

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Tina Klančar

**Ocenjevanje napora pri agilnem
razvoju programske opreme -
Sistematični pregled literature**

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM
PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: prof. dr. Viljan Mahnič

Ljubljana, 2017

COPYRIGHT. Rezultati diplomske naloge so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavo in koriščenje rezultatov diplomske naloge je potrebno pisno privoljenje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil L^AT_EX.

Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:
Ocenjevanje napora pri agilnem razvoju programske opreme - Sistematični pregled literature

Tematika naloge:

Izdelajte sistematični pregled znanstvene literature na temo ocenjevanja napora pri agilnem razvoju programske opreme za obdobje od leta 2014 do danes. Pri izdelavi pregleda upoštevajte smernice Guidelines for performing systematic literature reviews in software engineering avtorjev B. Kitchenham in B. Charters ter iskalne kriterije iz članka M. Usman et al., Effort estimation in agile software development: a systematic literature review. Omenjeni članek naj služi tudi kot osnova za klasifikacijo in razlago izbranih študij, tako da bo vaša naloga lahko podala odgovore na ista raziskovalna vprašanja, kot so si jih zastavili Usman in sodelavci. Na koncu izdelajte še primerjavo med študijami, ki so bile objavljene pred letom 2014, in študijami, ki ste jih obdelali v svojem pregledu.

Zahvaljujem se vsem, ki so kakorkoli pripomogli k nastanku tega diplomskega dela. Hvala mentorju prof. dr. Viljanu Mahničju za usmerjanje pri pisanju naloge, neverjetno odzivnost in potrpežljivost. Hvala Mojci za lektoriranje.

Hvala staršem, ki so mi omogočili študij in me vedno podpirali. Hvala bratoma, ki sta me usmerila v področje računalništva in mi dostikrat tako ali drugače priskočila na pomoč. Hvala zelo potrpežljivemu fantu, ki me vedno spodbuja, mi stoji ob strani in verjame vame. Hvala tudi vsem kolegom in prijateljem, ki so mi tekom študija pomagali na kakršenkoli način, še posebej se zahvaljujem Gašperju. Hvala tudi Evi za vse spodbudne besede v trenutkih, ko sem jih najbolj potrebovala.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Metodologija: Sistematični pregled literature	3
2.1	Načrtovanje sistematičnega pregleda	4
2.2	Izvedba sistematičnega pregleda	6
2.3	Poročanje	8
3	Opis predhodnega sistematičnega pregleda	9
3.1	Raziskovalna vprašanja	9
3.2	Kriteriji za izbor študij	10
3.3	Iskanje in izbor ustreznih študij	10
3.4	Glavne ugotovitve	13
4	Pridobivanje primerne literature	15
4.1	Viri podatkov in iskalna strategija	15
4.2	Proces izbiranja	16
4.3	Obrazec za izvleček podatkov	18
5	Rezultati	19
5.1	RV1: Kateri postopki so bili uporabljeni za ocenjevanje napora pri agilnem razvoju programske opreme?	19

5.2	RV2: Katere metrike obsega in glavni faktorji, ki vplivajo na oceno, so bili uporabljeni v študijah?	24
5.3	RV3: Kakšne so značilnosti nabora podatkov, ki so bili uporabljeni v študijah?	26
5.4	RV4: Katere agilne metode so bile predmet raziskav?	28
6	Primerjava ugotovitev s predhodnim sistematičnim pregledom	31
6.1	RV1: Kateri postopki so bili uporabljeni za ocenjevanje napora pri agilnem razvoju programske opreme?	31
6.2	RV2: Katere metrike obsega in glavni faktorji, ki vplivajo na oceno, so bili uporabljeni v študijah?	33
6.3	RV3: Kakšne so značilnosti nabora podatkov, ki so bili uporabljeni v študijah?	34
6.4	RV4: Katere agilne metode so bile predmet raziskav?	34
6.5	Primerjava pomanjkljivosti študij	35
7	Zaključek	37
	Priloga A. Seznam primarnih študij	39
	Literatura	42

Seznam uporabljenih kratic

kratica	angleško	slovensko
PICOC	Population, Intervention, Comparison, Outcome, Context	populacija, intervencija, primerjava, rezultati, kontekst
MMRE	Mean Magnitude of Relative Error	velikost povprečne relativne napake
MRE	Magnitude of Relative Error	velikost dejanske relativne napake
Pred(n)	Percentage of Predictions	delež napovedi, pri katerih je napaka manjša od n
RRSE	Root Relative Squared Error	koren relativne kvadratne napake
RAE	Relative Absolute Error	relativna absolutna napaka
RMSE	Root Mean Squared Error	koren povprečne kvadratne napake
MAE	Mean Absolute Error	povprečna absolutna napaka
MSE	Mean Square Error	povprečna kvadratna napaka
R^2	coefficient of determination	determinacijski koeficient
MER	Magnitude of Error Relative to the Estimate	velikost napovedane relativne napake

Povzetek

Naslov: Ocenjevanje napora pri agilnem razvoju programske opreme - Sistematični pregled literature

Agilne metode uvajajo nove načine ocenjevanja, ki največkrat temeljijo na primerjavi zahtevnosti in ne na uporabi časovnih enot.

V tem diplomskem delu je bil opravljen sistematični pregled literature na temo ocenjevanja napora pri agilnem razvoju programske opreme, ki še vedno predstavlja velik izziv.

Diplomsko delo je narejeno na osnovi sistematičnega pregleda literature od leta 2001 do 2013 in predstavlja nadaljevanje le-tega, saj zajema študije od leta 2014 do maja 2017. Sistematični pregled je prinesel odgovore na štiri glavna zastavljena vprašanja: kateri postopki so bili uporabljeni za ocenjevanje, katere metrike obsega (angl. size metrics) in glavni faktorji, ki vplivajo na oceno (angl. cost drivers), so bili uporabljeni, kakšne so značilnosti nabora podatkov in katere agilne metode so bile predmet raziskav. Opravljena je bila tudi primerjava z zgoraj omenjenim predhodnim sistematičnim pregledom.

Ključne besede: ocenjevanje napora, agilni razvoj programske opreme, sistematični pregled literature.

Abstract

Title: Effort estimation in agile software development - Systematic literature review

Agile methods introduce new effort estimation methods that are most often based on a comparison of effort rather than on the use of time units.

In this thesis, a review of existing scientific literature regarding the effort estimation in agile software development has been made. The effort estimation in agile software development still remains a challenge.

The thesis is based on a systematic literature review of studies between 2001 and 2013 and it is a continuation of the mentioned review covering studies between 2014 and May 2017.

The review brought forth answers to four previously set research questions: what techniques have been used for effort or size estimation in agile software development, what effort predictors (size metrics, cost drivers) have been used, what are the characteristics of the dataset and which agile methods have been investigated.

A comparison to the previously mentioned review was also made.

Keywords: effort estimation, agile software development, systematic literature review.

Poglavje 1

Uvod

Pri tradicionalnem načinu razvoja programske opreme je napor opredeljen kot skupen čas, v katerem člani razvojne ekipe opravijo določeno nalogo. Ta vrednost je pomembna, saj služi kot podlaga za ocenjevanje drugih vrednosti, kot so stroški ali skupen čas, potreben za izdelavo programske opreme [11]. Agilne metode se poslužujejo novih načinov ocenjevanja, ki ne uporabljajo časovnih enot, ampak temeljijo na primerjavi zahtevnosti in uporabi relativnih enot. Te ocene zato ne predstavljajo časa, ki je potreben za dokončanje dela, ampak odražajo obseg in kompleksnost dela. Dekompozicija uporabniških zahtev pri tradicionalnih metodah zahteva veliko napora in časa za ocenjevanje, primerjava zahtevnosti, ki se uporablja pri agilnem pristopu, pa je hitrejša. To sovпада z načeli agilnih metod, saj te zagovarjajo pristope, ki niso časovno zahtevni. Ker gre za nov pristop k ocenjevanju, ki se razlikuje od tradicionalnega, je na tem področju potrebna raziskava.

