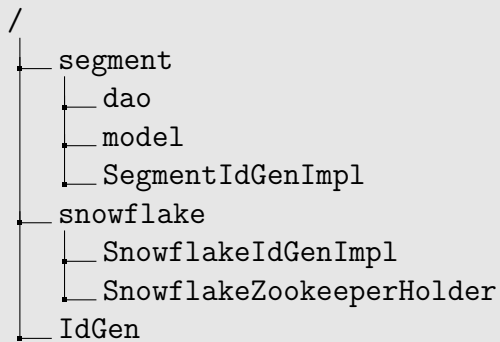
Šis šablonas turi vieną trūkumą – pakete su dideliu klasių kiekiu, bei keliomis skirtingomis sąsajomis sunku suprasti, kuri klasė kuria sąsają įgyvendina. Tai pastebima ir *Java* bibliotekos pavyzdyje – *utils* paketas iš viso turi daugiau nei 50 klasių, iš kurių bent 10 yra sąsajos, o visa kita – jų implementacijos, todėl greitai rasti norimą sąsajos įgyvendinimą vis tiek užtrunka.

Šio šablono trūkumas gali būti išspręstas, jį šiek tiek pakeitus – sąsaja, kuri turi daug skirtingų įgyvendinimų, turėtų turėti specifiškai jai sukurtą paketą, o kiekvienam paketo įgyvendinimui sukuriamas po subpaketas, kuriame yra tiek klasė įgyvendinanti minėtą sąsają, tiek kitos, įgyvendinimui reikalingos klasės. Naudojant tokią kodo struktūrą galima labai greitai rasti sąsajas bei galimus jos įgyvendinimus, beveik netyrinėjant sistemos struktūros, bei klasių kodo. Šis šablonas taip pat užtikrina, kad paketas turi vieną aiškų funkcionalumą, todėl laikosi bendro sąryšio principo. Šablone aprašyta paketų struktūra buvo sutikta tokiose sistemose kaip *HikariCP*⁹, skirtoje efektyviam *Java* programos prisijungimui prie duomenų bazių, bei *Leaf*¹⁰, naudojamoje unikalaus identifikatoriaus generavimui.

⁸<https://github.com/AdoptOpenJDK/openjdk-jdk11/tree/master/src/java.base/share/classes/java/util>

⁹<https://github.com/brettwooldridge/HikariCP/tree/dev/src/main/java/com/zaxxer/hikari>

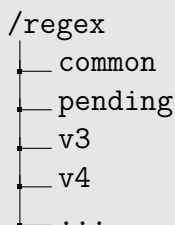
¹⁰<https://github.com/Meituan-Dianping/Leaf/tree/master/leaf-core/src/main/java/com/sankuai/inf/leaf>



10 pav. *Leaf* sistema, kur kiekvienas sąsajos įgyvendinimas yra aprašytas subpakete

2.2.4. Esybių pokyčių ir versijavimo valdymas

Sistemai egzistuojant ilgesnį laiką, jos pokyčiai pasidaro neišvengiami. Smulkūs pokyčiai nebūtinai paveikia bendrą sistemos struktūrą, tačiau didesniems pokyčiams kartais reikia sukurti naujas klasių ar funkcionalumų versijas. Jei reikia palaikyti atgalinį suderinamumą (angl. backward compatibility), sistemoje gali atsirasti kelios tų pačių esybių versijos. Tokiu atveju sistemoje egzistuojančios kelios tų pačių esybių versijos gali pridėti painumo. Ši problema dažnai sprendžiama sukuriant atskirus paketus skirtingoms versijoms (v1, v2, v3, ...) bei iškeliant bendrą, abiejų esybių naudojamą kodą į paketus taip, kad jas galėtų pasiekti abi versijos. Taip galima išvengti kodo duplikacijos bei perteklinio klasių skaičiaus paketuose, bei užtikrinant aiškią skirtingų versijų atskirtį. Toks skirstymas gali praplėsti skirstymą pagal dalykinės srities esybes – skirstomos versijos gali atspindėti reikalingus palaikyti besikeičiančius verslo reikalavimus.



