

testARS: Pregled zbirke optički čitanih testova

Predrag R. Jovanović, Andrija M. Bošnjaković, Igor I. Tartalja

Sadržaj — Rad opisuje unapređenja programa *testARS* za optičko prepoznavanje i bodovanje testova. Navedene su nove funkcionalnosti koje povećavaju kvalitet i upotrebljivost programa, kao i rešenja problema uočenih u dosadašnjem višegodišnjem korišćenju prethodne verzije programa. Posebno je obraćena pažnja na efikasnost prilikom pregledanja testova kojima je obuhvaćeno više stotina kandidata, uz podršku većeg broja načina skeniranja popunjenih obrazaca. U radu su ukratko predstavljeni najznačajniji koncepti upotrebe nove verzije programa, među kojima se posebno izdvaja podrška za rad sa zbirkama obrazaca.

Ključne reči — bodovanje testa, obrazac, skeniranje, test, zbirka testova.

I. UVOD

NA Elektrotehničkom fakultetu Univerziteta u Beogradu je već duže vreme u upotrebi programski sistem *test* [1], čija je namena pripremanje pitanja, sastavljanje i pregledanje testova, te pravljenje potrebnih izveštaja o rezultatima testova. Program *testARS* (skraćeniica potiče od *test Answer Recognition and Scoring*) je jedan od ključnih elemenata paketa *test*. Njegov zadatak je da skenira i pregleda obrasce sa odgovorima studenata na pitanja sa ponuđenim odgovorima, te da na osnovu prepoznatih odgovora i pravila bodovanja testa na koji se obrasci odnose, odredi odgovarajući broj poena.

Na razvoju programa *testARS* je radio veći broj programa. Zajednički deo svih verzija je algoritam prepoznavanja obrasca metodom kontrolnih tačaka, koji je u svakoj verziji po malo modifikovan. Prva verzija [2] je radila pod operativnim sistemom MS DOS, na računaru klase 286 sa Hercules grafičkom karticom. Sa napretkom tehnologije, program je prilagođen operativnom sistemu Windows [3], [4]. Zatim je program proširen u skladu sa idejama nastalim tokom korišćenja programa [5], [6].

II. MOTIVI ZA UNAPREĐENJE PROGRAMA *TESTARS*

U ranijim verzijama, korisnik programa je pomoću TWAIN izvora (najčešće ručnog skenera) unosio sliku obrasca koju je *testARS* potom odmah obrađivao. Nakon obrade obrasca, korisniku su na raspolaganju bili odgovori studenta na pitanja iz testa, indeks studenta i varijanta ispita. Nakon eventualnih izmena, dobijeni podaci su bili snimani u određenom formatu u izlazni fajl. Potom je korisnik skenirao naredni obrazac, pri čemu je slika upravo obrađe-

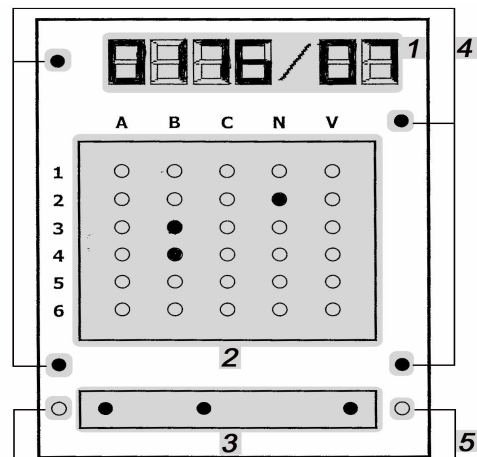
nog obrasca bila obrisana.

Prethodna verzija programa (verzija 2.0) se u praksi pokazala stabilnom i korisnom, ali se kroz njenu eksploataciju pojavilo nekoliko ideja za unapređenje. Osnovni ciljevi su bili povećanje pouzdanosti programa i povećanje brzine skeniranja popunjenih obrazaca, uz što veću fleksibilnost podešavanja parametara. Posebno je bilo značajno proširiti program mogućnošću rada sa obrascima dobijenim korišćenjem skenera sa automatskim prihvatanjem papira (engl. *ADF – Automatic Document Feeder*), koji bez intervencije korisnika skenira sve obrasce i snima ih u određeni grafički format (višestрани TIFF, engl. *Multi-page TIFF*). Zatim je trebalo realizovati vizuelni sistem podešavanja metričkih parametara obrasca i omogućiti napredno manipulisanje kako pojedinačnim obrascima, tako i zbirkama obrazaca. Pored korisniku vidljivih izmena, uočena je potreba za restrukturiranjem i doradom programskog koda tako da prati određene konvencije, koje olakšavaju timski rad i budući razvoj.