Agilne metode zagovarjajo zmanjšanje obsega dela, ki ni nujno potrebno. V knjigi Cohna [3] je na temo napora, ki ga vložimo v ocenjevanje, predstavljen graf, ki prikazuje, da se kakovost ocen z vloženim naporom izboljšuje, vendar samo do neke točke. Če za ocenjevanje porabimo preveč časa, se lahko ocene poslabšajo.

Čim bolj točne ocene so pomembne tako za razvijalca kot za stranko. Brez neke začetne ocene ne moremo vedeti, koliko časa in finančnih sredstev

bo potrebno za dokončanje projekta. Ocenjevanje napora ni lahko delo, še posebej pri agilnih metodah, kjer so prisotne stalne spremembe, ki jih je potrebno upoštevati.

Skozi leta je bilo napisanih veliko študij (npr. [8, 2, 7]) na temo ocenjevanja napora pri agilnem razvoju programske opreme. Te se nanašajo na posamezna specifična področja ocenjevanja napora, šele leta 2014 pa je bil izdelan prvi sistematični pregled literature, ki je na enem mestu zbral in združil ugotovitve o tej tematiki. V tem pregledu so bile zajete študije od leta 2001 (ko je bil izdan agilni manifest [1]) do leta 2013.

Cilj diplomske naloge je nadaljevati prvotni sistematični pregled s študijami od leta 2014 (vključno) naprej. V nalogi so zastavljena štiri glavna raziskovalna vprašanja, na katera odgovarjamo z metodo sistematičnega pregleda literature.

V prvotnem pregledu so poročali o nekaterih pomanjkljivostih, ki so jih zabeležili po analizi primarnih študij (točnosti postopkov niso bile navedene ali pa so bile nizke, nestrinjanje glede glavnih faktorjev, ki vplivajo na oceno ter premalo sodelovanja med podjetji). V sklopu naloge nas zanima, ali so te pomanjkljivosti v novejših študijah odpravljene.

Diplomska naloga ima sledečo strukturo. Po začetnem uvodnem poglavju sledi opis uporabljene metodologije: sistematični pregled literature, nato pa predstavitev članka, iz katerega izhajamo. V četrtem poglavju je opisan postopek pridobivanja primerne literature, v petem pa rezultati, kot odgovori na raziskovalna vprašanja. V šestem poglavju primerjamo svoje ugotovitve s tistimi iz prvotnega sistematičnega pregleda. Zadnje, sedmo poglavje je zaključek.

Poglavje 2

Metodologija: Sistematični pregled literature

V našem delu, kot tudi v članku, iz katerega izhajamo, je uporabljena metodologija sistematični pregled literature. Za pravilno uporabo le-te smo se zgledovali po smernicah avtorjev Kitchenham in Charters [6]. V njenem poročilu se pojavita naslednji razlagi sistematičnega pregleda:

- Sistematični pregled literature je sredstvo za vrednotenje in tolmačenje vseh razpoložljivih raziskav, ki so relevantne za določeno raziskovalno vprašanje ali temo.
- Sistematični pregled literature je oblika sekundarne študije, ki uporablja dobro opredeljeno metodologijo za prepoznavanje, analizo in interpretacijo vseh razpoložljivih dokazov, povezanih z določenim raziskovalnim vprašanjem na način, ki je nepristranski in (do določene mere) ponovljiv.

Cilj zgoraj omenjenega poročila je predlagati izčrpne smernice za sistematične preglede literature, primerne za raziskovalce na področju razvoja programske opreme, ki bodo predstavljene v tem poglavju. Smernice zajemajo tri faze sistematičnega pregleda literature:

- načrtovanje

- izvedba
- poročanje

Vsaka izmed faz vključuje več korakov, ki jim je potrebno slediti.

2.1 Načrtovanje sistematičnega pregleda

Faza načrtovanja zajema naslednje korake:

- potreba po sistematičnem pregledu
- naročilo sistematičnega pregleda
- raziskovalna vprašanja
- razvoj protokola za sistematični pregled
- ocenjevanje protokola za sistematični pregled

V nadaljevanju bomo na kratko predstavili vsakega od naštetih korakov.

Potreba po sistematičnem pregledu

Pred izvajanjem sistematičnega pregleda bi morali raziskovalci zagotoviti, da je sistematični pregled potreben. Potreba po sistematičnem pregledu izhaja iz želje raziskovalcev, da temeljito in nepristransko povzamejo vse obstoječe informacije o določenem pojavu na temeljit in nepristranski način. Morda je to potrebno za pridobitev bolj splošnih zaključkov o določenem pojavu, kot jih je mogoče pridobiti iz posameznih študij, ali pa se uporabljajo kot uvod v nadaljnje raziskovalne dejavnosti.

Naročilo sistematičnega pregleda

Včasih organizacija zahteva informacije o določeni temi, vendar nima časa ali strokovnega znanja za izvedbo sistematičnega pregleda. V tem primeru bodo to delo prepustili raziskovalcem. Ko se to zgodi, mora organizacija predložiti naročniški dokument, ki določa zahtevano delo. Ta korak seveda ni obvezen, saj se lahko raziskovalci lotijo dela za lastne potrebe.

Raziskovalna vprašanja

Določanje raziskovalnih vprašanj je najpomembnejši del vsakega sistematičnega pregleda, saj ta vodijo celoten proces. Petticrew in Roberts za postavljanje raziskovalnih vprašanj predlagata uporabo meril PICOC (angl. Population, Intervention, Comparison, Outcome, Context, kar v slovenščini pomeni populacija, intervencija, primerjava, rezultati, kontekst) [9].

- *Populacija* je lahko specifična vloga v razvoju programske opreme (npr. programer, tester), kategorija inženirja (npr. novinec ali izkušen inženir), področje uporabe (npr. sistemi vodenja in nadzora) ali industrijska skupina (npr. telekomunikacijska podjetja ali majhna podjetja za informacijsko tehnologijo).
- *Intervencija* je metodologija/orodje/tehnologija/postopek, ki obravnava določeno temo (npr. tehnologije za izvajanje določenih nalog, kot so specifikacija zahtev, sistemsko testiranje ali ocena stroškov programske opreme).
- *Primerjava* je metodologija/orodje/tehnologija/postopek, s katerim se primerja intervencija.
- *Rezultati* se morajo nanašati na dejavnike, pomembne za izvajalce (npr. izboljšana zanesljivost, zmanjšani proizvodni stroški).
- *Kontekst* pri razvoju programske opreme je domena, v kateri se izvaja primerjava (npr. akademsko okolje ali industrija), udeleženci, ki sodelujejo v študiji (npr. strokovnjaki, akademiki, svetovalci, študenti) in naloge, ki se izvajajo.

Razvoj protokola za sistematični pregled

Protokol za pregled določa metode, ki se bodo uporabljale za izvajanje specifičnega sistematičnega pregleda. Vnaprej določen protokol je potreben za zmanjšanje pristranskosti raziskovalcev. Sestavni deli protokola

vključujejo vse elemente sistematičnega pregleda in nekatere dodatne informacije o načrtovanju.

Ocenjevanje protokola za sistematični pregled

Raziskovalci se morajo dogovoriti o postopku ocenjevanja protokola.

2.2 Izvedba sistematičnega pregleda

Naslednja je faza izvedbe sistematičnega pregleda, ki vključuje korake:

- identifikacija raziskav
- izbor študij
- ocena kakovosti študij
- pridobivanje podatkov
- združitev podatkov

Identifikacija raziskav

Namen tega koraka je, da pridobimo potencialne študije, ki bi lahko vsebovale odgovore na naša raziskovalna vprašanja. Najprej je potrebno definirati iskalno strategijo. Splošni pristop je razčlenitev raziskovalnih vprašanj na posamezne vidike, tj. populacijo, intervencijo, primerjavo, rezultate in kontekst, kot je predhodno opisano v fazi načrtovanja, pri koraku raziskovalna vprašanja. Nato je potrebno pripraviti seznam sinonimov, okrajšav in alternativnih zapisov. Ostale izraze je mogoče pridobiti z upoštevanjem naslovov, ki se uporabljajo v revijah in zbirkah podatkov. Prefinjene iskalne nize lahko nato sestavimo z operatorjema IN ter AND. Primarne študije potem pridobimo tako, da sestavljen iskalni niz vnesemo v različne baze podatkov.