11 pav. *Mongo* duomenų bazės pavyzdys, kuriame skirtingos versijos patalpintos atskiruose paketuose

Toks skirstymo būdas matomas keliose repozitorijose – *Mongo*¹¹ duomenų bazėje, API skirtame kelionių valdymui *travels-java-api*¹² bei duomenų saugojimo įrankyje *nocodb*¹³. Šiose repozitorijose v1, v2 pavadinti paketai saugo skirtingas esybių versijas.

Apibendrinant, galima išskirti aukščiau aprašytus šablonus:

¹¹https://github.com/mongodb/mongo/tree/master/src/third_party/boost/boost/regex

¹²<https://github.com/mariazevedo88/travels-java-api/tree/master/src/main/java/io/github/mariazevedo88/travelsjavaapi/controller>

¹³<https://github.com/nocodb/nocodb/tree/c7cc1f92fd77f8b5daefceb7148aab4a69cb9b4e/packages/nocodb/src/meta/migrations>

Šablonas	Atskiras pagalbinių klasių paketas
Sprendžiama problema	Pagalbinių klasių skirstymas
Siūlomas sprendimas	Sukurti atskirą pagalbinių klasių paketą dalykinės srities esybių lygmenyje
Galimos variacijos	Gali būti papildomai sukuriami subpaketai pagal pagalbinių klasių dengiamą funkcionalumą

Šablonas	Pagalbinių klasių priskyrimas dalykinės srities paketui
Sprendžiama problema	Pagalbinių klasių skirstymas
Siūlomas sprendimas	klases laikyti <i>util</i> paketuose, žemesniame lygmenyje nei dalykinės srities esybės

Šablonas	Skirstymas pagal teikiamą funkcionalumą
Sprendžiama problema	Didelis klasių skaičius pakete
Siūlomas sprendimas	klases suskirstyti pagal smulkesnį jų teikiamą funkcionalumą
Galimos variacijos	Smulkūs moduliai su atskirose klasėse aprašytais kontraktais

Šablonas	Sąsajų atskyrimas
Sprendžiama problema	Daug skirtingų sąsajų implementacijų
Siūlomas sprendimas	Kiekvienam paketo įgyvendinimui sukurti subpaketą, kuriame yra tiek klasę įgyvendinanti sąsaja, tiek kitos, įgyvendinimui reikalingos klasės

Šablonas	Atskirų versijų skirstymas
Sprendžiama problema	Esybių pokyčių ir versijavimo valdymas
Siūlomas sprendimas	Sukurti atskirus paketus skirtingoms versijoms bei iškelti bendrą, abiejų esybių naudojamą kodą į paketus

3. Įrankiai šablonų analizei ir įvertinimui

Kompiuterinės sistemos, kurioms yra aktualu klasių ir paketų skirstymo metodai, paprastai yra labai didelės. Pilnai perprasti tokias sistemas, nustatyti jų įgyvendinimo kokybę, naudojamus šablonus kodui skirstyti, apskaičiuoti paketų kokybės matavimus yra sudėtingas procesas. Daryti tai rankiniu būdu užtrunka daug laiko bei yra paliekama daug vietos potencialioms klaidoms, todėl yra būtina šį procesą optimizuoti, skaitmenizuoti analizės procedūras pasitelkiant aiškiai apibrėžtą programinių įrankių pagalbą.

3.1. Reikalavimai įrankiams

Įrankių, leidžiančių paprasčiau atlikti sistemų analizę, atsakomybės galima suskirstyti į dvi grupes:

- Bendrinė sistemos analizė – įrankis ar įrankiai padeda atlikti bendrinę sistemos analizę. Šios atsakomybių grupės įrankių išvestis nėra objektyvūs, tiksliai apibrėžtų formulių rezultatai, o papildoma, aiškiai pavaizduota meta informacija apie sistemą – paketų struktūrą, juose esančių klasių priklausomybes, figūruojančių paketų bei klasių vardus. Ši papildoma informacija nėra aiškūs teiginiai, o tik pagalba analizę atliekančiams asmeniui, leidžianti priimti išvagas apie sistemą, kaip pavyzdžiui, kokiam paketų skirstymo šablonui yra artimiausia sistemos struktūra, arba kaip lengvai sistema yra suprantama.
- Paketų kokybės matų skaičiavimas – įrankis ar įrankiai turi padėti apskaičiuoti aprašytus paketų kokybės matavimus. Šių įrankių išvestis – tikslūs, formulėmis pagrįstų skaičiavimų rezultatai apie paketų atitikimą kriterijams, kuriuos galima lyginti tarpusavyje.