III. IZGLED OBRASCA

Slika 1 prikazuje osnovne tri logičke celine obrasca i kontrolne tačke. Logičke celine su polja za broj indeksa (koji predstavlja jednoznačnu identifikaciju kandidata), odgovori kandidata na pitanja i kodirana varijanta testa. Kontrolne tačke su neophodan element za otkrivanje rotacije i eventualnog izobličenja (čestog pri ručnom skeniranju), odnosno za pronalaženje pozicija ostalih elemenata od interesa na popunjenom obrascu.

Indeks je, za potrebe automatskog pridruživanja bodova odgovarajućem kandidatu i snimanja rezultata, predstavljen sedmosegmentnim cifarama koje popunjavaju kandidati. Test može imati nekoliko različitih varijanti sa istim



Slika 1 – Izgled tipičnog test obrasca, sa posebno označenim celinama: 1. indeks ili identifikacioni broj, 2. mesto za označavanje odgovora, 3. kodirana varijanta obrasca, 4. popunjene kontrolne tačke, 5. prazne kontrolne tačke.

P. R. Jovanović, Drenik, ND d.o.o, Deligradska 19, 11000 Beograd, Srbija (telefon: 381-11-2659-641; e mail: pjsoft@drenik.net).

A. M. Bošnjaković, I. I. Tartalja, Elektrotehnički fakultet u Beogradu, Bulevar kralja Aleksandra 73, 11120 Beograd, Srbija; (telefon: 381-11-3218-385; e mail: andrija@etf.rs, tartalja@rcub.bg.ac.yu).

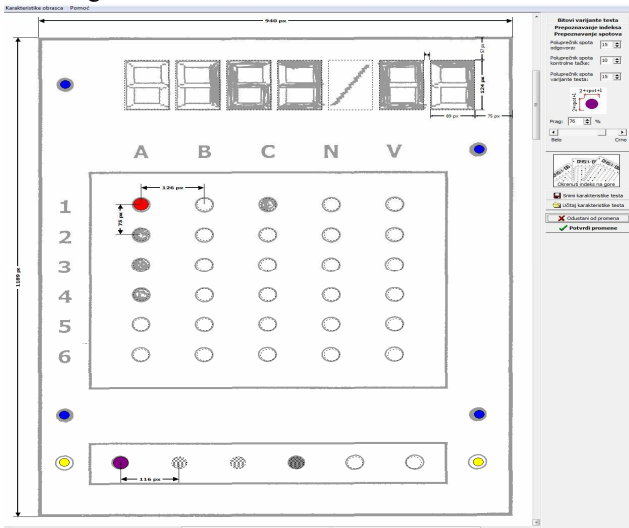
rao da koristi lenjir i unosi vrednosti u milimetrima, a nere-
tko i u desetim delovima milimetra. Zato je odlučeno da
unutar programa jedinica mere više ne bude metar, već pi-
ksel, jer, po prirodi elektronske obrade, slika koja predsta-
vlja obrazac ima upravo piksel kao osnovnu jedinicu.

Ovaj deo programa ima preduslov da se, pre nego što se
pokrene podešavanje, učita obrazac koji je postavljen tako
da je indeks pozicioniran u gornjem desnom delu slike obras-
ca. Kao prvi korak pripreme za vizuelna metrička pode-
šavanja, program vrši potrebnu preciznu rotaciju obrasca,
kako bi i vertikalna i horizontalna stranica obrasca bila na
monitoru idealno vertikalna i horizontalna, što je preduslov
za ispravan dalji rad.

Korisniku su na raspolaganju vizuelne komponente koje
predstavljaju kotne strelice i spotove kontrolnih tačaka,
odgovora i bitova u šifri varijante testa. Kotne strelice mo-
gu biti horizontale ili vertikalne, sa opcijom da stoje sa le-
ve ili desne strane objekta na koji se odnose.

Prilikom prelaska u režim grafičkog podešavanja metri-
čkih karakteristika, obrazac će biti prikazan preko celog
ekrana, dok će sa desne strane korisniku biti na raspolaga-
nju brojčane karakteristike koje određuju relativne prago-
ve, prečnike spotova i korisne bitove u varijantama testa. U
kontekstu programa *testARS*, spot je kružna ili pravouga-
ona površina na obrascu koja može, ali ne mora biti zacr-
njena i na taj način u sebi nosi korisnu informaciju. U ovoj
verziji programa postoji mogućnost posebnog definisanja
poluprečnika spotova kontrolnih tačaka, bitova varijante
testa i odgovora. Radi vizuelne ugodnosti, boja obrasca
prilikom podešavanja postaje svetlo siva i preko obrasca
korisnik može mišem pomerati kotne elemente. Karakteris-
tike testa se, kao i u prethodnim verzijama, mogu snimati i
učitavati.

Tokom podešavanja metričkih karakteristika korišće-
njem miša, grafički je prikazano i kako izgledaju površine
cifara indeksa koje su definisane aktuelnim vrednostima
rastojanja od ivica, razmakom između cifara, kao i širinom
i visinom cifri indeksa. Na ovakav način, korisnik vizuel-
nim načinom proverava da li se površina poklapa sa sed-
mosegmentnim ciframa koje program treba da prepozna.
Značajni bitovi kod varijante testa su prikazani kao šrafi-
rani krugovi.



Slika 3 – Pregled vizuelnog metričkog podešavanja obrasca.

C. Funkcionalne operacije nad entitetima studenata

Imajući u vidu česte potrebe korisnika programa, reali-
zovane su operacije pretrage po imenu studenta ili njego-
vom indeksu, operacije uvoženja spiska studenata i izvo-
ženja rezultata ispita sa različitim formatima indeksa. Na
raspolaganju su svi formati koji su se u praksi pokazali po-
trebnim.

V. ALGORITAM PREPOZNAVANJA OBRASCA

Jezgro sistema obrade predstavlja algoritam prepozna-
vanja obrasca, koji ima zadatak da analizom skenirane sli-
ke testa, a poštujući određene parametre, odredi broj inde-
ksa, označene odgovore i varijantu testa. Ovaj algoritam je
nekoliko puta unapređivan.

Na samom obrascu postoje već pomenute osnovne logi-
čke celine. Za svaku korisnu površinu, posebno je defini-
san prag. Prag predstavlja procenat crnih piksela jedne po-
vršine koji je potrebno da bude dostignut pa da ta cela po-
vršina bude smatrana zatamnjenom. Površina može biti
kružnog oblika (kontrolne tačke, odgovori na pitanja, vari-
janta testa) ili pravougaonog oblika (segmenti u cifri inde-
ksa).

Ulazni podatak algoritma obrade je 2D matrica piksela u
koju je učitana slika obrasca. Prvi korak je pretraga za
pravougaonim okvirom u kome su odgovori na pitanja. Po-
trebna je izrazita fleksibilnost u ovoj sekciji analize, koja
se zasniva na praćenju ivice obrasca i otkrivanju uglova,
jer su obrasci najčešće iskošeni, a ivice daleko od idealnih
zbog fotokopiranja testova pred ispit i nesavršenosti proce-
sa skeniranja. Nakon pronalazjenja okvira, biće provereno
postojanje kontrolnih tačaka na mestima koja su definisana
u metričkim parametrima obrazaca. Ukoliko su kontrolne
tačke na traženom mestu, tada program pristupa otkrivanju
šrafiranih površina koje predstavljaju broj indeksa, odgo-
vore na pitanja i varijantu testa.

Odnos definisanih dimenzija širine i visine obrasca pre-
ma dimenzijama trenutno analiziranog obrasca, daje raz-
meru koja može biti primenjena na sve elemente metričkih
karakteristika testa.

U ovoj verziji je sa apsolutnih vrednosti pragova učinjen
prelaz na relativne (procentualne) vrednosti. Na ovaj način
je, imajući u vidu i definisanje metričkih karakteristika u
pikselima, omogućena maksimalna sloboda korisnika u
podešavanju karakteristika samog obrasca. Ukratko, pro-
gram će, bez intervencije korisnika, prema situaciji, pove-
ćati ili smanjiti poluprečnike bitnih površina i tako omogu-
ćiti skeniranje obrasca u bilo kojoj DPI rezoluciji ili još
bolje, obrasca koji je skaliran za bilo koji faktor u odnosu
na očekivana podešavanja.

Pored ispravljanja manjih grešaka u prepoznavanju ob-
rasca, posebno je rađeno na optimizaciji analize. Na osno-
vu podatka da li je slika uspravna ili horizontalna, program
donosi odluku sa koje strane počinje da traži pravougaoni
okvir obrasca. Imajući na umu da je okvir uvek u sredini
ispitnog lista, pretraga za okvirom kreće od sredine, pa na-
zmenično prema ivicama.

VI. ALATI KORIŠĆENI ZA RAZVOJ SOFVERA

Za razvoj je upotrebljen Borland Delphi 2005. Za potrebe čitanja TIFF slika uz podršku za dekompresiju crno belih slika komprimovanih FAX CCITT G3 ili G4 metodom iskorišćena je besplatna LibTiff C++ biblioteka, posebno prevedena i prilagođena korišćenju u Delphi okruženju. Skeniranje je i dalje realizovano korišćenjem TWAIN standarda.

Program je povezan i sa bibliotekom *testCore*, jezgrom nove generacije programskih alata sistema *test*. Između ostalog, ova biblioteka sadrži mehanizam za bodovanje testa proizvoljne strukture, tako da je bilo jednostavno podržati testove koje obrađuje *testARS*. Za pojednostavljenje usluga koje pruža *testCore* (po ugledu na projektni uzorak *Facade*), napravljen je posebni DLL kojeg *testARS* koristi za bodovanje pregledanog testa umesto svog algoritma za bodovanje, ako korisnik tako želi. Za razvoj ovog DLL korišćeno je razvojno okruženje Microsoft Visual Studio.

VII. REZULTATI

Stablo obrazaca je prilagođeno korisniku, pa je rad sa formularima jednostavan i intuitivan, a mnoge moguće nedoumice u obradi predupređene su koncepcijom programa, osmišljenom tako da podaci o obrascima i kandidatima u svakom trenutku budu konzistentni. Uvid u sve formulare jednog ispita i njihovo čuvanje na disku omogućava, pored boljeg praćenja napredovanja same obrade, i lakši rad tokom uvida u radove kandidata. Podrška za rad sa višestranim TIFF fajlovima (po pravilu dobijenim automatskim skeniranjem) smanjuje mogućnost grešaka uzrokovanih upotrebom ručnog skenera i ubrzava proces pregledanja.

Korisniku je višestruko olakšano podešavanje metričkih karakteristika obrasca, jer je sada moguće da, samo upotrebom miša, bez lenjira, kotnim elementima preko slike obrasca precizno definiše sve neophodne osobine testa.

Nakon implementiranja ubrzanja u pristupu pikselima bitmap slike za tipičan obrazac postignuto je skraćivanje vremena obrade za oko 70% (sa prosečno 1.3 sekundi na 0.5 sekundi na prosečnom Pentium IV računaru).

Kod višestranih TIFF fajlova, prelazak na naredni obrazac je praktično trenutni. Svaka moguća intervencija operatera (na primer, ispravka loše prepoznatog indeksa) nakon analize obrasca traje do 5 sekundi. Obrada pojedinog obrasca traje od 3 sekunde (najčešći slučaj, kad operater samo proveriti da li je sve u redu) do 15 sekundi. Praksa pokazuje da srednja vrednost vremena obrade jednog obrasca iznosi 5 sekundi po obrascu. Nove funkcionalnosti programa uklanjaju ili umanjuju uzorke grešaka diskutovanih u [7], te iskustvo operatera znatno manje utiče na brzinu i tačnost rada. Do greške pri prepoznavanju indeksa dolazi jednom na oko 250 obrazaca. U toku upotrebe nove verzije programa na osnovu žalbi studenata nije konstatovana nijedna greška pri prepoznavanju izabranih odgovora.

VIII. ZAKLJUČAK

U ovoj verziji *testARS* je pored unapređenja načina opisanja i analize pojedinačnog obrasca, rađeno i na podršci za manipulisanje i analizu zbirke obrazaca, koje su opšta

karakteristika sprovođenja testova.

Iako su početna iskustva u korišćenju *testARS* 3.0 veoma dobra, i dalje postoji niz ideja za dalja unapređenja programa. Posebno bi bilo korisno razviti korisnički sistem sa različitim ulogama i njihovim privilegijama. Neke opcije koje bi dodatno unapredile vizuelni rad su: napredno uvećanje i umanjenje obrasca, specijalno obeležavanje (ikone) u stablu obrazaca, veća automatizacija obrade cele zbirke, te upoređivanje sličnosti između odgovora studenta radi kontrole mogućih malverzacija. Zbog činjenice da je algoritam bodovanja moguće poveriti spoljnoj komponenti, neće biti veliki problem naredne verzije programa učiniti opštijim, dodavanjem podrške za nove tipove pitanja sa ponuđenim odgovorima uz proizvoljan broj ponuđenih odgovora na svakom pitanju.

ZAHVALNICA

Autori rada duguju veliku zahvalnost dipl. inž. Dejanu Rizvanu koji je u toku višemesečnog razvoja nove verzije *testARS* razvojnim revizijama i finalnom revizijom nove verzije programa obradio preko 3500 obrazaca, ukazao nam na propuste u početnim zamislima i dao nekoliko vrednih saveta kako da *testARS* učinimo još ugodnijim za rad.

REFERENCE

- [1] D. Bojić, J. Protić, I. Tartalja, „test: programski alat za podršku pripreme testova, bodovanja rezultata i samostalnog učenja studenata“, *Zbornik XLV konferencije ETRAN-a*, Bukovička Banja, 2001.
- [2] I. Tartalja, J. Protić, „Razvoj programskog alata za automatsko pregledanje testova korišćenjem personalnog računara i ručnog skenera“, *Zbornik radova XXXVII konferencije ETAN-a*, Vol. 9, Beograd, 1993, str. 279-284.
- [3] P. Opačić, „Integracija programskog alata testARS za automatsko pregledanje testova korišćenjem personalnog računara i ručnog skenera u MS Windows 95 okruženje“, diplomski rad, Elektrotehnički fakultet Univerziteta u Beogradu, 1998.
- [4] P. Opačić, I. Tartalja, J. Protić, „testARS: programski alat za automatsko pregledanje testova korišćenjem skenera na MS Windows platformi“, *Zbornik radova XLII konferencije ETRAN-a*, Vrnjačka banja, 1998.
- [5] G. Anucojić, „testARS: Razvoj alata za bodovanje optički čitanog test-obrasca sa identifikacijom kandidata“, diplomski rad, Elektrotehnički fakultet Univerziteta u Beogradu, 2003.
- [6] P. Jovanović, „testARS 3.0: Razvoj nove verzije alata za pregled zbirke optički čitanih testova“, diplomski rad, Elektrotehnički fakultet Univerziteta u Beogradu, 2008.
- [7] A. Bošnjaković, J. Protić, I. Tartalja, „Unapređenje alata *testBase*, *testMix* i *testARS* za pripremu i pregled testova znanja“, *Zbornik radova 14. telekomunikacionog foruma TELFOR*, Beograd, 2006.

ABSTRACT

Paper describes improvements of the testARS software for optical recognition and scoring of tests. Presented are new functionalities that improve quality and usability of the software, and solutions to problems encountered in several years' long production use of previous versions. Paper focuses on scoring of tests with hundreds of candidates, with support for multiple ways of scanning of filled forms. Paper briefly introduces most important usage concepts of the new version of the software, with support for collections of forms as the most important feature.

testARS: SCORING OF COLLECTIONS OF OPTICALLY ACQUIRED TEST FORMS

Predrag R. Jovanović, Andrija M. Bošnjaković, Igor I. Tartalja