Pojavlja se pristranskost objav, ki se nanaša na težavo, da je bolj verjetna objava pozitivnih rezultatov kot negativnih. Zaradi pristranskosti objav

sistematični pregledi lahko vodijo do sistematične pristranskosti, razen, če je za to posebej poskrbljeno.

Proces izvajanja sistematičnega pregleda literature mora biti dokumentiran, pregleden in ponovljiv (kolikor je to mogoče).

Izbor študij

Namen tega koraka je, da iz vseh pridobljenih študij izločimo tiste, ki za našo raziskavo niso relevantne. Ko so potencialno pomembne študije pridobljene, jih je potrebno oceniti glede na njihovo dejansko uporabnost. Merila za izbiro študij so namenjena opredelitvi tistih primarnih študij, ki zagotavljajo neposredne dokaze o raziskovalnih vprašanjih. Kriterij vključitve in izključitve zato temelji na raziskovalnih vprašanjih. Vključeni so tudi ostali dejavniki, in sicer jezik, datum ter zasnova raziskav.

Izbor študij je večstopenjski proces. Če študije ne izključimo na podlagi naslova in povzetka, je potrebno pridobiti celoten izvod. Povzetki v študijah, ki se nanašajo na informacijske tehnologije in razvoj programske opreme, dostikrat niso dovolj kakovostni, da bi se nanje zanašali, zato je priporočeno, da se pregledajo tudi zaključki.

Včasih raziskovalci opravijo še eno fazo izbirnega postopka, in sicer oceno kakovosti. Ti sezname vključujejo vprašanja, katerih namen je oceniti, v kolikšnem obsegu so študije obravnavale pristranskost in veljavnost.

Pridobivanje podatkov

Cilj tega koraka je oblikovanje obrazcev za hranjenje izvlečkov najpomembnejših podatkov, pridobljenih iz študij. Da bi se zmanjšala možnost pristranskosti, je potrebno obrazce definirati ob določitvi protokola.

V večini primerov je potrebno za vsako študijo zabeležiti niz numeričnih vrednosti. Te so pomembne za vsak poskus povzetka rezultatov primarnih študij in so predpogoj za meta-analizo (uporaba statističnih tehnik za združitev rezultatov primarnih študij). Kadarkoli je mogoče, bi morala obrazce

samostojno izpolniti dva ali več raziskovalcev. Te obrazce je nato potrebno primerjati in rešiti morebitna nesoglasja.

Pomembno je, da v sistematični pregled ne vključimo več študij, v katerih so objavljeni rezultati iste raziskave, saj bi podvojena poročila negativno vplivala na zanesljivost podatkov. Če obstaja več objav iste raziskave, je potrebno uporabiti tisto, ki je najbolj popolna.

Združitev podatkov

Združitev podatkov vključuje zbiranje in povzemanje rezultatov vključenih primarnih študij. Ta je lahko samo opisna ali pa je dopolnjena še s kvantitativnim povzetkom.

2.3 Poročanje

Zadnja je faza poročanja, ki vključuje prikaz rezultatov sistematičnega pregleda in predstavitev le-teh potencialno zainteresiranim osebam. To je možno izvesti na različne načine, in sicer z objavo v revijah, s prospekti, ki vsebujejo povzetek raziskave, z objavo na spletnih straneh ali pa z neposredno komunikacijo z zainteresiranimi. Ponavadi se o sistematičnih pregledih poroča v vsaj dveh formatih:

- v tehničnem poročilu ali v delu doktorske disertacije in
- v članku ali referatu s konference.

Članki in referati imajo ponavadi omejeno dovoljeno velikost, zato je potrebno navesti, kje se lahko dostopa do celotne študije, v primeru, da bi bralec to želel storiti.

Poglavje 3

Opis predhodnega sistematičnega pregleda

Po pojavitvi agilnih metod leta 2001, je bilo izvedenih veliko študij (npr. [8, 2, 7]) na temo ocenjevanja napora pri agilnem razvoju programske opreme. Cilj članka z naslovom "Effort Estimation in Agile Software Development: A Systematic Literature Review" [10], ki ga lahko prevedemo kot "Ocenjevanje napora pri agilnem razvoju programske opreme: Sistematični pregled literature", je bil zajeti te študije in pripraviti podroben pregled področja na to temo. Za to je bila uporabljena metoda sistematičnega pregleda literature, ki se je ravnala po smernicah Kitchenhama in Chartersa [6], ki so predstavljene v drugem poglavju.

3.1 Raziskovalna vprašanja

Zastavljena raziskovalna vprašanja so bila naslednja:

- RV1: Kateri postopki so bili uporabljeni za ocenjevanje napora pri agilnem razvoju programske opreme?
 - RV1a: Na kakšen način je bila ovrednotena točnost ocen?
 - RV1b: Kako točne so bile ocene, pridobljene s temi postopki?

- RV2: Katere metrike obsega (angl. size metrics) in glavni faktorji, ki vplivajo na oceno (angl. cost drivers), so bili uporabljeni v študijah?
- RV3: Kakšne so značilnosti nabora podatkov, ki so bili uporabljeni v študijah?
- RV4: Katere agilne metode so bile predmet raziskav?
 - RV4a: Katere aktivnosti razvoja so bile raziskane?
 - RV4b: Kateri nivoji planiranja so bili raziskani?

3.2 Kriteriji za izbor študij

Glede na raziskovalna vprašanja je kriterij PICOC naslednji:

- Populacija (P): Aplikacije in projekti razvoja programske opreme, razviti z uporabo agilnih metod.
- Intervencija (I): Metode/tehnike/modeli/metrike velikosti ocenjevanja napora/glavni faktorji, ki vplivajo na ocenjevanje napora.
- Primerjava (C): /
- Rezultati (O): Točnost tehnik za ocenjevanje pri agilnem razvoju programske opreme.
- Kontekst (C): Agilni razvoj programske opreme.

3.3 Iskanje in izbor ustreznih študij

Iskanje se je začelo s ključnimi izrazi in njihovimi sopomenkami iz PICOC kriterija. Začetni iskalni niz je bil sestavljen iz zbranih izrazov in operatorjev Boolove algebre IN ter ALI. Ta je bil vnesen v iskalnike (Scopus, IEEE Explore, Science Direct), da se je preverila ustreznost nizov. Tudi ključne besede iz že znanih študij in ključne besede iz novo najdenih študij, so bile

vkjučene v iskalni niz. Končni niz, ki je moral biti prilagojen različnim bazam, je naslednji:

(Agile OR "extreme programming" OR "Scrum" OR "feature driven development" OR "dynamic systems development method" OR "crystal software development" OR "crystal methodology" OR "adaptive software development" OR "lean software development") AND (estimat OR predict* OR forecast* OR calculat* OR assessment OR measur* OR sizing) AND (effort OR resource OR cost OR size OR metric OR user story OR velocity) AND (software)*

Del primarne iskalne strategije je bil vnos iskalnega niza v različne digitalne knjižnice, in sicer: Scopus, IEEE Explore, EI Compendex, Web of Science, INSPEC, Science Direct in ACM DL. Iskanje je bilo omejeno z letom 2001, saj je bil takrat izdan agilni manifest [1]. Po izločitvi duplikatov, je bilo rezultatov 443.

Glede na raziskovalna vprašanja je bil oblikovan kriterij vključitve in izključitve, ki je bil uporabljen za izbiranje ustreznih študij.

Kriterij vključitve

Študije, ki

- so povezane z ocenjevanjem napora ali velikosti (tehnika ali model ali metrika ali napovedi) IN
- temeljijo na katerikoli agilni metodi IN
- so napisane v angleškem jeziku IN
- so objavljene v znanstvenih revijah ali na konferencah ali na strokovnih delavnicah IN
- temeljijo na dokazih

bodo vključene kot primarne študije.

Kriterij izključitve

Študije, ki

- niso povezane z ocenjevanjem napora ali velikosti ALI
- ne temeljijo na agilnih metodah ALI
- niso napisane v angleškem jeziku ALI
- niso objavljene v znanstvenih revijah ali na konferencah ali na strokovnih delavnicah ALI
- ne temeljijo na dokazih ALI
- obravnavajo samo fazo vzdrževanja programske opreme ALI
- obravnavajo samo merjenje hitrosti (angl. velocity)

ne bodo vključene kot primarne študije.