3.2. Reikalavimai bendrinės sistemos analizės įrankiui

Įrankis bendrinei sistemos analizei atlikti turėtų suteikti galimybę naudotojui nurodyti kelią iki *Java* programavimo kalba parašytos sistemos arba posistemės ir joje atlikti jos turinio analizę bei naudotojui pateikti naudingas išvadas, sudarytas iš:

- Klasių ir paketų skaičiaus
- Vidutinio klasių pakete skaičiaus
- Paketų ir klasių medžio, identifikuojančio abstrakčias klases ar sąsajas
- Paketų priklausomybių grafiką

Gautą rezultatą išvesti suprantamu formatu, leidžiant vartotojui susidaryti išvadas apie sistemos arba tam tikros posistemės struktūrą, naudojamus įrankius bei kokybę.

3.3. Reikalavimai įrankiui paketo kokybei skaičiuoti

Įrankis paketo kokybei skaičiuoti turėtų suteikti galimybę naudotojui nurodyti kelią iki *Java* programavimo kalba parašytos sistemos arba posistemės ir joje apskaičiuoti kiekvieno paketo kokybės matavimus:

- Klasų skaičių
- Aferentinių jungčių skaičių
- Eferentinių jungčių skaičių
- Nestabilumo santykį
- Abstrakcijos santykį
- Atstumo nuo pagrindinės sekos santykį
- Žiedinių priklausomybių skaičių

Gautą rezultatą išvesti vartotojui suprantamu formatu, kuriame matytųsi individualių paketų matai, bei šių matų vidurkis sistemoje (arba posistemėje). Išvedimo formatas turėtų būti toks, jog skirtingų analizių rezultatai būtų lengvai palyginami su kitais.

Abieiose įrankiuose vienas iš palaikomų išvesties formatų turėtų būti *latex*, taip suteikiant galimybę analizės rezultatus pateikti tolesniame šio dokumento turinyje.

3.4. Įrankių įgyvendinimas

Nors beveik visiems reikalavimuose minimiems funkcionalumams galima rasti jau sukurtų įrankių, greit ir efektyviai pritaikyti juos skirtingoms sistemoms (arba posistemėms) nėra patogu – kiekvieną įrankį reikėtų vykdyti atskirai, skirtingais vykdymo procesais ir argumentais. Taip išvestų rezultatų formatai yra skirtingi. Todėl, norint palengvinti šį procesą – suvienodinti procesų vykdymą bei gautus rezultatus, visi įrankiai, reikalingi analizei, įgyvendinti kaip viena programa, kuri apdoroja failus nurodytoje sistemos direktorijoje, nuskaityto *java* failų turinį ir išveda informaciją apie sistemos paketus bei klases. Surinkta informacija naudojama įgyvendinti kiekvienam aprašytam įrankio funkcionalumui, ten, kur galima, naudojant jau parašytus įrankius, taip programiškai supaprastinant skirtingų įrankių vykdymą.

4. Esamų sistemų pertvarkymas

Apibrėžus šablonus paketų skirstymui galima pradėti vertinti jų įtaką sistemai, juos pritaikant esamoms sistemoms. Šio skyriaus tikslas – pasirinkti ir išnagrinėti kelias sistemas, pasitelkus paketų kokybės matus, bei bendrą sistemos struktūros analizę, taip identifikuojant programiniam kodui būdingas problemas, rastas problemas išspręsti pritaikant aprašytus šablonus. Atlikus pertvarkymus sistemose, dar kartą paskaičiuoti paketų kokybės matus, taip gaunant įrodymus, ar gauti šablonai yra efektyvūs ir iš tiesų sprendžia jiems priskirtas problemas.

