

Puheenvuoro

Piiri pieni pyörii

Tarja Systä, Maarit Harsu & Kai Koskimies

Tampereen teknillinen yliopisto

Ohjelmistotekniikka

etunimi.sukunimi@tut.fi

1 Johdanto

Julkaiseminen on tutkimuksen kulmakivi: tutkimustulokset pyritään tuomaan tutkimusyhteisön saataville, niin että niitä voidaan hyödyntää uudessa tutkimuksessa ja yhteisö voi jakaa saman ymmärryksen tieteenalasta. Toisaalta julkaisemista käytetään lisääntyvässä määrin myös mittaamaan tutkijoiden ja tutkimuslaitosten tuottavuutta. Kun samaan aikaan erilaisten mittareiden merkitys resurssien jaossa on kasvamassa, ollaan päädytty tilanteeseen, missä julkaisulla on rahoitukseen ja urakehitykseen vaikuttavaa merkitystä, joka riippuu julkaisufoorumista enemmän kuin julkaisun sisällöstä [5]. Vaikka useimpia tutkijoita — toivon mukaan — motivoi aito kiinnostus tutkimusongelmien ratkaisuun pikemmin kuin tämän tapaiset sekundääriset seikat, on olemassa ilmeinen vaara, että tutkijat muodostavat julkaisufoorumien ympärille sisäänpäinlämpiäviä yhteisöjä, joiden jäsenet palkitsevat toisiaan julkaisuilla ja viittauksilla. Tämän kaltainen kehitys voi tapahtua täysin tiedostamatta

On huomattava, että tutkijayhteisön sisäänpäinlämpiävyys ei välttämättä johda julkaisujen laadun heikkenemiseen. On jopa oletettavaa, että sisäänpäinlämpiävä

yhteisö kehittää ja ylläpitää korkealaatuisia julkaisukulttuuria, mitä tulee julkaisujen rakenteeseen, esitystapaan, tieteellisten menetelmien käyttöön ja yleiseen argumentaatioon. Sisäänpäinlämpiävän yhteisön suurin uhka ei ole julkaisujen laadun heikkeneminen sinänsä, vaan tutkimuksen jähmettyminen ja kasvava irrelevanttisuus muun maailman kannalta.

Konferenssiartikkelit ovat keskeisessä asemassa tietojenkäsittelytieteen julkaisukulttuurissa, kun sen sijaan muilla tieteenaloilla alan tieteelliset lehdet ovat usein se ensi sijainen julkaisufoorumi. Tämän vuoksi keskitymme tässä katsauksessa konferenssijulkaisukäytäntöihin.

Tutkijayhteisön sisäänpäinlämpiävyyttä on vaikea tutkia ja mitata. Ohjelmatoimikunnat (program committee, PC) noudattavat tavallisesti enemmän tai vähemmän tarkkoja sääntöjä, joiden tarkoitus on sulkea pois yksittäiset sisäänpäinlämpiävyyteen johtavat päätökset. Tästä seuraa, että sisäänpäinlämpiävyys ei ole suoraan nähtävissä tilastoissa ja dokumenteissa, vaan se ilmenee lähinnä ihmisten asenteissa ja ajatuksissa. Väitämme kuitenkin, että on olemassa tapoja havaita yhteisön sisäänpäinlämpiävyyttä epäsuorasti.

Koska ohjelmatoimikunta on se taho, joka tekee päätökset julkaisemisesta, oh-

ohjelmatoimikunnan jäsenillä on avainrooli tutkijayhteisön evoluutiossa. Tätä roolia korostaa vielä se, että ohjelmatoimikunnan jäsenyys jo itsessään on usein mitaroitu meriitti. Niinpä keskitymme tässä tutkimuksessa pääasiassa analysoimaan ohjelmatoimikunnan jäsenyyksiin liittyvää tilastotietoa. Lähtökohtamme on oletamus, että terveessä tutkijayhteisössä ohjelmatoimikunta ei suosi itseään julkaisujen hyväksymisessä.

Keskitymme tässä katsauksessa kuutta ohjelmistotuotannon (software engineering) tunnettua konferenssia koskevaan julkisesti saatavilla olevaan tilastotietoon vuosilta 2004–2009. Oletamme kuitenkin, että nämä konferenssit edustavat tyypillisiä käytäntöjä ohjelmistotuotannon alueella. Korostamme, että sen enempää tämän paperin tarkoitus kuin johtopäätökset eivät ole suunnattuja johonkin tiettyyn tai tiettyihin konferensseihin. Huomattakoon myös, että vaikka rajoitumme tässä tarkastelemaan kirjoittajille tuttuja ohjelmistotuotannon konferensseja, voidaan olettaa, että muilla tietojenkäsittelytieteiden aloilla konferenssien käytännöt ovat samankaltaisia.