Nadaljnji izbor študij je potekal v dveh korakih. V prvem koraku je bil kriterij vključitve in izključitve uporabljen na naslovih in povzetkih vseh študij. Ta korak so ločeno izvedli trije raziskovalci. Ti so našli študije, ki so ustrezali kriteriju vključitve. Nekatere študije so bile označene kot take, ki bi mogoče ustrezale. O teh so se raziskovalci posvetovali in sledil je ponoven, bolj podroben pregled. S tem korakom se je število študij znatno zmanjšalo, in sicer na 32. V naslednjem koraku je bil kriterij vključitve in izključitve uporabljen na celotnih besedilih študij. Raziskovalci so svoje rezultate primerjali in se nato odločili za primerne študije.

Ocena kakovosti je bila izvedena s seznamom trinajstih vprašanj, na katera je bilo možno odgovoriti z odgovori, ki so bili točkovani: "Da" (1 točka), "Ne" (0 točk) in "Delno" (0.5 točk). Če je študija dosegla manj kot 3.25 točk, ni bila vključena v nabor izbranih primarnih študij.

Končen nabor študij je bil nato uporabljen za analizo in pridobivanje odgovorov na zastavljena raziskovalna vprašanja. Vsako od vprašanj je bilo predstavljeno s tabelo, v kateri je zapisano nekaj pogostih odgovorov, v katerih študijah se ti odgovori pojavljajo in število teh študij.

3.4 Glavne ugotovitve

V sistematičnem pregledu so ugotovili, da večina uporabljenih postopkov za ocenjevanje napora temelji na subjektivni strokovni oceni. Ti postopki so poker planiranje (angl. planning poker), ocena eksperta (angl. expert judgment) in metoda točk primerov uporabe (angl. use case points method).

Večina ocen, pridobljenih s temi postopki, ni dosegla dobre točnosti.

Najbolj pogoste metrike obsega so bile točke uporabniških zgodb (angl. user story points) in točke primerov uporabe (angl. use case points).

Glede najbolj primernih glavnih faktorjev, ki vplivajo na oceno, študije niso bile enotne.

Najbolj pogosto so študije izvedene v industrijskem okolju, in sicer znotraj enega podjetja.

Agilni metodi, ki sta bili predmet raziskav, sta bili Scrum ali ekstremno programiranje.

V pregledu je navedenih tudi nekaj pomanjkljivosti oziroma premalo raziskana področja, ki so jih zabeležili po analizi primarnih študij:

- točnosti postopkov niso bile navedene ali pa so bile nizke,
- nestrinjanje glede glavnih faktorjev, ki vplivajo na oceno in
- premalo sodelovanja med podjetji.

Poglavje 4

Pridobivanje primerne literature

4.1 Viri podatkov in iskalna strategija

V okviru naloge nas zanimajo odgovori na ista raziskovalna vprašanja, kot so bila zastavljena v prvotnem sistematičnem pregledu, zato je isti tudi iskalni niz, ki smo ga uporabili.

(Agile OR "extreme programming" OR "Scrum" OR "feature driven development" OR "dynamic systems development method" OR "crystal software development" OR "crystal methodology" OR "adaptive software development" OR "lean software development") AND (estimat OR predict* OR forecast* OR calculat* OR assessment OR measur* OR sizing) AND (effort OR resource OR cost OR size OR metric OR user story OR velocity) AND (software)*

Vnesen je bil v več digitalnih knjižnic, z namenom pridobitve študij, ki bi podale odgovore na zastavljena raziskovalna vprašanja. Iskanje je bilo dodatno omejeno, saj so nas zanimale le študije od leta 2014 (vključno) naprej. Baze podatkov, ki smo jih uporabili, so naslednje:

- Scopus (<https://www.scopus.com/>)
- IEEE Xplore (<http://ieeexplore.ieee.org/Xplore/home.jsp>)
- Web of Science (<https://apps.webofknowledge.com/>)
- Science Direct (<http://www.sciencedirect.com>)
- ACM DL (<http://dl.acm.org/>)
- Springer Link (<https://link.springer.com/>)

Iskalni niz je moral biti za nekatere baze prilagojen (npr. pri IEEE Xplore je največje dovoljeno število izrazov 15).

4.2 Proces izbiranja

Proces izbiranja študij smo izvedli v petih korakih, in sicer na podoben način, kot je opisano v originalnem sistematičnem pregledu, ki je predstavljen v tretjem poglavju.

V prvem koraku, ko še nismo izločili nobene od študij, smo skupno dobili 784 zadetkov. V te zadetke so vključene študije od leta 2014 do maja 2017.

V drugem koraku smo glede na naslove izločili veliko študij, ki za nas niso bile relevantne. Pri tem smo si pomagali s kriterijem vključitve in izključitve, ki je opisan v tretjem poglavju. Tako smo tu, kolikor se je dalo razbrati iz naslovov, izključili študije, ki niso bile v angleščini, se niso osredotočale na agilne metode ali so ustrezale kateremu drugemu delu kriterija izključitve. Za pregled je po tem koraku ostalo 60 študij.

V vmesnem, tretjem koraku smo samo izločili študije, ki so se pojavile v več bazah. Skupno je ostalo 40 študij.

V četrtem koraku smo izločili študije na podlagi povzetka in zaključka, kot je predlagano v smernicah [6], ki so opisane v drugem poglavju. Pri tem smo spet upoštevali kriterij vključitve in izključitve. V tem koraku je bilo izbranih 18 študij, ki jih je bilo potrebno v zadnjem koraku v celoti prebrati.

Glede na vsebino smo potem izbrali končno število študij, in sicer 12. Te predstavljajo primarne študije, ki smo jih nato analizirali.

Tabela 4.1 prikazuje koliko študij smo pridobili na začetku in koliko jih je ostalo po posameznem izvedenem koraku.

Baza	1. korak	2. korak	3. korak	4. korak	5. korak
Scopus	301	16	16	7	3
IEEE Xplore	73	20	15	9	7
Web of Science	179	15	5	1	1
Science direct	74	3	1	1	1
ACM DL	52	3	2	0	0
Springer Link	105	3	1	0	0
Skupno	784	60	40	18	12

Tabela 4.1: Proces izbiranja študij

Izbrane študije so glede na predhodno postavljene kriterije, članki in referati s strokovnih konferenc ali delavnic. Večina študij, in sicer kar tri četrtine, spada v kategorijo referatov s konferenc ali delavnic. Ostale študije pa so članki, kot lahko vidimo v tabeli 4.2.

Vrsta študije	Število študij	Delež
Članek	3	25%
Referat s konference	8	67%
Referat z delavnice	1	8%

Tabela 4.2: Delitev študij glede na izvor

Seznam izbranih primarnih študij se nahaja v prilogi A. Na njih se sklicujemo z veliko črko Š in številko študije (npr. [Š1]).

4.3 Obrazec za izvleček podatkov

Po izboru primarnih študij, je potrebno le-te ustrezno analizirati. Za hranjenje izvlečkov najpomembnejših podatkov smo pripravili ustrezen obrazec. Ker nas je zanimala primerjava s predhodnim sistematičnim pregledom, smo ohranili ista raziskovalna vprašanja in sestavili temu primeren obrazec, ki sledi tem vprašanjem. Kot primer je za študijo [Š3] priložen izpolnjen obrazec v obliki tabele 4.3. Na enak način smo vnesli podatke iz drugih študij. Ko smo imeli podatke ustrezno zabeležene, smo jih združili v rezultate, ki so predstavljeni v naslednjem poglavju.