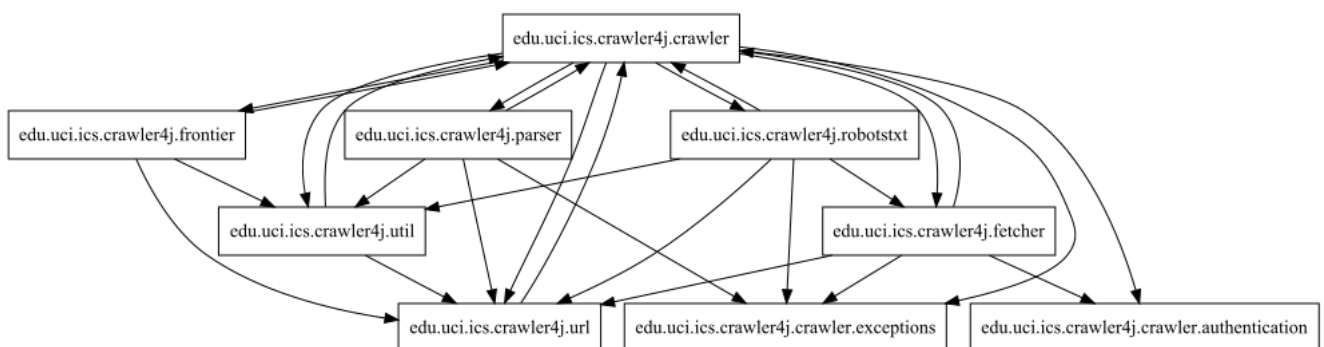
4.1. Sistemų pasirinkimas

Sistemų pertvarkymui pasirinktos atviro kodo sistemos, kurių kodas yra viešai prieinamas *github* platformoje. Pasirinktos sistemos yra vidutinio dydžio, todėl nėra labai sudėtinga jas suprasti ir pertvarkyti, bet taip pat jos nėra tokios paprastos, kad neturėtų sistemos dizaino problemų. Pasirinktos sistemos yra skirtingo tipo projektai, taip užtikrinant didesnę problemų įvairovę ir objektyvesnius įvertinimus. Per visą pasirinktų sistemų imtį yra sutinkamos visos aprašytos problemos, taip įvertinant visus aprašytus šablonus.

4.2. *crawler4j* sistema

crawler4j¹⁴ yra atvirojo kodo žiniatinklio tikrinimo (angl. *crawling*) programa parašyta su *Java* programavimo kalba, leidžianti efektyviai tikrinti žiniatinklį naudojant daugiagijį angl-multithreaded mechanizmą. Tai taikomosios programinės įrangos tipas, kurios naudotojas yra žmogus.

Prieš visus pakeitimus *crawler4j* sistemos, paketų struktūra atrodo taip:



12 pav. *crawler4j* sistemos struktūra

Atlikus sistemos kokybės matų analizę galima pamatyti sistemai būdingas problemas:

¹⁴<https://github.com/yasserg/crawler4j>

Paketo vardas	N	A	E	S	A	D	C
edu.uci.ics.crawler4j.crawler.authentication	4	2	0	0.0	0.25	0.75	0
edu.uci.ics.crawler4j.url	4	6	1	0.143	0.0	0.857	1
edu.uci.ics.crawler4j.crawler.exceptions	3	4	0	0.0	0.0	1.0	0
edu.uci.ics.crawler4j.crawler	5	6	8	0.571	0.2	0.229	5
edu.uci.ics.crawler4j.robotstxt	6	1	5	0.833	0.0	0.167	1
edu.uci.ics.crawler4j.frontier	6	1	3	0.75	0.0	0.25	1
edu.uci.ics.crawler4j.fetcher	6	2	4	0.667	0.0	0.333	1
edu.uci.ics.crawler4j.util	3	4	2	0.333	0.0	0.667	0
edu.uci.ics.crawler4j.parser	12	1	4	0.8	0.167	0.033	1

\bar{N}	\bar{A}	\bar{E}	\bar{S}	\bar{A}	\bar{D}	$sum(C)$
5	3	3	0.455	0.069	0.476	5

Iš bendros sistemos analizės bei matų lentelės rezultatu matomos kelios sistemos problemos –

- **Problema numeris 1** – nors ši tai nesimato kokybės matuose, bet atlikus bendra sistemos kodo analizę matoma, kad sistemos paketuose yra daug sąsajų su skirtingais jų įgyvendinimais, todėl yra sunku rasti visus specifinės sąsajos įgyvendinimus, esama paketų struktūra tam negelbėja
- **Problema numeris 2** – paketo *edu.uci.ics.crawler4j.parser* viduje yra 12 klasių, todėl sunku suprasti už kokius funkcionalumus jis atsakingas ir kada jį reikėtų nagrinėti bei keisti.
- **Problema numeris 3** – Sistema turi daug ciklinių prilausomybių kur du arba daugiau paketų priklauso vienas nuo kito. Tai aiškiai matosi sistemos paketų diagramoje¹², ciklinėse priklausomybėse išviso dalyvauja 10 paketų ir kartu jie sudaro 5 ciklinių priklausomybių sekas, kurios turėtų būti pašalintos.