2 Tutkimusmenetelmä

Olemme käyttäneet kvantitatiivista lähestymistapaa ohjelmatoimikunnan ominaisuuksien tutkimiseen. Kerätty data koskee seuraavia konferensseja/workshoppeja pääasiassa vuosilta 2004–2009:

- ICSE (International Conference on Software Engineering)
 - ICSM (International Conference on Software Maintenance)
 - ICPC (International Conference on Program Comprehension), formerly IWPC (International Workshop on Program Comprehension)
 - CSMR (European Conference on Software Maintenance and Reengineering)
 - WCRE (Working Conferences on Reverse Engineering)
 - GPCE (International Conference on Generative Programming and Component Engineering).
- Konferenssit valittiin erilaisista syistä. ICSE valittiin, koska se on ohjelmistotuotannon johtava konferenssi ja sinne lähetettävät paperit kattavat laajasti kyseisen tutkimusalan. Mukaan haluttiin myös ohjelmistotuotannon osa-alue, joksi valitsimme ohjelmien ymmärtämisen, ylläpidon ja takaisinmallinnuksen (jatkossa käytämme tästä osa-alueesta nimeä ylläpito). Tämä osa-alue tuli valituksi, koska se on aika yhtenäinen ja sillä on hyvin vakiintunut konferenssien joukko (ICSM, ICPC/IWPC, CSMR, WCRE). Kullakin konferenssilla on tietenkin oma erityisalueensa. Listan viimeinen konferenssi, GPCE on nuorempi ja pienempi kuin muut, ja se otettiin mukaan vertailukonferenssiksi muille valituille konferensseille.
- Kaikista yllä mainituista konferensseista kerättiin seuraavat tiedot (suluissa vastaava englanninkielinen termi, jonka perusteella hakasuluissa oleva lyhenne on muodostettu):
- lähetettyjen papereiden määrä (number of submitted papers [#*subm*])
 - hyväksytyjen papereiden määrä (number of accepted papers [#*acc*])
 - PC-papereiden määrä eli sellaisten papereiden määrä, joissa ainakin yksi kirjoittaja on PC:n jäsen (number of accepted paper (co-)authored by PC members [#*PCacc*])
 - PC-jäsenten määrä eli PC:n koko (number of PC members [#*PCmem*])
 - uusien PC-jäsenten määrä (number of

new PC members; compared to the previous year [#newPCmem])

- oikeasti uusien PC-jäsenten määrä (number of new PC members; compared to all previous years starting from 2004 [#realnewPCmem]).

Pääasiallisimpana tietolähteenä käytimme kunkin konferenssin web-sivuja. Lisäksi hyväksymisprosentit sekä lähetettyjen ja hyväksytyjen paperien määrät löytyvät esimerkiksi web-sivulta [7]. Koska rajoitimme vain julkiseen dataan, emme pysty pohtimaan esimerkiksi hylättyihin papereihin liittyviä asioita.

Kerätyn datan pohjalta laskimme seuraavat metriikat:

- hyväksymisprosentti (acceptance rate [$accRate = \frac{\#acc}{\#subm}$])
- PC-papereiden osuus hyväksytyistä papereista (proportion of papers (co-)authored by PC member from all accepted papers [$PCaccProp = \frac{\#PCacc}{\#acc}$])
- uusien PC-jäsenten osuus koko PC:stä (proportion of new PC members [$newPCprop = \frac{\#newPCmem}{\#PCmem}$])
- oikeasti uusien PC-jäsenten osuus koko PC:stä (proportion of real new PC members [$realnewPCprop = \frac{\#realnewPCmem}{\#PCmem}$])
- tarkastuskerroin (review coefficient [$revCoeff = \frac{\#subm}{\#PCmem}$]).

Kertätty data ja lasketut metriikat löytyvät liitteestä, jossa kullekin konferenssille on oma taulukkonsa. Olemme julkaisseet datan myös web-sivulla [3].

3 Datat analysointi

3.1 Ohjelmatoimikunnan suhteellinen koko

Aloitamme kerätyn datan analysoinnin tutkimalla konferenssien ohjelmatoimi-

kuntien kokoa. Luonnollinen oletus olisi, että ohjelmatoimikunnan koko kasvaisi lähetettyjen papereiden määrän kasvaessa. Tällöin eri konferenssien ohjelmatoimikuntien jäsenten ”arviointitaakka” olisi suhteelliseen samankokoinen. Näin ei kuitenkaan kerätyn datan perusteella selvästikään ole, mikä näkyy vertailtaessa tarkastuskertoimia ($revCoeff$), jotka lasketaan jakamalla lähetettyjen papereiden lukumäärä ($\#subm$) ohjelmatoimikunnan koolla ($\#PCmem$). Mikäli tarkastuskerroin on suuri, jokaisen ohjelmatoimikunnan jäsenen keskiarvoinen arvioitavien papereiden lukumäärä (tyypillisesti $3 * revCoeff$) on suuri, ja toisaalta siis heidän vaikutusvaltansa papereiden hyväksymisen ja hylkäämisen suhteen on suuri.