RV1	Uporabljen postopek za ocenjevanje napora	<i>Različice poker planiranja (kombinacija poker planiranja in algoritmov strojnega učenja)</i>
RV1a	Način vrednotenja točnosti ocen	<i>MMRE, točnost (delež pravilno klasificiranih primerov)</i>
RV1b	Dosežena točnost uporabljenih postopkov	<i>MMRE: 91.75%, točnost: 31%</i>
RV2a	Metrike obsega	<i>točke uporabniških zgodb (angl. user story points)</i>
RV2b	Glavni faktorji, ki vplivajo na oceno	<i>ključne besede na karticah uporabniških zgodb</i>
RV3a	Domena, v kateri je bila izvedena študija	<i>industrijska</i>
RV3b	Vrsta domene (znotraj podjetja ali več podjetij)	<i>znotraj podjetja</i>
RV4	Agilna metoda, ki je bila predmet raziskave	<i>ni opisano</i>
RV4a	Aktivnost razvoja	<i>implementacija</i>
RV4b	Nivo planiranja	<i>ni opisano</i>

Tabela 4.3: Primer izpolnjenega obrazca

Poglavje 5

Rezultati

Rezultati so zbrani in prikazani v tabelah, v katerih stolpec z oznako P označuje število študij, stolpec z oznako D pa delež študij (v odstotkih) z določeno lastnostjo.

5.1 RV1: Kateri postopki so bili uporabljeni za ocenjevanje napora pri agilnem razvoju programske opreme?

Za ocenjevanje napora so bili uporabljeni različni postopki, ki so navedeni v tabeli 5.1. V vseh primarnih študijah je bil naveden in opisan uporabljen postopek. V treh od dvanajstih študij [Š1, Š4, Š6] je bil predstavljen neki novi algoritem, ki so ga avtorji predlagali za ocenjevanje napora. Opazimo, da si avtorji večine študij prizadevajo poiskati nove programske podprte metode. To ne pomeni, da se sedaj v praksi ne uporablja več postopkov za ocenjevanje napora, ki temeljijo na subjektivni strokovni oceni (poker planiranje (angl. planning poker), ocena eksperta (angl. expert judgment) in metoda točk primerov uporabe (angl. use case point method)), ampak, da je bilo o njih že veliko napisanega in je sedaj pozornost pri raziskovanju in pisanju študij usmerjena predvsem v iskanje novih metod, ki bi zmanjšale potrebo

po ekspertnem znanju.

Dve od treh študij, v katerih predlagajo neki novi algoritem, imata iste avtorje. Ti v prvotni študiji [Š1] predstavljajo algoritem, katerega ocena temelji na faktorjih, ki so povezani s projektom in faktorjih, ki so povezani z ljudmi. V njihovi naslednji študiji [Š6] je dodana še ena skupina faktorjev.

Različice poker planiranja se pojavijo v dveh študijah, vendar na povsem drugačen način. V eni [Š3] je poker planiranje kombinirano z algoritmi strojnega učenja, v drugi [Š10] pa je ocenjevanje pri poker planiranju, poleg točk uporabniških zgodb (angl. user story points), dopolnjeno s točkami vrednosti (angl. value points).

Postopek	Študije	P	D
Nov algoritem za ocenjevanje napora	Š1, Š4, Š6	3	25%
Nevronske mreže	Š5, Š11	2	17%
Različice poker planiranja	Š3, Š10	2	17%
Bayesovske mreže	Š2, Š12	2	17%
Ostali	Š9 (metoda točk primerov uporabe (angl. use case points method)), Š7 (regresijski modeli), Š8 (ocena eksperta (angl. expert judgment))	3	25%

Tabela 5.1: Postopki za ocenjevanje napora

5.1.1 RV1a: Na kakšen način je bila ovrednotena točnost ocen?

V tabeli 5.2 opazimo, da je najbolj pogosto uporabljena metrika MMRE, in sicer v štirih od dvanajstih študij. Presenetljivo je, da se v več kot polovici študij ni analiziralo točnosti ocen uporabljenega postopka. Seštevek v stolpcu

P ni enak številu študij, ki smo jih zajeli, saj je bilo v nekaterih študijah uporabljenih več metrik točnosti.

Metrika	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
MMRE	Š2, Š3, Š5, Š7	5	20%	4	33%
MRE	/	5	20%	0	0%
Pred(n)	Š2, Š5, Š7	2	8%	3	25%
Nobene	Š1, Š4, Š6, Š8, Š9, Š10, Š12	6	24%	7	58%
Ostale	Š2 (RRSE, RAE, RMSE, MAE, točnost (delež pravilno klasificiranih primerov)), Š5 (MSE, R^2), Š11 (MER)	10	40%	3	25%

Tabela 5.2: Metrike za vrednotenje točnosti postopkov ocenjevanja napora

5.1.2 RV1b: Kako točne so bile ocene, pridobljene s temi postopki?

V tabeli 5.3 so navedeni uporabljeni postopki za ocenjevanje napora in točnost ocen, pridobljenih s temi postopki. Prikazani so le tisti postopki, za katere so bili vsaj v eni izmed študij, navedeni rezultati o doseženi točnosti. Da je postopek ocenjevanja sprejemljiv, naj bi dosegel vrednost MMRE, ki je manjša ali enaka 0.25 in/ali vrednost PRED(0.25), ki je večja ali enaka 0.75 [4, 5]. Primerjava rezultatov je rahlo otežena, saj so pridobljene vrednosti za MMRE iz različnih študij nekonsistentne, kot lahko opazimo v tabeli 5.3.

V vseh študijah, kjer so bili predlagani novi algoritmi za ocenjevanje napora [Š1, Š4, Š6], ni bilo navedeno, kako točni so bili.

V študiji [Š3], kjer so predlagane različice poker planiranja (angl. planning poker), so bile dosežene točnosti v splošnem slabe. Najboljša različica, kombinacija algoritma J48 in poker planiranja, je imela MMRE 91.75% in točnost (delež pravilno klasificiranih primerov) 31%, kar pa je boljše od "navadnega" poker planiranja, kjer je bila MMRE 106.81% in točnost 24.5%. Obe vrednosti MMRE očitno nista v skladu z zgoraj omenjenim kriterijem, saj sta veliko večji od 25%.

Pri vseh ostalih postopkih so rezultati veliko boljši.

Uporaba Bayesovskih mrež, opisanih v študiji [Š2], je dosegla točnost (delež pravilno klasificiranih primerov) 99.375%, MMRE 6.21 in PRED(25%) 100%. Slednja ustreza zgoraj omenjenemu kriteriju, da mora biti vrednost PRED(0.25) večja ali enaka 0.75. Vsi rezultati, razen MMRE, kažejo na dobro doseženo točnost. Tu opazimo zgoraj omenjeno nekonsistentnost rezultatov, saj bi bila bolj smiselna vrednost za MMRE v tem primeru 6.21% (tudi glede na študijo, ki pravi, da izmerjena MMRE predstavlja dobre rezultate). Postopek, opisan v tej študiji, se uporablja v podjetju, kjer ga projektni vodja smatra za izredno uporabnega.

Podobna MMRE, in sicer 6.28, je bila izmerjena tudi v enem izmed regresijskih modelov, opisanih v študiji [Š7]. Tudi tu se pojavlja isto vprašanje glede navedene vrednosti kot prej, saj so rezultati v študiji predstavljeni kot

dobri, kar kaže tudi vrednost PRED(30).

Pri nevronskih mrežah iz študije [Š5] del rezultatov (dve od štirih vrst predlaganih nevronskih mrež) po točnosti ustreza zgoraj omenjenemu kriteriju sprejemljivosti (vrednost MMRE je manjša ali enaka 0.25). Eden ima vrednost 0.15, drugi pa 0.16, kar predstavlja dobre rezultate.

Študij, ki so navedle doseženo točnost ocen, je bilo malo, tako da pomanjkanje študij, ki merijo to točnost, še vedno predstavlja vrzel na tem področju.

Postopek	Točnost
Različice poker planiranja (Š3)	MMRE: 91.75% Točnost (delež pravilno klasificiranih primerov): 31%
Bayesovske mreže (Š2)	MMRE: 6.21 MAE: 0.026 RMSE: 0.065 RAE: 9.71% RRSE: 17.81% Pred(25)%: 100% Pred(10)%: 100% Točnost (delež pravilno klasificiranih primerov): 99.375%
Nevronske mreže	MSE: 0.0059-0.0244 (Š5) R^2 : 0.6259-0.9303 (Š5) MMRE: 0.1486-0.3581 (Š5) PRED(%) 85.9182-94.7649 (Š5) MER: 12% (Š11)
Regressijski modeli (Š7)	MMRE: 6.28-23.14 PRED(30): 66.67-100

Tabela 5.3: Točnost ocen

5.2 RV2: Katere metrike obsega in glavni faktorji, ki vplivajo na oceno, so bili uporabljeni v študijah?

5.2.1 Metrike obsega

V večini primerov, in sicer devetih od dvanajstih, je uporabljena metrika obsega točke uporabniških zgodb (angl. user story points). Opazimo, da so v študijah zahteve pogosto podane v obliki uporabniških zgodb, za katere je največkrat uporabljena prej omenjena metrika. Zelo redko se v študijah uporabijo druge metrike, in sicer točke primerov uporabe (angl. use case points), funkcijske točke (angl. function points) in ostale, ki so navedene v tabeli 5.4.