Sistemos paketas *edu.uci.ics.crawler4j.parse*, turi sąsają *ParseData* bei 4 skirtingų jos įgyvendinimus – *CssParseData*, *TextParseData*, *HtmlParseData*, *BinaryParseData*, šios klasės pakete dalinasi vieta dar su 9 klasėmis, todėl yra gan nepatogu rasti ieškomo paketo įgyvendinimo.

```

/parser
├── AllTagMapper
├── ParseData
│   ├── TextParseData
│   └── HtmlParseData
├── CssParseData
├── BinaryParseData
├── ExtractedUrlAnchorPair
├── HtmlParser
├── HtmlContentHandler
├── TikaHtmlParser
└── Parser

```

13 pav. *parser* paketo struktūra

Šiam paketui sutvarkyti taikomas *Sąsajų atskyrimo* šablonas – sąsaja iškeliamą į atskirą paketą *parse*, jos įgyvendinimai perkliami į minėto paketo subpaketus. Taip pat sukuriamą *ParseDataFactory* klasę, kuri grąžina tinkamą *ParseData* sąsajos implementaciją, pagal nurodytą turinio tipą, paslepiant įgyvendinimo pavadinimą nuo funkcijos naudotojo.

```

public final class ParseDataFactory {

    public static ParseData parseDataOf(String contentType) {
        if (Util.hasBinaryContent(contentType)) {
            return new BinaryParseData();
        }

        if (Util.hasCssTextContent(contentType)) {
            return new CssParseData();
        }

        if (Util.hasPlainTextContent(contentType)) {
            return new TextParseData();
        }
        throw new UnsupportedContentTypeException(contentType);
    }
}

```

14 pav. *ParseDataFactory* klasė įgyvendinimas

Šie pertvarkymai pakeičią *parser* paketo struktūrą:


```

/parser
├── parse
│   ├── ParseData // Sąsaja
│   ├── text
│   │   └── TextParseData // Pirmas sąsajos įgyvendinimas
│   ├── html
│   │   └── HtmlParseData // Antras sąsajos įgyvendinimas
│   ├── css
│   │   └── CssParseData // Trečias sąsajos įgyvendinimas
│   ├── binary
│   │   └── BinaryParseData // Ketvirtas sąsajos įgyvendinimas
│   └── factory
│       └── ParseDataFactory // Klasė grąžinanti tinkamą sąsajos įgyvendinimą
├── ExtractedUrlAnchorPair
├── HtmlParser
├── HtmlContentHandler
├── TikaHtmlParser
└── Parser

```

15 pav. *parser* paketo struktūra po pirmojo pertvarkymo

Dėl atskirtos sąsajos ir jos įgyvendinimų tampa daug paprasčiau ją rasti, bei suprasti sistemoje egzistuojančius jos įgyvendinimus. Tokia struktūra išsprendžia sistemos **problemą numeris 1**. Šis pertvarkymas taip išsprendė **problemą numeris 2**, sumažinant klasių skaičių pakete į 5. Palyginus pertvarkytos ir buvusios sistemos matavimus matome, kad nors sistema tapo aiškesnė, tai nepadarė nesumažino paketų kokybės matų:

Paketas <i>edu.uci.ics.crawler4j.parser</i>	<i>N</i>	<i>A</i>	<i>E</i>	<i>S</i>	<i>A</i>	<i>D</i>	<i>C</i>
Prieš pakeitimus	12	1	4	0.8	0.167	0.033	1
Po pakeitimų	5	1	7	0.875	0.167	0.042	1

	\bar{N}	\bar{A}	\bar{E}	\bar{S}	\bar{A}	\bar{D}	$sum(C)$
Prieš pakeitimus	5	3	3	0.455	0.069	0.476	5
Po pakeitimų	4	3	3	0.505	0.108	0.409	5