Taulukkoon 1 on koottu tarkasteltujen konferenssien tarkastuskertoimet. Taulukosta havaitaan, että ICSE:lla tarkastuskertoimet ovat huomattavasti suurempia kuin muilla konferensseilla. Selityksenä on se, että vaikka lähetettyjen papereiden määrä ICSE:lla on huomattavasti suurempi kuin muilla tarkastelluilla konferensseilla, ICSE ei ohjelmatoimikuntansa koon puolesta suuresti poikkea muista konferensseista. Itse asiassa ICSE:n ohjelmatoimikunta on jopa huomattavasti pienempi kuin ICSM:n, joka on kuitenkin selvästi pienempi konferenssi ja joka edustaa vain yhtä ohjelmistotuotannon osa-aluetta.

Tarkastellaan ICSE-konferenssia vuodelta 2006, jolloin $revCoeff = 12.0$. Jos pidetään oletuksena yleistä käytäntöä, eli että jokaisen lähetetyn paperin arvioi kolme ohjelmatoimikunnan jäsentä ja että ”arviointitaakka” on tasaisesti jakaantunut ohjelmatoimikunnan jäsenten kesken, saadaan tulokseksi, että kukin ohjelmatoimikunnan jäsen oli vastuussa 36 arviosta. Vastaavilla oletuksilla IWPC 2004 -tapauksessa kukin ohjelmatoimikunnan

jäsen oli vastuussa kolmesta arviosta.

Kun tarkastuskerroin on hyvin suuri, on todennäköisempää ja oletettavaa, että ohjelmatoimikunnan jäsen delegoi joidenkin papereiden arvioinnin esimerkiksi jatko-opiskelijoilleen ja mahdollisesti itse tarkastelee kyseisiä papereita vain pintapuolisesti. On syytä kuitenkin korostaa, että näin ei tietenkään välttämättä ole.

Vaikkakin tarkastuskerrointen huomattavat erot ovat mielenkiintoinen havainto sinänsä, emme kyetneet löytämään erityisiä syitä siihen. ICSE:n tapauksessa pieni ohjelmatoimikunnan koko on jokseenkin yllättävää, koska konferenssi on isoin ohjelmistotuotannon alan konferenssi katkaen koko alan. ICSE on niitä harvoja konferensseja, joissa ovat vielä käytössä fyysiset ohjelmatoimikunnan kokoukset. Sellaisten järjestäminen suurelle ohjelmatoimikunnalle saattaa olla hankalaa, mikä voi osittain vaikuttaa asiaan. Lisäksi ohjelmatoimikunnan koon pitäminen pienenä saattaa olla ICSE:n perinteinen käytäntö.

3.2 PC-papereiden hyväksyminen

Seuraavaksi tutkimme PC-papereiden osuutta hyväksytyistä papereista eli sitä, mikä osuus kaikista konferenssiin hyväksytyistä papereista on niillä papereilla, joiden kirjoittajista vähintään yksi kuuluu ohjelmatoimikuntaan. Tämä osuus vaihtelee huomattavasti konferenssien kesken: se voi suurimmillaan olla yli 70 % (esimerkiksi vuoden 2006 WCRE-konferenssilla 71 %), ja pienimillään se on nolla (esimerkiksi GPCE 2007).

PC-papereiden suuri osuus herättää luonnollisesti kysymyksen: Onko hyväksymisprosentit samat PC-papereille ja muille lähetetyille papereille? Kerätyn datan perusteella emme kuitenkaan pysty vastaamaan tähän kysymykseen, koska ainoastaan tieto hyväksymisprosentteista yleisesti kaikille papereille on julkises-

ti saatavilla. Jos oletamme, että hyväksymisprosentit PC-papereille ja muiden kuin ohjelmatoimikunnan jäsenten kirjoittamille papereille ovat samat, tarkoittaisi se esimerkiksi vuoden 2009 ICSE-konferenssille, että

$$\frac{\#PCacc}{accRate} = \frac{10}{0.12} \approx 83,3$$

lähetetyistä papereista olisi PC-papereita. Koska kyseisen konferenssin ohjelmatoimikunnan koko oli 41, tarkoittaisi se, että jokainen ohjelmatoimikunnan jäsen oli (keskimääräisesti laskettuna) kirjoittajana hieman yli kahdessa (83.3/41) paperissa. Tämä esimerkki ei ole erikoistapaus, vaan myös muita vastaavia esimerkkejä löytyy kerätystä datasta. Luku on niin suuri, että on varsin epätodennäköistä, jos oletus samoista hyväksymisprosentteista pitäisi paikkansa. Sen sijaan vaikuttaa todennäköiseltä, että PC-papereiden hyväksymisprosentti on selvästi suurempi kuin muiden papereiden.

Luonnollinen jatkokysymys onkin: Jos edellä esitetty johtopäätelmä pitää paikkansa, niin miksi ohjelmatoimikunnan jäsenten on helpompi saada paperinsa hyväksytyiksi? Ensinnäkin kuten edellä diskussoimme, ohjelmatoimikunnan jäsenet ovat todennäköisesti keskimääräistä kokeneempia tutkijoita, ja siksi oletettavasti myös saavat keskimääräistä parempia tutkimustuloksia ja kirjoittavat siten myös keskimääräistä parempia papereita. Toinen, tosin spekulatiivinen selitys voisi olla, että papereiden kirjoittajien nimet vaikuttavat arviointeihin ja hyväksymispäätöksiin, erityisesti kun arvioinnissa käytetään yhdensuuntaisen anonymiteetin periaatetta (blind review), jolloin siis arvioijat näkevät papereiden kirjoittajien tiedot. Tämä selitys voi olla realistinen erityisesti, kun paperien arviointi delegoidaan tai arvioijat ovat kokemattomia tai epävarmoja. Tässäkin on kuitenkin syytä korostaa,

Taulukko 1: Tarkastuskertoimet kerättyinä yhteen taulukkoon

Vuosi	2004	2005	2006	2007	2008	2009
ICSE	9.3	7.3	12.0	10.1	9.8	9.9
ICSM	2.1	3.1	1.9	2.5	2.0	2.1
ICPC/IWPC	1.0	1.8	2.0	1.5	1.1	1.5
CSMR	1.7	1.7	1.6	2.4	1.7	1.4
WCRE	2.1	1.6	2.1	2.3	1.3	1.7
GPCE	4.2	5.1	3.4	2.5	2.6	2.1

ettei niin välttämättä tietenkään ole. Delegoitaessa arviointi, on PC-jäsen kuitenkin itse lopulta vastuussa arviosta. Tällöin arviointi saattaa olla erityisen perusteellinen, koska sitä on ollut kirjoittamassa kaksi asiantuntijaa: PC-jäsen itse sekä PC:n ulkopuolinen arvioija, jolle PC-jäsen on arvioinnin delegoinut.

3.3 Ohjelmatoimikunnan vaihtuvuus

Seuraavaksi tarkastelemme ohjelmatoimikuntien evoluutiota ja vaihtuvuutta. Tyypillisesti ohjelmatoimikuntaa muodostettaessa osa edellisen vuoden ohjelmatoimikunnan jäsenistä kutsutaan mukaan, ja lisäksi kutsutaan uusia jäseniä. Uudet jäsenet voivat puolestaan olla sellaisia, jotka eivät ole toimineet kyseisen konferenssin ohjelmatoimikunnissa aiempinakaan vuosina (*#realnewPCmem*), tai sellaisia, jotka ovat (mutta eivät edellisenä vuonna) (*#newPCmem*).

Tutkimalla kerättyä dataa ja laskettuja metriikka-arvoja voimme havaita useita eroavaisuuksia, jotka koskevat ohjelmatoimikuntien vaihtuvuutta. GPCE on suhteellisen nuori konferenssi, joka järjestettiin ensimmäisen kerran vuonna 2002. Sen ohjelmatoimikunta ei ole kovin stabiili. Tämän voi havaita tarkastelemalla *#realnewPCmem*-arvoja suhteessa *#PCmem*-arvoihin.

ICSE on vanhin konferenssi, joka puolestaan järjestettiin ensimmäisen ker-

ran vuonna 1975. Voimme havaita, että sen ohjelmatoimikunnan vaihtuvuus on pienempi kuin GPCE:n, vaikkakin myös ICSE:n ohjelmatoimikunnassa on havaittavissa selvää vaihtuvuutta. Tarkempi analyysi paljastaa kuitenkin, että ICSE:ssä suhteellisen usein uudet ohjelmatoimikunnan jäsenet ovat ”uudelleenkutsuttuja”, ts. he ovat toimineet ohjelmatoimikunnassa aiempina vuosina. Tämä selviää vertaamalla *#realnewPCmem*- ja *#newPCmem*-arvoja sekä niiden suhteellista osuutta ohjelmatoimikunnassa.

Ylläpitöyhteisö on puolestaan melko stabiili, ja ehkäpä siksi myös sen konferenssien ohjelmatoimikuntien vaihtuvuus on pientä. Nämä konferenssit sijoittuvat iältään GPCE- ja ICSE-konferenssien väliin: ICSM-konferensseja on järjestetty vuodesta 1983 lähtien, IWPC/ICPC-konferensseja vuodesta 1992, WCRE-konferensseja vuodesta 1994 ja CSMR-konferensseja vuodesta 1997. Näiden konferenssien välisten suhteiden tarkemman tarkastelun jätämme jatkotutkimusaiheeksi.

3.4 PC-papereiden hyväksymisasteen ja ohjelmatoimikunnan vaihtuvuuden suhde

Havaitsimme mielenkiintoisen yhteyden PC-papereiden hyväksymisprosentin ja ohjelmatoimikunnan vaihtuvuuden välillä. Datan ja metriikoiden avulla on selvästi

todettavissa, että mitä pienempi ohjelματοimikunnan vaihtuvuus on, sitä suurempi on PC-papereiden määrä ja osuus kaikista papereista. Tämä korrelaatio näkyy kuvassa 1. Huomionarvoista on myös se, että jos tutkitaan kaikkia tarkasteltuja konferensseja kokonaisuutena, tämä korrelaatio on erittäin selvä (p-value: $1,532e-06$). Ilmiö ei kuitenkaan päde kaikille konferensseille erikseen. Esimerkiksi CSMR-konferenssin suhteen tilanne on kutainkin päinvastainen.

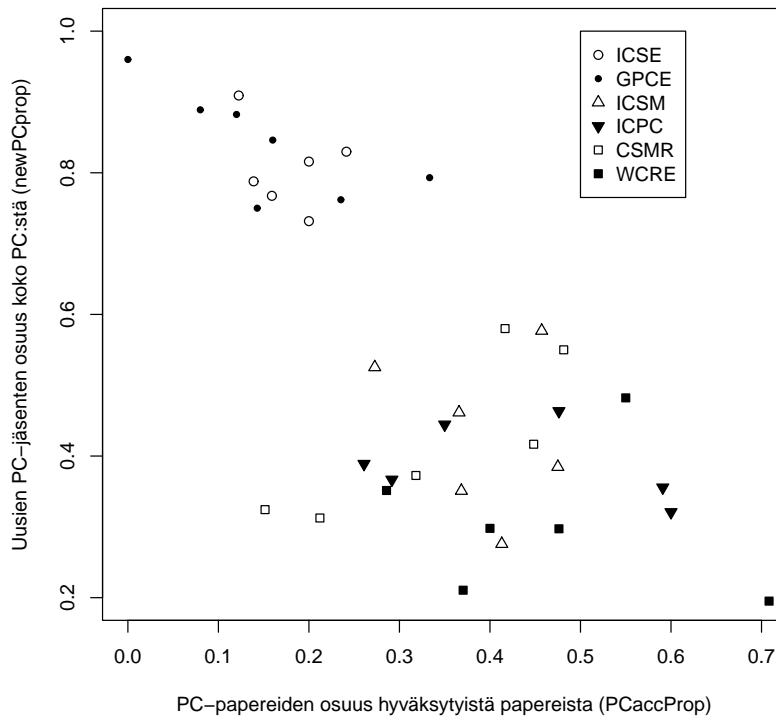
Tarkastellaan aluksi GPCE-konferenssia. Kuten edellä todettiin, sen ohjelματοimikunnan vaihtuvuus on suuri. Lisäksi data (liitetaulukko 6) selvästi osoittaa, että PC-papereiden osuus kaikista hyväksytyistä papereista on varsin pieni koko tarkastellun ajanjakson 2004–2009 aikana (pienimmillään jopa nolla). Vaikuttaa siis siltä, että korkea ohjelματοimikunnan vaihtuvuus johtaa PC-papereiden pieneen osuuteen.

Tarkastelemalla ICSE-konferenssia hieman tarkemmin havaitaan myös selvästi, että korrelaatio ohjelματοimikunnan vaihtuvuuden ja PC-papereiden osuuden välillä on olemassa. Taulukko 2 osoittaa ohjelματοimikunnan evoluution ja sen korrelaation PC-papereiden osuuden suhteen. Vuosille 2004–2010 laskimme lisäksi kumulatiivisen ohjelματοimikunnan (*cumPC*), joka koostuu henkilöistä, jotka toimivat ohjelματοimikunnassa kyseisenä vuotena sekä edellisinä (tarkasteluajanjakson) vuosina. Duplikaatit luonnollisesti poistettiin. Toisin sanoen vuoden 2010 *cumPC* koostuu henkilöistä, jotka toimivat ICSE:n ohjelματοimikunnassa vähintään kerran vuosina 2004–2010. Lisäksi laskimme uusien ohjelματοimikunnan jäsenten määrän verrattuna edellisen vuoden kumulatiiviseen ohjelματοimikuntaan (*newcumPCmem*).

Esimerkiksi vuonna 2005 ohjelmatoi-

mikunnan koko oli 43 (ks. liitetaulukko 1). Verrattuna vuoden 2004 ohjelματοimikuntaan, jolloin *#PCmem* = 47, vuonna 2005 ohjelματοimikunnassa oli 33 uutta jäsentä (siten *#cumPC* = 80) ja 10 vanhaa (sekä vuoden 2004 että vuoden 2005 ohjelματοimikunnassa toiminutta). Vuonna 2006 puolestaan ohjelματοimikunnassa oli 11 sel-laista jäsentä, jotka olivat myös joko vuoden 2004, vuoden 2005 tai molempien ohjelματοimikunnassa/-kunnissa. Huomaa, että *#cumPC*-arvoja ei voida laskea suoraan edellisen vuoden *#cumPC*-arvosta ja tarkastellun vuoden *#newcumPCmem*-arvosta. Lopuksi las-kimme niiden hyväksytyjen papereiden lukumäärän (*cumPCacc*), joissa vähintään yksi kirjoittajista kuuluu kumulatiiviseen ohjelματοimikuntaan.

Taulukosta 2 voimme havaita jälleen korrelaation ohjelματοimikunnan vaihtuvuuden ja PC-papereiden osuuden välillä: kun ohjelματοimikuntaan otetaan suhteellisen monta uutta jäsentä, PC-papereiden osuus (*#cumPCacc*) laskee. Toisaalta, kun ohjelματοimikuntaan otetaan vähän uusia jäseniä, PC-papereiden osuus kasvaa. Esimerkiksi vuonna 2005 *#newcumPCmem*-arvo oli melko korkea (33) ja *#oldcumPCmem*-arvo suhteellisen pieni (10). Tällöin *cumPCacc*-arvo laski neljästätoista yhdeksään (verrattuna edelliseen vuoteen). Vuonna 2006 trendi oli vastaava: korkea ohjelματοimikunnan vaihtuvuus ja *#cumPCacc*-arvon lasku (yhdeksästä viiteen, edelliseen vuoteen verrattuna). Vuonna 2007 sen sijaan ohjelματοimikuntaan otettiin suhteellisen vähän uusia jäseniä (*newcumPCmem* = 14 ja *oldcumPCmem* = 19), jolloin *#cumPCacc*-arvossa havaitaan selvä kasvu (viidestä neljääntoista). Vuonna 2010 50 % kaikista hyväksytyistä papereista oli PC-papereita (26/52) eli tässä tapaukses-



Kuva 1: Hyväksytyjen PC-papereiden ja PC:n vaihtuvuuden välinen korrelaatio

sa papereita, joiden kirjoittajista vähintään yksi kuului kumulatiiviseen ohjelmakuntaan. Vuonna 2010 kumulatiivisen ohjelmakunnan koko oli 140, mikä on silti vielä suhteellisen pieni verrattuna lähetettyjen papereiden lukumäärään: $revCoeff = 380/140 \approx 2.7$ (kts. taulukko 1).

Kuten edellä diskussoimme, korkealle PC-papereiden määrälle on monta mahdollista selitystä. Koska ohjelmatoimikuntien jäsenet ovat tyypillisesti kokeneita ja hyviä tutkijoita, heidän voidaan olettaa kirjoittavan korkeatasoisia papereita, joilla on suhteellisen hyvä mahdollisuus tulla hyväksytyiksi. Tämä ei kuitenkaan näyttäisi olevan riittävä selitys korkeil-

le PC-papereiden määrälle, koska keräämämme data ja lasketut metriikat osoittavat selvää korrelaatiota PC-papereiden hyväksymisprosenttien ja ohjelmatoimikuntien vaihtuvuuden välillä: vähäinen ohjelmatoimikunnan vaihtuvuus implikoi PC-papereiden korkeaa osuutta, kun taas ohjelmatoimikunnan vaihtuvuuden ollessa suuri on PC-papereiden osuus pienempi.

4 Katsaus julkaisukäytäntöihin

Tiedossamme ei ole muita esittämämme kaltaisia tutkimuksia. Läheisiä aihe-

Taulukko 2: PC:n evoluutio ICSE:n osalta

Vuosi	2004	2005	2006	2007	2008	2009	2010
#acc	58	44	36	47	55	50	52
#cumPC	47	80	80	94	113	127	140
#cumPCacc	14	9	5	14	15	15	26
#newcumPCmem	-	33	22	14	21	14	13
#oldcumPCmem	-	10	11	19	17	27	28

ta ovat kuitenkin tutkimuksen arviointitavat ja paperien hyväksymiskriteerit, joita on käsitelty lähteessä [4]. Kirjoittajat tarkastelevat tietojenkäsittelyn julkaisukulttuuria ja sitä, miten voitaisiin arvioida erilaisten julkaisutyyppeiden (lehtijulkaisut, konferenssijulkaisut, kirjat) todellista arvoa. Kirjoittajien mukaan myös arviointikriteerit kaipaisivat arviointia.

Toisena läheisenä aiheena tarkastellaan Pattersonin julkaisua, jossa kirjoittaja on huolissaan siitä, että tavanomaiset paperit hyväksytään helpommin kuin ”suuren idean” sisältävät paperit [6]. Pattersonin oleellisin sanoma on se, miten jonkin uuden suuntauksen esittely voitaisiin saada helpommin julkaistua, mutta työssä kiinnitetään myös huomiota lisääntyneeseen lähetettyjen paperien määrään ja siihen, että myös PC:n pitäisi kasvaa vastaavasti. Jos PC on hyvin suurin (mahdollisesti jopa 300 jäsentä) osallistuminen kokouksiin ja niissä keskustelu on hankalaa. Lisäksi, jos PC ei kasva lähetettyjen paperien määrän kasvaessa, paperien tarkistaminen tulee hyvin työlääksi tarkastajille.

Kuten tässä työssä tuli esiin, PC-papereiden hyväksymisprosentti voi olla hyvin korkea. Yksi syy tähän voi olla se, että tunnettujen kirjoittajien papereita hyväksytään helpommin. Tähän liittyen Hassan ja Holt [2] ovat keränneet tietoa siitä, ketkä kirjoittajat tuntevat toisensa takaisinmallinnuksen alalla. Tosin he eivät tarkastele asiaa PC-jäsenyyden kan-

nalta vaan pelkästään sen mukaan, ketkä ovat kirjoittaneet papereita toistensa kanssa. Walker *et al.* [8] ovat puolestaan tutkineet paperien arvosanojen jakaumaa ja ns. ”harmaan alueen” leveyttä.

5 Pohdintaa ja suosituksia

Perinteiseen vertaisarviointiin (peer-to-peer review) liittyviä ongelmia on aina silloin tällöin otettu esille, mutta tavallisesti johtopäätös on kuitenkin ollut, että huolimatta heikkouksistaan tämä on kuitenkin paras tunnettu tapa valikoida julkaistavat työt. Myöskään käsillä oleva tutkimus ei anna aihetta kyseenalaistaa tätä prosessia sinänsä. Sen sijaan tutkimus mielestämme viittaa siihen, että julkaisuforumien ympärille syntyneet tutkijayhteisöt ovat jossain määrin alttiita sisänpäinlämpiävyydelle, joka ilmenee PC-papereiden korkeahkona määränä. Erityisesti PC-papereiden määrän ja PC:n pienen vaihtuvuuden välinen korrelaatio on huolestuttava. Johtopäätöksemme on, että arviointiprosessiin tulisi lisätä sääntöjä ja mekanismeja, jotka toimivat sisänpäinlämpiävyyttä vastaan.

Yleisesti arviointiprosessin läpinäkyvyyden lisääminen todennäköisesti vähentäisi ohjelmatoimikunnan sisänpäinlämpiävyyttä. Tähän suuntaan vaikuttavia konkreettisia käytäntöjä olisivat esimerkiksi seuraavat:

Konferenssien järjestäjien tulisi

- julkaista lähetettyjä ja hyväksytyjä PC-papereita ja ei-PC-papereita koskevat tilastot
- ottaa käyttöön PC:n vaihtuvuutta koskevat selkeät ja sitovat säännöt
- varmistaa, että PC-paperit saavat osakseen riittävän tiukan arvioinnin.

Eräs mahdollinen keino jälkimmäisen varmistamiseksi olisi edellyttää, että PC-paperien hyväksymisprosentti ei saa ylitää yleistä hyväksymisprosenttia. Koska voidaan olettaa, että näiden papereiden taso on keskimääräisen tason yläpuolella, tämä takaa sen, että PC-paperit arvioidaan todennäköisesti tiukemmin kuin muut paperit, mikä mielestämme on terve tilanne.

Ohjelmatoimikunnan jäsenten tulisi

- olla hyväksymättä kutsua saman konferenssin ohjelmatoimikuntaan enempää kuin kolme kertaa peräkkäin, vaikka kyseisen konferenssin omat säännöt sen sallisivatkin
- välttää ajatusta, että ohjelmatoimikunnan jäsenyys on erityinen tilaisuus julkaista omia tai omien jatko-opiskelijoiden papereita.

Periaatteessa ohjelmatoimikunnan tulisi aina pyrkiä hakemaan uusia jäseniä. Vaikka ohjelmatoimikunnan puheenjohtajilla on taipumus enemmän tai vähemmän automaattisesti kutsua useimmat edellisen vuoden jäsenistä uudelleen, jäsenet itse voisivat huolehtia siitä, että ohjelmatoimikunnan pieni vaihtuvuus ei johdu ainaakaan heistä. Viimeinen suositus voi kuulostaa hurskaalta toiveelta, koska se liittyy enemmänkin asenteisiin kuin käytäntöihin. Mielestämme se on kuitenkin siinä mielessä paikallaan, että tässä tutkimuksessa käsitelty ongelma saattaa juuri-kin johtua asenteiden pienestä muutoksesta, joka on seurausta Parnas'in [5] kritisoinnista numeropelistä.

Kiitokset

Kirjoittajat haluavat kiittää Ira Baxteria, Imed Hammoudaa ja Mika Kataraa hyödyllisistä kommentteista ja palautteesta, joka koski paperin englanninkielistä versiota. Kiitämme myös Jorma Tarhiota suomenkielistä versiota koskevista kommentteista.

Viitteet

1. EasyChair Conference System. Available from <http://www.easychair.org/> [Viitattu 27.4.2010]
2. Hassan, A., Holt, R., The small world of software reverse engineering, *Proceedings of the 11th Working Conference on Reverse Engineering (WCRE 2004)*, Delft University of Technology, the Netherlands, November 2004, pp. 278–283.
3. Inbreeding in software engineering conferences? Available from <http://www.cs.tut.fi/~tsysta/ConfPractises.html> [Viitattu 27.8.2010]
4. Meyer, B., Choppy, C., Staunstrup, J., van Leeuwen, J., Research evaluation for computer science, *Communications of the ACM* **52** (4) 2009 pp. 31–34.
5. Parnas, D., Stop the numbers game: counting papers slows the rate of scientific progress, *Communications of the ACM* **50** (11) 2007 pp. 19–21.
6. Patterson, D., The health of research conferences and the dearth of big idea papers, *Communications of the ACM* **47** (12) 2004 pp. 23–24.
7. Software Engineering Conferences, Available from <http://people.engr.ncsu.edu/txie/seconferences.htm> [Viitattu 16.2.2010]
8. Walker, H., Ma, W., Mboya, D., Variability of referees' ratings of conference papers, *Proceedings of the 7th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE 2002)*, Aarhus, Denmark, June 2002, pp. 178–182.

Liite: Kerätty data**Liitetaulukko 1: ICSE-data**

Vuosi	2004	2005	2006	2007	2008	2009
#subm	436	313	395	334	371	405
#acc	58	44	36	49	55	50
#PCacc	14	7	5	6	11	10
#PCmem	47	43	33	33	38	41
#newPCmem	39	33	26	30	31	30
#realnewPCmem			22	12	19	11
accRate	0.13	0.14	0.09	0.15	0.15	0.12
PCaccProp	0.24	0.16	0.14	0.12	0.20	0.20
newPCprop	0.83	0.77	0.79	0.91	0.82	0.73
realnewPCprop			0.67	0.36	0.50	0.27
revCoeff	9.28	7.28	11.97	10.12	9.76	9.88

Liitetaulukko 2: ICSM-data

Vuosi	2004	2005	2006	2007	2008	2009
#subm	122	180	147	214	156	162
#acc	38	55	41	46	40	35
#PCacc	14	15	15	19	19	16
#PCmem	57	59	78	87	78	78
#newPCmem	20	31	36	24	30	45
#realnewPCmem			17	20	22	19
accRate	0.31	0.31	0.28	0.21	0.26	0.22
PCaccProp	0.37	0.27	0.37	0.41	0.48	0.46
newPCprop	0.35	0.53	0.46	0.28	0.38	0.58
realnewPCprop			0.22	0.23	0.28	0.24
revCoeff	2.14	3.05	1.88	2.46	2.00	2.08

Liitetaulukko 3: ICPC/IWPC-data

Vuosi	2004	2005	2006	2007	2008	2009
#subm	42	54	73	66	57	74
#acc	21	24	23	22	20	20
#PCacc	10	7	6	13	12	7
#PCmem	41	30	36	45	53	51
#newPCmem	19	11	14	16	17	20
#realnewPCmem			12	10	14	14
accRate	0.50	0.44	0.32	0.33	0.37	0.27
PCaccProp	0.48	0.29	0.26	0.59	0.60	0.35
newPCprop	0.46	0.37	0.39	0.36	0.32	0.39
realnewPCprop			0.33	0.22	0.26	0.27
revCoeff	1.02	1.80	2.03	1.47	1.08	1.45

Liitetaulukko 4: CSMR-data

Vuosi	2004	2005	2006	2007	2008	2009
#subm	62	81	65	85	87	70
#acc	33	33	27	29	24	22
#PCacc	5	7	13	13	10	7
#PCmem	37	48	40	36	50	51
#newPCmem	12	15	22	15	29	19
#realnewPCmem			20	11	23	11
accRate	0.53	0.41	0.42	0.34	0.28	0.31
PCaccProp	0.15	0.21	0.48	0.45	0.42	0.32
newPCprop	0.33	0.31	0.55	0.42	0.58	0.37
realnewPCprop			0.55	0.31	0.46	0.22
revCoeff	1.72	1.69	1.63	2.36	1.74	1.37

Liitetaulukko 5: WCRE-data

Vuosi	2004	2005	2006	2007	2008	2009
#subm	78	58	84	87	70	79
#acc	28	21	24	27	20	20
#PCacc	8	10	17	10	11	8
#PCmem	37	37	41	38	56	47
#newPCmem	13	11	8	8	27	14
#realnewPCmem			7	7	24	7
accRate	0.36	0.36	0.29	0.31	0.29	0.25
PCaccProp	0.29	0.48	0.71	0.37	0.55	0.40
newPCprop	0.35	0.30	0.20	0.21	0.48	0.30
realnewPCprop			0.17	0.18	0.43	0.15
revCoeff	2.11	1.57	2.05	2.29	1.25	1.68

Liitetaulukko 6: GPCE-data

Vuosi	2002	2003	2004	2005	2006	2007	2008	2009
#subm	39	62	75	87	88	62	55	62
#acc	18	21	25	25	25	19	17	18
#PCacc	1	3	2	3	4	0	4	6
#PCmem	22	20	18	17	26	25	21	29
#newPCmem		15	16	15	22	24	16	23
#realnewPCmem			14	11	14	21	10	18
accRate	0.46	0.34	0.33	0.29	0.28	0.31	0.31	0.29
PCaccProp	0.06	0.14	0.08	0.12	0.16	0	0.24	0.33
newPCprop		0.75	0.89	0.88	0.85	0.96	0.76	0.79
realnewPCprop			0.78	0.65	0.54	0.84	0.48	0.62
revCoeff	1.77	3.10	4.17	5.12	3.38	2.48	2.62	2.14