

Automatinio ir pusiau automatinio vertinimo ypatumai IT valstybiniame brandos egzamine

Bronius Skūpas

Matematikos ir informatikos institutas

Akademijos g. 4, LT-08663 Vilnius

E. paštas: bskupas@ktl.mii.lt

Santrauka. Informacinių technologijų valstybinis brandos egzaminas (ITVBE) pasižymi specifine savybe – praktinėmis programavimo užduotimis, atliekamomis prie kompiuterio. Užduočių sprendimai vertinami tiek automatiškai, tiek pusiau automatiškai būdu. Straipsnyje analizuojamos priežastys, kurios verčia taikyti pusiau automatinį vertinimą šio egzamino vertinime. Tiriamos gerų programų, neįveikusių automatinio testavimo testų, būdingos klaidos.

Raktiniai žodžiai: pusiau automatinis programavimo užduočių sprendimų vertinimas, automatinis programavimo užduočių sprendimų vertinimas, programavimo mokymas.

Įvadas

Šiuolaikinis informacinių technologijų brandos egzamino projektas pradėtas rengti 2002 metais. Jame numatyti du informatikos egzaminai: informacinių technologijų ir programavimo [6, 2]. 2009 metais mokyklinis ir valstybinis egzaminai sujungti į vieną [3], tačiau egzamino struktūra liko panaši. Todėl galima sakyti, kad panašus informacinių technologijų valstybinis brandos egzaminas (anksčiau: valstybinis informacinių technologijų (programavimo) egzaminas) vyksta jau penketą metų.

Informatikos brandos egzaminai vyksta ne vien Lietuvoje, tačiau dauguma valstybių šių egzaminų programoje neturi praktinių programavimo užduočių. Šiuo požiūriu Lietuvoje vykdomas ITVBE artimas Lenkijoje vykdomam informatikos brandos egzaminui. Kolczyk [5] pateikia Lenkijos egzamino sudėtį, analizuoja egzamino rezultatus ir pateikia išvadą, kad toks egzaminas negali būti skirtas daugumai mokinių, tačiau jis skatina mokinius domėtis informatikos dalyku, leidžia mokiniams tikslingiau pasirinkti informatikos studijas.

Praktinės programavimo užduotys Lietuvos ITVBE sprendžiamos naudojantis kompiuteriu bei lokalizuota Free Pascal 0.6.4a programavimo terpe (FPS). Visi praktiniai užduočių sprendimai vertinami bent po du kartus skirtingų vertintojų. Taškams daugiau išsiskyrus, sprendimas vertinamas trečio vertintojo.

Kandidatų pateiktos programos vertinamos panaudojant specialią vertinimo programinę įrangą [7]. Ji pradžioje įvertina sprendimus dinaminės analizės būdu, panaudodama juodos dėžės testavimo metodiką su vertinimo schemeje pateiktais testais. Jei programa įveikia šiuos testus ji gauna 4/5 viso įvertinimo. Likusi 1/5 taškų gaunama už sąlygoje pateiktus kriterijus, kaip pvz. nurodytų funkcijų ar duomenų struktūrų sukūrimą bei naudojimą [4].

Vertinimo kriterijai	Taškai	Pastabos
Testai	20	Visi taškai skiriami, jeigu programa pateikia teisingus visų testų rezultatus
Teisingai skaitomi duomenys iš failo	4	Vertinama tada, kai neskiriama taškų už testus
Teisingai spausdinamas rezultatas	2	
Sukurta funkcija, kuri suskaičiuoja, kiek šachmatų žaidimo komplektų galima sudaryti iš mokinių atneštų baltų figūrų	5	
Teisingos kitos procedūros ir funkcijos, jeigu jų yra, ir pagrindinė programa	9	
Teisingai aprašyti vienmačio masyvo duomenų tipas (tipai) ir kintamieji	1	
Sukurta nurodytus skaičiavimus atliekanti funkcija	1	Visada vertinama
Prasmingai pavadinti kintamieji. Komentuojamos programos dalys, laikomasi rašybos taisyklių	1	
Išlaikomas vientisas programos rašymo stilius, nėra sakinių, skirtų darbui su ekranu	2	
Iš viso taškų	25	

1 pav. 2010 m. ITVBE I praktinės užduoties vertinimo kriterijai [4].