Kad išspręsti **problemą numeris 3** – ciklines priklausomybes, reikia identifikuoti klases, kurių priklausomybės sudaro ciklus ir iškelti jas į atskirus paketus. todo: pridėti citavimą iš straipsnio apie ciklinių priklausomybių eliminavimą

Kuriant naujus paketus vadovaujamosi *skirstymo pagal teikiamą funkcionalumą* šablonu, užtikrinant, kad kiekvienas paketas teikia vieną, aiškiai apibrėžtą funkciją, kuri yra pasiekama per pakete aprašytą sąsają. Paketo vidus yra pasleptas su *Java* kalbos pasiekiamumo modifikatoriais – konkrečių klasių negalima inicializuoti už paketo ribų, nes jų konstruktoriai privatus. Sąsajos įgyvendinimas gaunamas *static factory* pagalba. todo: aprašyti kas tai yra

```

public interface HtmlParser {

    HtmlParseData parse(ParsedPage page, String contextURL)
        throws ParseException;

    // ąŽGrinamas ąssajos įgyvendinimas
    static HtmlParser newHtmlParser(CrawlConfig config, TLDList
        tldList) throws InstantiationException,
        IllegalAccessException {
        return new TikaHtmlParser(config, tldList);
    }
}

// ąSsajos įgyvendinimas. ėNra švieas, pasiekiamas tik paketo
viduje,
// nes ėklas ir konstruktorius nenaudoja public žžųraktaodi
class TikaHtmlParser implements HtmlParser {
    ...
    TikaHtmlParser(CrawlConfig config, TLDList tldList) throws
        InstantiationException, IllegalAccessException {
        ...
    }
}

```

16 pav. skirstymo pagal teikiamą funkcionalumą šablono sąsaja

Peržiūrėjus paketus su ciklinėmis priklausomybės iš jų buvo išskirti šie nauji funkcionalumai, kuriems buvo sukurti atskiri paketai:

- Iš *parser* paketo iškeltas funkcionalumas *html* turiniui apdoroti, patalpintas į *parser/html* paketą. Po šio pertvarkymo *parser* paketas pasidaro pakankamai mažas turėti viena funkcionalumą – printi puslapį ir deleguoti jį apdorojimui kitai klasei, pagal puslapio tipą.
- Iš *robotstxt* paketo iškeltas funkcionalumas patikrinti ar sistema autorizuota apdoroti pasirinktą puslapį į *robotstxt/permissions* paketą.
- Iš *crawler* paketo iškeltas funkcionalumas, programiškai aprašantys internetinį puslapį į *page* paketą.
- Iš *crawler* paketo iškeltas funkcionalumas, nustatyti įrankio konfigūraciją į *config* paketą.
- Iš *url* paketo iškeltas funkcionalumas, gauti internetinių domenų pavadinimus į *tld* paketą.

```

/
├─ crawler4j
│   ├── crawler
│   │   ├── authentication
│   │   └── exceptions
│   ├── fetcher
│   ├── frontier
│   ├── parser
│   │   └── parse
│   │       ├── binary
│   │       ├── css
│   │       ├── factory
│   │       ├── html
│   │       └── text
│   ├── robotstxt
│   ├── url
│   └── util