5.2.2 Glavni faktorji, ki vplivajo na oceno (angl. cost drivers)

Glavni faktorji, ki vplivajo na oceno in se pojavijo v več študijah, so kompleksnost zahtev in razvoja ter tehnične sposobnosti razvijalcev, kot lahko vidimo v tabeli 5.5. Oba faktorja se pojavita v tretjini izbranih študij. V četrtini študij se kot eden izmed glavnih faktorjev pojavi poznavanje projekta in nalog. V treh študijah faktorji, ki vplivajo na oceno, niso bili opisani.

V dveh študijah [Š1, Š6] so predlagali nove algoritme, ki temeljijo na izbranih faktorjih, ki vplivajo na oceno. Razdelili so jih na faktorje, povezane s projektom in faktorje, povezane z ljudmi, kot lahko vidimo v tabeli 5.5. Ugotovili so, da je za projekt potrebno manj časa, če so faktorji, povezani z ljudmi (komunikacijske sposobnosti, sposobnosti vodenja, poznanstva v ekipi), dobri. V nadgradnji [Š6] prvotne študije [Š1] je bilo za izboljšanje ocen predlagano še nekaj novih faktorjev (sestava ekipe, neudobje na delovnem mestu, prehod v agilnost, dinamika v ekipi, spremembe v ekipi, predstavitev nove tehnologije, uporabnost, pomanjkanje jasnosti v zahtevah, ne-

Metrika	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
Točke uporabniških zgodb (angl. user story points)	Š1, Š3, Š4, Š5, Š6, Š7, Š8, Š10, Š11	6	24%	9	75%
Točke primerov uporabe (angl. Use Case Points)	Š9	7	28%	1	8%
Funkcijske točke (angl. Function Points)	Š7	3	12%	1	8%
Ostale	Š2 (dejanske ure dela), Š8 (ure dela na osebo), Š10 (točke vrednosti)	3	12%	3	25%
Ni opisano	Š12	5	20%	1	8%

Tabela 5.4: Metrike obsega

stabilnost zahtev, spremembe v delovnem okolju). Vsi faktorji skupaj so bili nato upoštevani v novem algoritmu.

Poleg faktorjev, navedenih v tabeli 5.5, so ostali faktorji, ki se še pojavijo v študijah:

- Š3 (ključne besede, navedene na karticah z uporabniškimi zgodbami)
- Š2 (kakovost specifikacije)
- Š4 (različne lokacije, tveganja (vprašanja pri integraciji, vprašanja pri upravljanju virov), ostali faktorji (komunikacija v ekipi v različnih

Faktorji	Študije	P	D
Kompleksnost (razvoj, zahteve)	Š1, Š2, Š4, Š8	4	33%
Poznavanje projekta in nalog	Š8, Š9, Š11	3	25%
Tehnične sposobnosti razvijalcev	Š1, Š2, Š6, Š8	4	33%
Faktorji, povezani s projektom (tip projekta, kompleksnost, kakovost zahtev, potreba po programski in strojni opremi, zahtevnost operacije), faktorji, povezani z ljudmi (komunikacijske sposobnosti, poznanstva v ekipi, sposobnosti vodenja, varnost, delovni čas, izkušnje iz prejšnjih projektov, tehnične sposobnosti)	Š1, Š6	2	17%
Podpora različnim platformam	Š9, Š11	2	17%
Nestabilnost zahtev	Š4, Š6	2	17%
Ni opisano	Š5, Š7, Š10	3	25%

Tabela 5.5: Glavni faktorji, ki vplivajo na oceno

časovnih pasovih, katere člani govorijo različne jezike, poznavanje znotraj ekipe, pretekle izkušnje z agilnimi projekti, tehnične in vodstvene sposobnosti ekipe))

Opazimo, da je naštetih veliko različnih faktorjev, saj študije niso enotne glede najbolj primernih glavnih faktorjev, ki vplivajo na oceno.

5.3 RV3: Kakšne so značilnosti nabora podatkov, ki so bili uporabljeni v študijah?

Da bi odgovorili na tretje raziskovalno vprašanje, nas je v izbranih študijah zanimalo, v katerem okolju (industrijskem ali akademskem) so bile izvedene.

V primeru, da je bila študija izvedena v industrijskem prostoru, nas je zanimalo še, ali so bili pri tem uporabljeni podatki zajeti znotraj podjetja ali so podatki prišli iz več različnih podjetij. V tabeli 5.6 opazimo, da je bila večina študij, in sicer osem od dvanajstih, izvedena v industrijskem okolju. Ker so bile študije izvedene v podjetjih, torej na realnih podatkih, so te bolj uporabne za nadaljnje raziskave.

Postopek, predlagan v [Š2] se v podjetju, kjer je bila študija izvedena, tudi dejansko uporablja.

Domena	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
Industrijska	Š2, Š3, Š5, Š7, Š8, Š9, Š10, Š11	19	76%	8	67%
Akademska	Š12	3	12%	1	8%
Ni opisano	Š1, Š4, Š6	3	12%	3	25%

Tabela 5.6: Domena, v kateri so bile izvedene študije

V tabeli 5.7 bolj podrobno vidimo, kako so bile porazdeljene študije v industrijskem okolju. Skoraj vse, sedem od osmih so bile izvedene znotraj enega podjetja, samo v eni pa so bili uporabljeni podatki iz več različnih podjetij.

Vrsta	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
Znotraj podjetja	Š2, Š3, Š7, Š8, Š9, Š10, Š11	18	90%	7	87.5%
Več podjetij	Š5	2	10%	1	12.5%

Tabela 5.7: Porazdelitev študij znotraj industrijske domene

5.4 RV4: Katere agilne metode so bile predmet raziskav?

V tabeli 5.8 opazimo, da na žalost pri 58% študijah ni navedeno, katera agilna metoda je bila uporabljena. V ostalih primerih je bila uporabljena priljubljena agilna metoda Scrum. V študiji [Š8] je bilo poleg ocenjevanja napora pri metodi Scrum raziskano še ocenjevanje pri ekstremnem programiranju.

Študija [Š2] opisuje predlagan postopek na primeru Scruma, vendar navaja, da je uporaben tudi za druge agilne metode.

Opazimo, da se je uporaba metode Scrum povečala, uporaba ekstremnega programiranja pa zmanjšala. Ti rezultati so smiselni, saj je v zadnjih letih Scrum najbolj priljubljena in uporabljena agilna metoda, ekstremno programiranje pa ni več tako aktualno kot včasih.

Agilna metoda	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
XP	Š8	7	28%	1	8%
Scrum	Š2, Š7, Š8, Š10, Š12	8	32%	5	42%
Ni opisano	Š1, Š3, Š4, Š5, Š6, Š9, Š11	10	40%	7	58%

Tabela 5.8: Agilne metode, ki so bile predmet raziskav

5.4.1 RV4a: Katere aktivnosti razvoja so bile raziskane?

Iz tabele 5.9 je razvidno, da so bile v več kot polovici študij raziskane vse aktivnosti razvoja, in sicer analiza, načrtovanje, implementacija in testiranje. Zgolj v dveh od dvanajstih študij je bila navedena samo ena izmed omenjenih aktivnosti iz tabele 5.9, in sicer implementacija.

Aktivnost	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
Analiza	/	0	0%	0	0%
Načrtovanje	/	0	0%	0	0%
Implementacija	Š3, Š9	3	12%	2	17%
Testiranje	/	4	16%	0	0%
Implementacija in testiranje	/	3	12%	0	0%
Vse	Š1, Š2, Š5, Š6, Š7, Š10, Š11	9	36%	7	58%
Ni opisano	Š4, Š8, Š12	6	24%	3	25%

Tabela 5.9: Raziskane aktivnosti razvoja

5.4.2 RV4b: Kateri nivoji planiranja so bili raziskani?

V agilnem kontekstu se planiranje lahko izvaja na treh nivojih, in sicer na nivoju izdaje, iteracije in na dnevnem nivoju [3]. V nobeni izmed študij ni bil raziskan dnevni nivo planiranja. V tretjini primerov je bilo raziskano ocenjevanje napora na nivoju izdaje in v četrtni primerov na nivoju iteracije in izdaje, kot lahko vidimo v tabeli 5.10.

Nivo planiranja	Študije	Prvotni pregled		Naš pregled	
		P	D	P	D
Dnevno planiranje	/	0	0%	0	0%
Iteracija	Š2, Š8	6	24%	2	17%
Izdaja	Š4, Š5, Š6, Š9	5	20%	4	33%
Iteracija in izdaja	Š1, Š10, Š11	2	8%	3	25%
Ni opisano	Š3, Š7, Š12	12	48%	3	25%

Tabela 5.10: Raziskani nivoji planiranja

Poglavje 6

Primerjava ugotovitev s predhodnim sistematičnim pregledom

6.1 RV1: Kateri postopki so bili uporabljeni za ocenjevanje napora pri agilnem razvoju programske opreme?

V prvotnem sistematičnem pregledu je postopek, ki je uporabljen v največ študijah (tretjini), ocena eksperta (angl. expert judgment). V naših rezultatih opazimo, da je ocena eksperta uporabljena samo v eni izmed izbranih študij [Š8]. Največji delež (četrtnina) je novih predlaganih algoritmov za ocenjevanje napora [Š1, Š4, Š6]. Kot smo že omenili, to ne pomeni, da se v praksi postopki za ocenjevanje napora, ki temeljijo na subjektivni strokovni oceni (poker planiranje (angl. planning poker), ocena eksperta (angl. expert judgment) in metoda točk primerov uporabe (angl. use case point method)), ne uporabljajo več, ampak, da je bilo o njih že veliko napisanega in je sedaj pozornost pri raziskovanju in pisanju študij usmerjena predvsem v odkrivanje novih postopkov za čim bolj točno ocenjevanje napora.

6.1.1 RV1a: Na kakšen način je bila ovrednotena točnost ocen?

Najbolj pogosto uporabljeni metriki točnosti sta v prvotnem sistematičnem pregledu MMRE in MRE. V naši analizi opazimo, da MRE ni bil uporabljen v nobeni izmed študij, MMRE pa je bila uporabljena v tretjini študij. Že prej je bilo veliko študij, kjer točnost ni bila izmerjena (24%), sedaj pa je takih študij še veliko več, in sicer kar 58%. Ta pomanjkljivost je bila zabeležena že pri predhodnem sistematičnem pregledu, in kot vidimo, se je stanje še poslabšalo.

6.1.2 RV1b: Kako točne so bile ocene, pridobljene s temi postopki?

V tistih študijah, ki navajajo točnost, je bila ta praviloma boljša kot v prvotnem sistematičnem pregledu. Kriteriju, ki pravi, da je postopek ocenjevanja sprejemljiv, če je dosežena vrednost MMRE, ki je manjša ali enaka 0.25 in/ali vrednost PRED(0.25), ki je večja ali enaka 0.75 [4, 5], v prvotnem pregledu ustrezata dva postopka (ocena eksperta (angl. expert judgment) in metoda točk primerov uporabe (angl. use case point method)). V naših študijah točnost za postopka ocena eksperta in metoda točk primerov uporabe ni bila navedena, tako da jih ne moremo primerjati.

Uporabljena modela (linearna regresija in nevronske mreže) v prvotnem pregledu dosegata slabe rezultate, kar je v nasprotju z našimi rezultati. Vsi modeli, ki so bili raziskani v naših študijah in so imeli navedeno točnost, so ustrezali zgoraj omenjenemu kriteriju, kar pomeni, da so bili rezultati dobri.

6.2 RV2: Katere metrike obsega in glavni faktorji, ki vplivajo na oceno, so bili uporabljeni v študijah?

6.2.1 Metrike obsega

V prvotnem sistematičnem pregledu sta bili najbolj pogosto uporabljeni metrike obsega točke primerov uporabe (angl. use case points) - v 28% primerih in točke uporabniških zgodb (angl. user story points) - v 24% primerih. Sedaj prevladuje uporaba točk uporabniških zgodb (75%). Opazimo, da so v študijah zahteve pogosto podane v obliki uporabniških zgodb, za katere je največkrat uporabljena prej omenjena metrika.

Samo v eni izmed naših študij uporabljena metrika ni bila opisana, v originalnem članku pa je bilo takih študij petina.

6.2.2 Glavni faktorji, ki vplivajo na oceno (angl. cost drivers)

Glavni faktorji, ki vplivajo na oceno in so se pojavili v več študijah v prvotnem pregledu, so bili velikost naloge, predhodne izkušnje ekipe in sposobnosti razvijalcev. Slednji se je pojavil tudi pri zdajšnji analizi. Poleg omenjenih sta bila pogosta še poznavanje projekta in kompleksnost zahtev in razvoja. Kot v predhodnem pregledu tudi tu študije niso ravno enotne glede primernih glavnih faktorjev, ki vplivajo na oceno. Pri obeh pregledih je bilo v izbranih študijah navedenih veliko različnih faktorjev.

V četrtini naših izbranih študij faktorji niso bili opisani, v prejšnjem pregledu pa v več kot tretjini študij.

6.3 RV3: Kakšne so značilnosti nabora podatkov, ki so bili uporabljeni v študijah?

Kar se tiče značilnosti nabora podatkov, torej domena (akademska ali industrijska (znotraj podjetja ali več vključenih podjetij)), so rezultati obeh pregledov zelo podobni. Večino študij je bilo izvedenih v industrijskem okolju, in sicer znotraj podjetja. V predhodnem pregledu je bila ena izmed pomanjkljivosti ta, da je bilo malo študij, kjer bi sodelovalo več podjetij, kar opazimo tudi v naši analizi.

6.4 RV4: Katere agilne metode so bile predmet raziskav?

Pri obeh analizah v večini študij ni bilo opisano, katere agilne metode so bile predmet raziskav. V prvotnem sistematičnem pregledu je bilo takih primerov 40%, v naši analizi pa kar 58%. V ostalih študijah, ki smo jih izbrali za naš sistematični pregled, je bila raziskana metoda Scrum (42% primerov). V eni od teh študij, kar predstavlja 8% primerov, je bilo predmet raziskave poleg metode Scrum tudi ekstremno programiranje. V prvotnem sistematičnem pregledu je bila v 32% študij uporabljena metoda Scrum, v ostalih 28% pa ekstremno programiranje. V splošnem se delež uporabe ekstremnega programiranja niža, Scrum pa je že vrsto let najbolj priljubljena agilna metoda, kar je razvidno tudi iz naših rezultatov.

6.4.1 RV4a: Katere aktivnosti razvoja so bile raziskane?

Ocenjevanje napora v naših študijah se je nanašalo na implementacijo ali pa vse aktivnosti razvoja, in sicer analiza, načrtovanje, implementacija in testiranje. V prvotnem pregledu je bilo tudi nekaj študij, ki so se osredotočale samo na testiranje ali na implementacijo in testiranje. Pri obeh analizah v

25% primerov ni bilo opisano, na katero aktivnost se nanaša ocenjevanje.

6.4.2 RV4b: Kateri nivoji planiranja so bili raziskani?

Nivo planiranja, ki je bil v prvotnem pregledu največkrat raziskan, je bil iteracija, v naši analizi pa izdaja. V obeh primerih dnevno planiranje ni bilo omenjeno v nobeni izmed študij.

6.5 Primerjava pomanjkljivosti študij

Že pri primerjavi ugotovitev posameznih raziskovalnih vprašanjih smo navedli, katere pomanjkljivosti oziroma premalo raziskana področja so se pojavila v prvotnem sistematičnem pregledu in katera v naši analizi. Če povzamemo te ugotovitve, opazimo, da večina slabosti študij iz prejšnjega pregleda ni odpravljenih. To se vidi predvsem v deležu študij, ki niso navedle točnosti ocen uporabljenih postopkov, saj se je v našem pregledu pojavilo še več takih, ki točnosti ne navajajo. So pa tiste točnosti ocen, ki so navedene, v primerjavi s prejšnjim pregledom, dosti boljše.

Glede glavnih faktorjev, ki vplivajo na oceno, še vedno ni nekega splošnega strinjanja in se v študijah pojavi veliko število faktorjev.

Večina študij je še vedno izvedenih znotraj enega samega podjetja. Samo v eni od študij v študiji sodelujeta dve podjetji.

Poglavje 7

Zaključek

Ocenjevanje napora pri agilnem razvoju programske opreme zaradi nenehnega spreminjanja še vedno predstavlja velik izziv. Zanimalo nas je, kaj se je v zadnjih letih dogajalo na tem področju in kaj se je spremenilo v primerjavi s prvotnim sistematičnim pregledom.

Ugotovili smo, da se v objavljenih študijah ne piše več toliko o postopkih za ocenjevanje napora, ki temeljijo na subjektivni strokovni oceni (poker planiranje (angl. planning poker), ocena eksperta (angl. expert judgment) in metoda točk primerov uporabe (angl. use case point method)), ampak je poudarek na odkrivanju novih postopkov za čim bolj točno ocenjevanje napora.

Pri veliki večini študij ni podatka o tem, kako točni so bili uporabljeni postopki. Pri ostalih (razen pri enem) je točnost dobra in ustreza kriteriju za sprejemljiv postopek ocenjevanja. Pojavlja se vprašanje, v kolikšni meri je smiselna uporaba predlaganih modelov, saj je v primerjavi s postopki, ki temeljijo na strokovnih ocenah, potrebno vložiti kar nekaj napora v pripravo podatkov. To se ne sklada z načeli agilnosti, ki zagovarjajo čim manjši obseg dela, ki ni nujno in pa pristope, ki niso preveč časovno zahtevni.

Analizirali smo metrike obsega in glavne faktorje, ki vplivajo na oceno. V izbranih študijah je bila najbolj pogosto uporabljena metrika obsega točke uporabniških zgodb (angl. user story points). Glede najbolj primernih glav-

nih faktorjev, ki vplivajo na oceno, študije niso bile enotne, saj je bilo uporabljenih veliko različnih faktorjev. Tisti, ki so se v študijah največkrat pojavili, so bili znanje razvijalcev, kompleksnost zahtev in razvoja ter poznavanje projekta.

Najbolj pogosto so študije izvedene v industrijskem okolju, in sicer znotraj enega podjetja.

V velikem številu študij ni bilo navedeno, katera agilna metoda je bila predmet raziskave. Pri ostalih je bila največkrat raziskana agilna metoda Scrum, kar je smiselno, ker je to najbolj priljubljena metoda v zadnjih letih.

V skoraj vseh študijah so bile raziskane vse aktivnosti razvoja (analiza, načrtovanje, implementacija in testiranje). Največ študij je bilo osredotočenih na izdajo kot nivo planiranja.

Opazimo, da se večina pomanjkljivosti, ki so bile navedene v prvotnem sistematičnem pregledu, še vedno pojavlja. Edina izboljšava, ki izstopa, je dobra točnost ocen predlaganih modelov.

Priloga A. Seznam primarnih študij

- Š1: Rashmi Popli and Naresh Chauhan. Agile estimation using people and project related factors. In Computing for Sustainable Global Development (INDIACom), 2014 International Conference on, pages 564-569. IEEE, 2014.
- Š2: Srdjana Dragicevic, Stipe Celar, and Mili Turic. Bayesian network model for task effort estimation in agile software development. *Journal of Systems and Software*, 127:109-119, 2017.
- Š3: Kayhan Moharreri, Alhad Vinayak Sapre, Jayashree Ramanathan, and Rajiv Ramnath. Cost-effective supervised learning models for software effort estimation in agile environments. In Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual, volume 2, pages 135-140. IEEE, 2016.
- Š4: Mohd Owais and R Ramakishore. Effort, duration and cost estimation in agile software development. In Contemporary Computing (IC3), 2016 Ninth International Conference on, pages 1-5. IEEE, 2016.
- Š5: Aditi Panda, Shashank Mouli Satapathy, and Santanu Kumar Rath. Empirical validation of neural network models for agile software effort estimation based on story points. *Procedia Computer Science*, 57:772-781, 2015.

- Š6: Rashmi Popli and Naresh Chauhan. Estimation in agile environment using resistance factors. In *Information Systems and Computer Networks (ISCON), 2014 International Conference on*, pages 60-65. IEEE, 2014.
- Š7: Erdir Ungan, Numan Çizmeli, and Onur Demirörs. Comparison of functional size based estimation and story points, based on effort estimation effectiveness in scrum projects. In *Software Engineering and Advanced Applications (SEAA), 2014 40th EUROMICRO Conference on*, pages 77-80. IEEE, 2014.
- Š8: Binish Tanveer, Liliana Guzmán, and Ulf Martin Engel. Understanding and improving effort estimation in agile software development: An industrial case study. In *Proceedings of the International Workshop on Software and Systems Process*, pages 41-50. ACM, 2016.
- Š9: Sunil Kumar Khatri, Sanjana Malhotra, and Prashant Johri. Use case point estimation technique in software development. In *Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), 2016 5th International Conference on*, pages 123-128. IEEE, 2016.
- Š10: CJ Torrecilla-Salinas, Jorge Sedeño, MJ Escalona, and Manuel Mejías. Estimating, planning and managing agile web development projects under a value-based perspective. *Information and Software Technology*, 61:124-144, 2015.
- Š11: Alaa El Deen Hamouda. Using agile story points as an estimation technique in cmmi organizations. In *Agile Conference (AGILE), 2014*, pages 16-23. IEEE, 2014.
- Š12: Alan Ramirez-Noriega, Reyes Juarez-Ramirez, Raul Navarro, and Janeth Lopez-Martinez. Using bayesian networks to obtain the task's parameters for schedule planning in scrum. In *Software Engineering*

Research and Innovation (CONISOFT), 2016 4th International Conference in, pages 167-174. IEEE, 2016.

Literatura

- [1] Manifesto for agile software development. <http://agilemanifesto.org/>. Accessed: 2017-08-15.
- [2] Lan Cao. Estimating agile software project effort: an empirical study. *AMCIS 2008 Proceedings*, page 401, 2008.
- [3] Mike Cohn. *Agile estimating and planning*. Pearson Education, 2005.
- [4] Samuel Daniel Conte, Hubert E Dunsmore, and Vincent Y Shen. *Software engineering metrics and models*. Benjamin-Cummings Publishing Co., Inc., 1986.
- [5] Norman E Fenton and Shari Lawrence Pfleeger. Software metrics: a rigorous and practical approach. 1997. *Brooks/Cole Pub Co*, 1993.
- [6] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. In *Technical report, Ver. 2.3 EBSE Technical Report*. Keele University and University of Durham, 2007.
- [7] S Kuppuswami, Kalimuthu Vivekanandan, Prakash Ramaswamy, and Paul Rodrigues. The effects of individual xp practices on software development effort. *ACM SIGSOFT Software Engineering Notes*, 28(6):6–6, 2003.

-
- [8] Viljan Mahnič and Tomaž Hovelja. On using planning poker for estimating user stories. *Journal of Systems and Software*, 85(9):2086–2095, 2012.
 - [9] Mark Petticrew and Helen Roberts. *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons, 2008.
 - [10] Muhammad Usman, Emilia Mendes, Francila Weidt, and Ricardo Britto. Effort estimation in agile software development: a systematic literature review. In *Proceedings of the 10th International Conference on Predictive Models in Software Engineering*, pages 82–91. ACM, 2014.
 - [11] Jovan Živadinović, Zorica Medić, Dragan Maksimović, Aleksandar Damjanović, and Slañana Vujčić. Methods of effort estimation in software engineering. In *International Symposium Engineering Management And Competitiveness. Zrenjanin, Serbia*, 2011.