Vertinant egzamino darbus nepasitenkinama tik juodos dėžės testavimo rezultatais – jei programa neveikia dalies ar visų testų, ji vertinama alternatyviu būdu. Kandidatas gali surinkti jau minėtas 4/5 taškų pagal kriterijus. Tai ypač svarbu, jei programos tekste yra smulkių klaidų. 1 pav. pateikiama 2010 m. pirmosios praktinės užduoties vertinimo kriterijai, kuriuose 2–5 punktai ir yra šio alternatyviojo vertinimo pavyzdžiai. Vertintojai pradėdami vertinimą analizuoja pateiktus kriterijus ir dar papildomai juos skaido bei detalizuoja, tuo būdu siekdami tikslesnio vertinimo.

Tačiau net ir gerai parengta vertinimo sistema, detalizuoti kriterijai neleidžia pasiekti, kad skirtingų vertintojų balai sutaptų. Straipsnio tikslas – analizuoti užduočių parengimo vertinimui bei paties vertinimo procesą; ištirti gautų rezultatų neatitikimų mastą ir priežastis; pasiūlyti galimus vertinimo proceso tobulinimo būdus.

1 Automatinis vertinimas

Automatinis mokinių programų vertinimas yra jau senai tiriamas [1]. Vertinimas yra objektyvus, nes viskas vyksta pagal aiškų algoritmą.

Automatinio vertinimo sistema vertindama konkrečią programą ją kompiliuoja, parengia vykdymui saugią aplinką, paruošia pradinių duomenų failus, vykdo kandidato programą tuo pačiu stebėdama vykdymo laiką, nutraukia per ilgai užsitęsusių programų vykdymą, vykdo specialias užduočių sprendimų pateiktų atsakymų vertinimo programas, sukaupus vertinimo rezultatus išsaugo tolimesnei analizei.

Rengiantis automatiniam vertinimui reikia parengti testinius duomenų rinkinius bei vertinimo programą, vertinančią užduočių sprendimų pateiktus atsakymus.

Testiniai duomenų rinkiniai turi būti:

- pakankamai trumpi ir aiškūs, suprantami vertintojams,
- tiksliai identifikuoti galimus blogus užduoties sprendimus,
- demonstruoti programos funkcionavimą įvairiose situacijose.

Vertinimo programa vertinanti užduočių sprendimų pateiktus atsakymus turi:

- vertinti tolerantiškai smulkiems netikslumams,
- diagnozuoti tikslią problemos ar klaidos vietą,
- pateikti aiškius pranešimus žmogui–vertintojui.

2 Alternatyvusis vertinimas ir pusiau automatinis vertinimas

Pusiau automatinio vertinimo metu vertintojas gali pasinaudoti statinės ir dinaminės analizės būdu gautais duomenimis, modifikuoti pateiktą sprendimą ir pabandyti pakartoti automatinį vertinimą. Šioje stadijoje vertintojas turi apsispręsti, kiek reikšmingas klaidas padarė kandidatas, kiek svarbių užduoties dalių jis išsprendė, kokias programos konstrukcijas jis moka. Čia iškyla nemažai problemų dėl vertinimo subjektyvumo, vertintojų nuomonių nevienodumo, skirtingos vertintojų patirties ir pan.

Labai svarbu tinkamai pasirengti šiai vertinimo daliai: tinkamai paruošti ir patikslinti vertinimo schemą. Tačiau pasirengti alternatyviajam vertinimui nėra paprasta, nes:

Kriterijai turi būti:

- aiškūs, vienareikšmiškai suprantami vertintojams,
- tiksliai atitinkantys galimus užduoties sprendimus,
- numatantys taškus už kandidato pademonstruotas žinias.

Taškai už kriterijus turi:

- atitikti egzamino matricą,
- būti proporcingi žinių svoriui.

Kita šio vertinimo problema yra ta, kad jis reikalauja daug žmogiškojo darbo laiko.

3 Tyrimas

Buvo analizuojamos 2010 m. ITVBE programos, kurios neįveikė testų, tačiau bent vienas iš vertintojų skyrė visus taškus už alternatyviusius kriterijus. 1 lentelėje pateikiami duomenys apie šių tiriamų programų skaičių.

Minimos programos buvo pasirinktos, nes jos turėjo būti pakankamai artimos veikiančioms (kitaip vertintojai nebūtų davę daug alternatyvių balų). Kadangi jos sudaro arti 5% visų pateiktų programų, tai patvirtina prielaidą, kad vien automatinio vertinimo neužtenka ITVBE vertinimui.

Taip pat šis tyrimas išryškino tam tikras vertintojų pozicijas, nes jei vienas vertintojas davė visus balus už alternatyvų vertinimą, tai kitas galėjo visų balų neduoti. Buvo pasigilinta į skirtingai įvertintus kriterijus. Dažniausiai galima buvo identifikuoti arba ne iki galo tikslų kriterijų taikymą, arba ne iki galo aiškią kriterijaus formulotę ją taikant konkrečiam sprendimui.

Tokių kriterijų pavyzdys galėtų būti kriterijus apie globalių kintamųjų naudojimą. Skirtingi vertintojai vertindami programą su globaliu masyvu tai interpretavo skirtingai. Šios problemos iliustracijai tiktų ir sumos kintamojo inicializavimo kriterijus.

1 lentelė. 2010 m. valstybinio IT egzamine pateiktų programų kiekis.

	Pateikta iš viso	Programų su tiriamu vertinimu
Pirmos užduoties	1224	34
Antros užduoties	1019	62
Per abi užduotis	2243	96

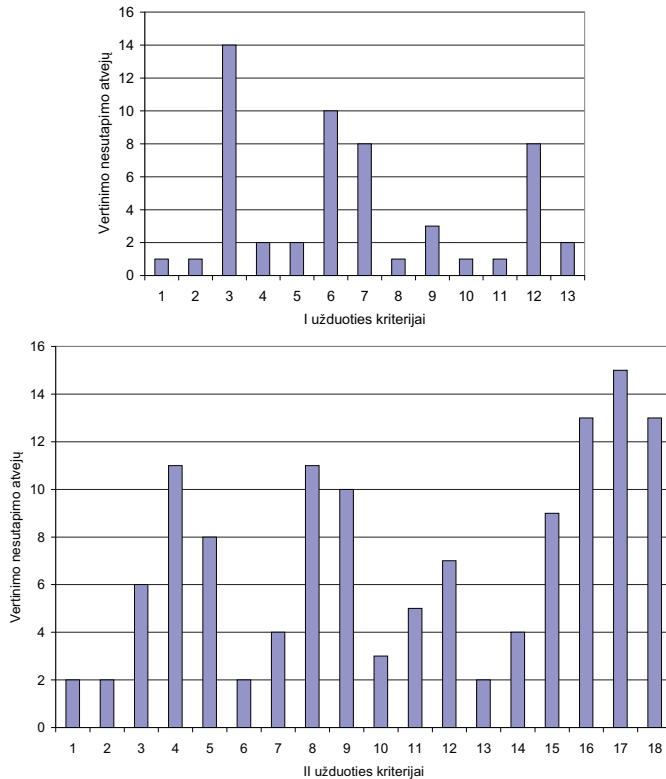
Kai kurie vertintojai laikė, kad globalius kintamuosius neprivaloma inicializuoti nuliu, nes Free Pascal juos inicializuoja automatiškai.

Buvo analizuojamos tipingos klaidos bei priežastys, dėl kurių automatinis vertinimas nepateikė balų už testus.

Viena būdingiausių situacijų buvo skirtingo Pascal kalbos dialekto naudojimas. Vertinimo sistemoje programos testuojamos naudojant Turbo Pascal 7.0 dialektą, kuris yra standartas de facto mokymui ir įvairioms varžyboms. Free Pascal turi įvairių kalbos išplėtimų, kuriuose leidžiama funkcijos vardą naudoti kaip kintamąjį, deklaruo- jant funkcijas be argumentų galima naudoti tuščius skliaustelius ir pan. Vertintojai pertestuodavo minimaliai pataisę programas ir gaudavo testų teisingus atsakymus. Šioje situacijoje tos pačios programos taisymai vertinti skirtingai. Matyt, tam įta- kos turėjo vertintojų patirtis, požiūris į šios užduoties paskirtį egzamine (mokinio algoritmas užrašytas teisingas, neteisingas tik programavimo kalbos dialektas).

Kita būdinga programų klasė tai programos su įvairiomis skubėjimo, neatidumo klaidomis: netikslūs failų vardai, dokumentuojant programą netyčia užkomentuoti svarbūs fragmentai, neinicializuoti kintamieji. Galima spėti, kad šių klaidų gali būti sunku išvengti dėl egzamine kandidatų patiriamo streso ir riboto laiko.

2 pav. pateikiamos diagramos demonstruojančios, kad skirtingų vertintojų verti- nimai įvairiuose kriterijuose skyrėsi netolygiai. Be jau minėtų bendrų problemų la-



2 pav. Alternatyvių vertinimo kriterijų nesutampančių vertinimų histogramos (tiriamų programų aibėje).

biau skyrėsi vertintojų vertinimai apie funkcijų gražinamos reikšmės inicializavimą, skaičiavimą, rezultato gražinimą. Tai greičiausiai rodo, kad šių kriterijų vertinimas vertintojams sukelia problemų analizuojant pateiktas programas.

Histogramose taip pat išsiskiria ir grynai techniniai kriterijai, pvz. teisingas failo skaitymas. Galbūt vertintojams jie kėlė abejones, nes programos pilnai nefunkcionavo, o klaidos vertė nebuvo lengvai įvertinama.

Taip pat teigtina, kad trečiasis vertinimas mažiau skyrėsi nuo antrojo vertinimo, nei antrasis nuo pirmojo. Tai galima paaiškinti kylančia vertintojų kvalifikacija duotai užduočiai.

4 Išvados

Automatinis programų vertinimas naudingas greitesniam ITVBE užduočių vertinimui. Tačiau kandidatų nepasiruošimas griežtai specifikuotam automatiniam programų vertinimui naudojant testus sukelia problemas. Būdingas problemų sprendimas – pusiau automatinis vertinimas yra neišvengiamas šioje vertinimo situacijoje.

Pusiau automatinis vertinimas kelia aukštus reikalavimus kriterijų kokybei bei vertintojų bendrai pozicijai vertinimo situacijose. Pusiau automatinio vertinimo kokybė pagerėja vertinant daugiau darbų. Didelę įtaką vertinimo kokybei gali turėti vertintojų patirtis ir pasirengimo vertinimui etapas, kurio metu analizuojamos populiariausios klaidos, priimami sprendimai dėl kriterijų detalizavimo.

Literatūra

- [1] K.M. Ala-Mutka. A survey of automated assessment approaches for programming assignments. *Computer Science Education*, **15**:83–102, June 2005.
- [2] J. Blonskis, V. Dagienė, R. Burbaitė. Lyginamoji informacinių technologijų valstybinių brandos egzaminų analizė. *Liet. mat. rink.*, **48/49**:92–98, 2008. Adresas internete: ftp://ftp.science.mii.lt/pub/publications/48-49_TOMAS%282008%29/MAT_INF.%20DESTYMAS/Blon_bur.pdf.
- [3] *Informacinių technologijų brandos egzaminų programa*. Nacionalinis egzaminų centras, patvirtinta Lietuvos švietimo ir mokslo ministro 2005 m. liepos 21 d. įsakymu Nr. ISAK-1524 (Lietuvos Respublikos švietimo ir mokslo ministro 2009 m. rugsėjo 4 d. įsakymo Nr. ISAK-1801 redakcija), Vilnius, 2005. Adresas internete: http://www.egzaminai.lt/failai/1257_programa_IT_2009_09_04.pdf.
- [4] *Informacinės technologijos, 2010 m. valstybinio brandos egzamino užduotis, pagrindinė sesija*. Nacionalinis egzaminų centras, Vilnius, 2010. Adresas internete: http://www.egzaminai.lt/failai/1602_IT-pagr-2010.pdf.
- [5] E. Kolczyk. Examiner's remarks on informatics matura examination in Poland. *Informatics in Education*, **8**(2):251–260, 2009. Available from Internet: http://www.mii.lt/informatics_in_education/pdf/INFE150.pdf.
- [6] *Lietuvos bendrojo lavinimo mokyklos bendrosios programos ir išsilavinimo standartai XI–XII klasėms*. Švietimo plėtotės centras, patvirtinta Lietuvos Respublikos švietimo ir mokslo ministro 2002 m. rugpjūčio 21 d. įsakymu Nr. 1465, Vilnius, 2002. Adresas internete: <http://www.pedagogika.lt/puslapis/bps.htm>.
- [7] B. Skūpas and V. Dagienė. Is automatic evaluation useful for the maturity programming exam. In *Proceedings of 8th International Conference on Computing Education Research, Koli Calling'2008*, pp. 117–118, 2008.

SUMMARY

Automatic and semi-automatic program evaluation specifics in IT maturity exam

B. Skūpas

National IT maturity exam in Lithuania have important practical programming part. Task solutions are evaluated in automatic and semi-automatic way. Article analyses reasons for semi-automatic evaluation in exam. Good programs with poor automatic testing results are analysed also and errors are classified.

Keywords: semi-automatic program evaluation, automatic program evaluation, teaching of programming.