```

17 pav. *crawler4j* paketų medis prieš iškėliant minėtus funkcionalumus

```

/
├─ crawler4j
│   ├── config
│   ├── crawler
│   │   ├── authentication
│   │   └── exceptions
│   ├── fetcher
│   ├── frontier
│   ├── page
│   ├── parser
│   │   ├── html
│   │   └── parse
│   │       ├── binary
│   │       ├── css
│   │       ├── factory
│   │       ├── html
│   │       └── text
│   ├── robotstxt
│   │   └── permissions
│   ├── url
│   │   ├── canonicalizer
│   │   └── tld
│   └── util

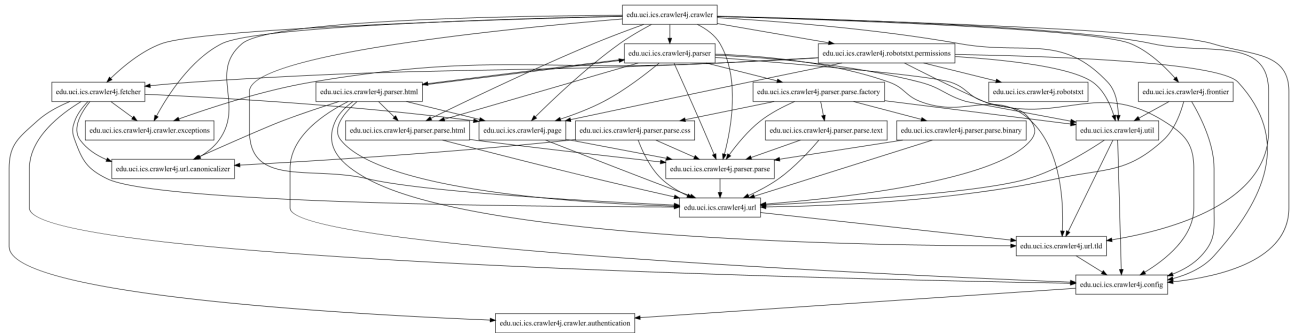
```

18 pav. *crawler4j* paketų medis iškėlus minėtus funkcionalumus

Atlikus šiuos funkcionalumų ištraukimus į atskirus paketus buvo panaikintos žiedinės pri-

klausomybės. Naujoje paketų diagramoje galime matyti, jog dabar sistema turi daug daugiau paketų, tačiau jų priklausomybių kryptys daug aiškesnės:

4.3. *Crawler sistema*



19 pav. *crawler4j* sistemos struktūra po antro pertvarkymo

todo: ranka perpašyti sugeneruotus paveikslėlius, kad jie geriau tilptų į puslapį

Atlikus matų analizę ant naujai pertvarkytos sistemos yra matoma, kad sistemos kokybė nenukrito:

Paketo vardas	N	A	E	S	A	D	C
edu.uci.ics.crawler4j.parser.parse.binary	1	1	2	0.667	0.0	0.333	0
edu.uci.ics.crawler4j.url	1	12	1	0.077	0.0	0.923	0
edu.uci.ics.crawler4j.crawler.exceptions	1	3	0	0.0	0.0	1.0	0
edu.uci.ics.crawler4j.crawler	4	0	13	1.0	0.25	0.25	0
edu.uci.ics.crawler4j.parser.html	2	1	7	0.875	0.5	0.375	0
edu.uci.ics.crawler4j.robotstxt.permissions	2	1	7	0.875	0.5	0.375	0
edu.uci.ics.crawler4j.config	1	8	1	0.111	0.0	0.889	0
edu.uci.ics.crawler4j.frontier	6	1	3	0.75	0.0	0.25	0
edu.uci.ics.crawler4j.page	2	5	2	0.286	0.5	0.214	0
edu.uci.ics.crawler4j.parser.parse.factory	1	1	5	0.833	0.0	0.167	0
edu.uci.ics.crawler4j.parser	6	2	8	0.8	0.0	0.2	0
edu.uci.ics.crawler4j.crawler.authentication	4	2	0	0.0	0.25	0.75	0
edu.uci.ics.crawler4j.parser.parse.css	1	1	3	0.75	0.0	0.25	0
edu.uci.ics.crawler4j.robotstxt	5	1	0	0.0	0.0	1.0	0
edu.uci.ics.crawler4j.url.tld	2	5	1	0.167	0.5	0.333	0
edu.uci.ics.crawler4j.parser.parse.html	1	3	2	0.4	0.0	0.6	0
edu.uci.ics.crawler4j.parser.parse	1	8	1	0.111	1.0	0.111	0
edu.uci.ics.crawler4j.fetcher	6	2	6	0.75	0.0	0.25	0
edu.uci.ics.crawler4j.util	3	5	3	0.375	0.0	0.625	0
edu.uci.ics.crawler4j.parser.parse.text	1	1	2	0.667	0.0	0.333	0
edu.uci.ics.crawler4j.url.canonicalizer	3	4	0	0.0	0.333	0.667	0

	\bar{N}	\bar{A}	\bar{E}	\bar{S}	\bar{A}	\bar{D}	$sum(C)$
Prieš pakeitimus	5	3	3	0.455	0.069	0.476	5
Po pirmos iteracijos	4	3	3	0.505	0.108	0.409	5
Po visų pakeitimų	3	3	3	0.452	0.183	0.471	0

Todėl po šio pertvarkymo galima teigti, kad šablono *skirstymas pagal teikiamą funkcionalumą* ir *sąsajų atskyrimas*, išsprendė visas 3 sistemoje matytas problemas, padarė sistemą lengviau suprantamą žmogui, panaikino ciklinės priklausomybės bet taip pat neturėjo neigiamos įtakos matams, nurodantiems sistemos įgyvendinimo kokybę.

todo: parašyti, pertvarkymo tikslas nėra gerinti matus, o jie yra naudojami tik tam kad isitikinti, jog kokybė nesuprastejo. Svarbiau skaitomumas, o ne kokybė paremta matais todo: pridėti figures prie lentelių

Rezultatai

Rezultatų skyriuje išdėstomi pagrindiniai darbo rezultatai: kažkas išanalizuota, kažkas sukurta, kažkas įdiegta. Tarpinių žingsnių išdavos skirtos užtikrinti galutinio rezultato kokybę neturi būti pateikiami šiame skyriuje. Kalbant informatikos terminais, šiame skyriuje pateikiama darbo išvestis, kuri gali būti įvestimi kituose panašios tematikos darbuose. Rezultatai pateikiami sunumeruotų (gali būti hierarchiniai) sąrašų pavidalu. Darbo rezultatai turi atitikti darbo tikslą.

Išvados

1. Išvadų skyriuje daromi nagrinėtų problemų sprendimo metodų palyginimai, siūlomos rekomendacijos, akcentuojamos naujovės.
2. Išvados pateikiamos sunumeruoto (gali būti hierarchinis) sąrašo pavidalu.
3. Darbo išvados turi atitikti darbo tikslą.

Šaltiniai

- [Aga15] B. Aga. A Tool for Breaking Dependency Cycles Between Packages. 2015 [žiūrėta 2024-04-04]. Prieiga per internetą: <https://scg.unibe.ch/archive/masters/Aga15a.pdf>.
- [Eli10] M. Elish. Exploring the Relationships between Design Metrics and Package Understandability: A Case Study. 2010 [žiūrėta 2024-03-04]. Prieiga per internetą: https://www.researchgate.net/publication/221219583_Exploring_the_Relationships_between_Design_Metrics_and_Package_Understandability_A_Case_Study.
- [Eva03] E. Evans. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison Wesley, 2003.
- [Jan21] M. Jang. Two Different Ways To Package Your Code. 2021 [žiūrėta 2024-04-04]. Prieiga per internetą: <https://medium.com/ryanjang-devnotes/two-different-ways-to-package-your-code-e9cb12d1b6ea>.
- [Kle17] M. Kleppmann. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media, 2017.
- [Mar02] R. C. Martin. *Agile Software Development, Principles, Patterns, and Practices*. Pearson, 2002.
- [Sad21] M. Sadowski. Divide Your Codebase by Domains and Features To Keep It Scalable. 2021 [žiūrėta 2024-05-04]. Prieiga per internetą: <https://betterprogramming.pub/divide-code-by-domains-and-features-and-keep-it-scalable-bb5bd66cf3d2>.
- [Sho19] M. A. Shouki A. Ebad. Investigating The Effect of Software Packaging on Modular Structure Stability. 2019 [žiūrėta 2024-03-04]. Prieiga per internetą: https://d1wqtxts1xzle7.cloudfront.net/75578419/pdf-libre.pdf?1638471828=&response-content-disposition=inline%3B+filename%3DInvestigating_the_Effect_of_Software_Pac.pdf&Expires=1709553794&Signature=BWU2DSSXDNUjfvT1lUWYspcqn1bW1Fgg~doSlc6JoeK7XXJ5bGLPB1B1yBD0tnojJ0yNuWZzQP9fpTjd~yff0hlxnM4GyB2gNMuGZyXsDBXWQuD66kZpwWdluJ63GWjvs28T2ArRpaekqk9JAc-Icun18nyonYJ~W4pIViibjHA4k9yN5r1FZRSWFeUSHpRmflEpJ1opD2Nh889TquLxsA2DhPP3L5_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA.