

Mr. Srđan Dumičić

UDK: 31:681.3

Stručni rad

Republički zavod za statistiku
SR Hrvatske, Zagreb,

KONTROLA I ISPRAVLJANJE PODATAKA PRIKUPLJENIH STATISTIČKIM ISTRAŽIVANJIMA

U ovom radu prikazano je jedno rješenje za povećanje produktivnosti rada na dizajnu, realizaciji i održavanju računalskih aplikacija namijenjenih kontroli i ispravljanju podataka prikupljenih statističkim istraživanjima. Analizirana je organizacija rada na provođenju velikog broja (oko 200) statističkih istraživanja i predložena je organizacija primjerenija raspoloživoj informatičkoj opremi. U nastavku je, za proces kontrole i ispravljanja, definiran jedinstven dizajn računalske aplikacije u kojoj su svi uočeni postupci uključeni. Dan je i model ulaznih podataka koji se koriste za opis onih karakteristika po kojima se istraživanja razlikuju. Unos i kontrola tih podataka realiziran je uz korištenje relacijske baze podataka CA UNIVERSE. Za dio jednostavnijih postupaka napisani su programski paketi koji se malim brojem parametara prilagođuju konkretnim istraživanjima. Za najsloženije postupke, kontrolu konzistentnosti i on-line ispravljanje podataka, razvijeni su generatori programa. Ulaz u generatore programa su parametri koji opisuju konkretno istraživanje, a izlaz je programski kod (PL/I) za kontrolu i ispravljanje podataka tog istraživanja. Iskustvo u jednogodišnjoj primjeni standardiziranog dizajna standardizirane programske podrške pokazuje da se oni mogu koristiti kod 80% statističkih istraživanja.

Statističko istraživanje; kontrola podataka; generator programa.

UVOD

U poslovima vezanim uz korištenje računala već je duže vremena prisutan trend da je količina novih zahtjeva za izradom programske podrške bitno veća od količine istovremeno napravljenih programske podrške. Time se stalno povećava obujam poslova koje tek treba napraviti, a i produžava se vrijeme čekanja do njihove realizacije. Dio tako nastalih problema želi se riješiti uvođenjem u rad programskih jezika i alata koji

osiguravaju bitno veću produktivnost u realizaciji aplikacija, kao i onih koji omogućavaju krajnjim korisnicima da samostalno, bez posredstva informatičara, rješavaju svoje zadatke na računalu. Tako smo sada svjedoci vrlo brzog razvoja CASE alata (computer-aided software engineering) koji pružaju podršku informatičarima u svim fazama nastanka aplikacije: od najranijih faza dizajna i planiranja pa sve do generiranja programskog koda na kraju tog procesa.

Problem djelotvornosti u izradi programske podrške postoji i u statističkim zavodima. Oni su zaduženi za provođenje više stotina statističkih istraživanja koja se tokom vremena mijenjaju pa je godišnje potrebno realizirati programsku podršku za pedesetak novih ili bitno izmijenjenih istraživanja. Pri tome se, prvenstveno zbog velikog obujma podataka koji se obrađuje, zahtijeva i visoka efikasnost u fazi korištenja dobivene programske podrške na računalu.

Programska podrška za obradu podataka prikupljenih statističkim istraživanjima mora pokriti dvije osnovne faze rada: kontrolu i ispravljanje podataka i izradu rezultata (analitičku obradu). Za potrebe izrade rezultata razvijena je, već duže vrijeme, kvalitetna programska podrška od strane niza komercijalnih proizvođača, a dijelom (npr. izrada statističkih tabela) i u okviru nekih većih statističkih zavoda. No, za područje kontrole i ispravljanja podataka još ne postoje programski alati koji bi bitno olakšali i ubrzali razvoj i održavanje odgovarajućih aplikacija. To se, s jedne strane može objasniti činjenicom da se radi o aplikacijama koje su specifične za mali segment tržišta, praktično samo za institucije koje se bave statističkim istraživanjima, a s druge strane, radi se o problemeima koji su složeniji od izrade statističkih tabela, pa u okviru samih statističkih zavoda još nisu u potpunosti i kvalitetno riješeni.

Dodatni je problem, dijelom obrađen u ovom radu, potreba da se korisnicima, statističarima koji su vlasnici podataka, omogući da samostalno upravljaju cijelim tokom obrade podataka. U tu svrhu potrebno im je osigurati dovoljno informacija o toku obrade i statusu njihovih podataka tako da oni mogu donositi kvalitetne odluke o daljnjim aktivnostima.

U ovom radu je prikazano jedno rješenje ovog problema. Ono je prilagođeno uvjetima koji postoje u Republičkom zavodu za statistiku SR Hrvatske. Taj Zavod je zadužen za provođenje 220- 250 statističkih istraživanja godišnje, od kojih se oko 70 obrađuje uz korištenje računala, s ciljem da se taj broj u idućih par godina bitno poveća. Godišnje se unese oko tri miliona slogova prosječne dužine 80 znakova. Na 20-30% slogova postoje pogrešni podaci. Njih je potrebno "otkriti" i zatim ispraviti. Sva istraživanja obrađuju se centralizirano u Zavodu za cijelu Hrvatsku. Raspoloživa oprema je IBM 4381/12 s 32 Mb pod MVS/SP operativnim sistemom.

I. ORGANIZACIJA RADA NA KONTROLI I ISPRAVLJANJU

Kod izrade programske podrške za kontrolu i ispravljanje podataka prikupljenih statističkim istraživanjima potrebno je uvažiti slijedeće karakteristike:

- brza koncentracija,
- velik broj podataka,
- podaci se koriste za statističke, grupne obrade.

Brza koncentracija i velik broj podataka ukazuju na potrebu da prvi prijenos podataka s obrazaca na magnetski medij obavljaju profesionalni radnici u odjelu za unos podataka. Da bi se postigla zadovoljavajuća brzina, u toku unosa se postavljaju samo one kontrole na koje radnici unosa mogu reagirati bez velikog gubitka vremena. Takav postupak nameće postojanje programskog rješenja za provođenje kontrole u batch modu. Pokazalo se, a i u skladu je s mogućnostima suvremenih računala, da ispravljanje treba provoditi kao on-line aplikaciju. Pri tome je poželjno da se u toku ispravljanja provode i sve potrebne kontrole kako bi se izbjegle nove pogreške. Ako to nije osigurano, mora se nakon ispravljanja ponovno izvesti batch program kontrole.

Činjenica da se prikupljeni podaci, u pravilu, ne koriste individualno, već samo za statističke obrade, donekle smanjuje potrebu za njihovom individualnom točnošću. Treba uočiti da je taj zahtjev kod većine drugih, na primjer poslovnih i bankarskih, sistema izrazito naglašen te da je to jedna od značajnih specifičnosti statističkog sistema. Zahtjev za individualnom točnošću ublažuje se tako da svaki podatak mora zadovoljavati uvjete kontrole postavljene od strane predmetnih statističara, ali na individualnom nivou ne mora uvijek odgovarati činjeničnom stanju. Posljedica toga je ta da se ispravljanje uočenih pogrešaka, kao i umetanje nedostajućih vrijednosti, ne mora uvijek rješavati ponovnim kontaktom s izvorom podataka, već se mogu koristiti različite determinističke i probabilističke metode. Te metode je gotovo nemoguće koristiti bez upotrebe računala, pa proizlazi zahtjev da se kod izrade programske podrške za ispravljanje omogući i njihova primjena.

Da bi se povećala produktivnost radnika Zavoda u cjelini, poslove kontrole i ispravljanja je potrebno analizirati i podijeliti, ne samo po sadržaju statističkih istraživanja kao što je uobičajeno, već i po složenosti i potrebnom znanju za njihovo obavljanje. Uočavaju se dvije grupe poslova:

- složeni, inicijalni poslovi kod definiranja postupka kontrole i ispravljanja za pojedino istraživanje; oni se obavljaju jednokratno i zahtijevaju dobro poznavanje pojave na koju se istraživanje odnosi i
- jednostavni, operativni poslovi koji se ciklički ponavljaju u toku distribucije, prikupljanja te kontrole i ispravljanja podataka.

Tok poslova na kontroli i ispravljanju podataka može se jednostavno prikazati slikom (slika 1). Taj tok odgovara kretanju obrazaca od dolaska u Zavod do završetka poslova na kontroli i ispravljanju. On je istovjetan za veliku većinu istraživanja.

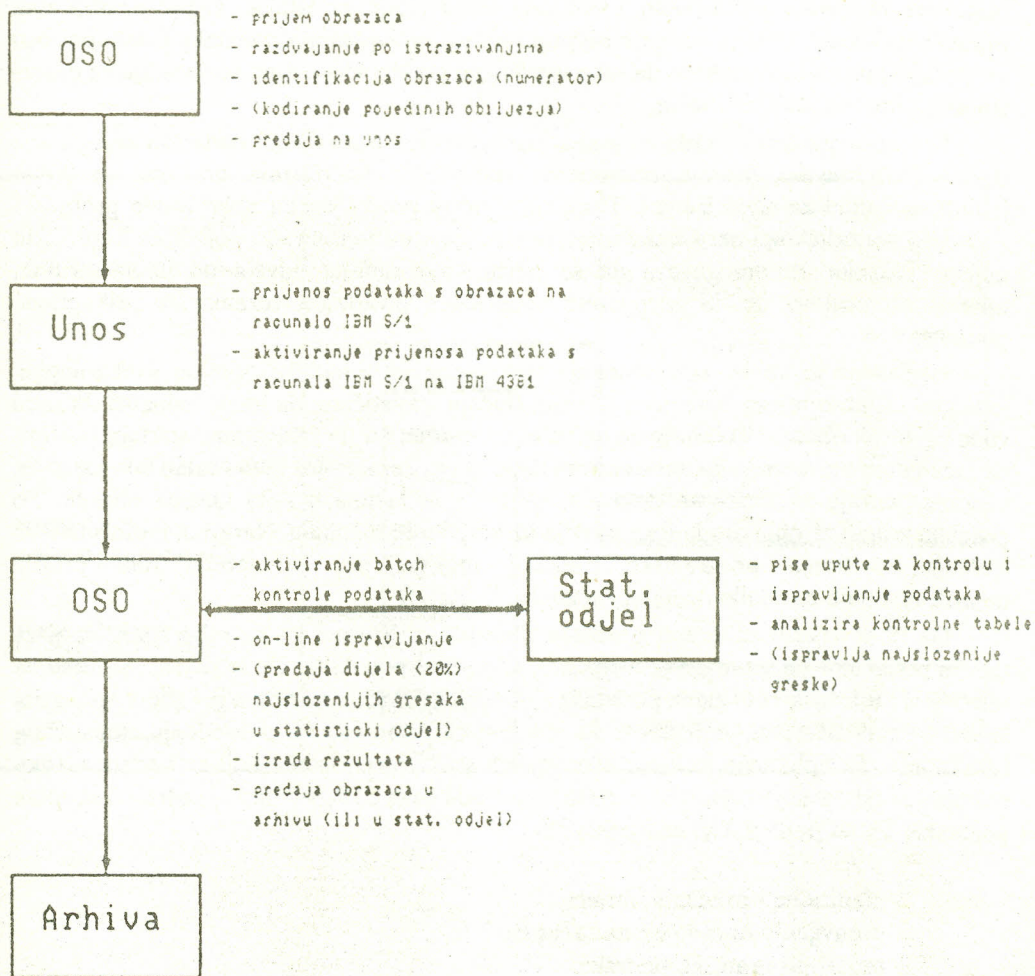
Na slici je naznačeno da u tim poslovima malen udio imaju odjeli u kojima rade najstručniji radnici. No, bitan preduvjet za takvu organizaciju rada je taj da upravo ti najstručniji radnici, dobri poznavaoци pojava koje se prate statističkim istraživanjima, detaljno opišu sve postupke koji se javljaju u provođenju kontrole i ispravljanja prikupljenih podataka. Takvi opisi moraju imati standardan oblik i predstavljaju osnovu za izradu potrebne programske podrške, kao i za provođenje svih postupaka u toku kontrole i ispravljanja. Pored toga, to je i dio dokumentacije o istraživanju koja statističke odjele i računski centar čini manje ovisnima o radnicima koji su radili na nekom istraživanju.

Prijem obrazaca i priprema za unos obavlja se u Odjelu statističke obrade (OSO) za većinu istraživanja u Zavodu. Radnici tog Odjela nemaju specijalistička znanja o pojavama na koje se istraživanja odnose. Naravno, za svako istraživanje postoji precizan opis kako ga treba obrađivati, a pored toga održava se i kratka instruktaža na samom početku obrade pojedinog istraživanja. Ta faza obuhvaća poslove prijema svih obrazaca, njihovo razdvajanje po istraživanjima i predaju u Odjel za unos podataka. Teži se tome da pristigli izvještaji idu direktno na unos, bez ikakve dodatne intervencije u Zavodu, no nekad je potrebno obaviti šifriranje onih obilježja za koja to nikako nije moguće provesti na terenu. To u nekim slučajevima mogu biti: šifre naselja, zanimanja, škola i sl. Pored toga, u OSO se numeracijom identificiraju pristigli izvještaji i time olakšava operativni rad s njima. Ta se identifikacija nalazi na svim izvještajima na istom mjestu i razlikuje se od prave, sadržajne identifikacije.

Unos podataka obavljaju profesionalni radnici u Odjelu za unos podataka, i to centralizirano za sva istraživanja. U toku unosa koriste se samo one kontrole koje ne usporavaju rad radnika na unosu i kod kojih takvi radnici mogu vršiti odgovarajuće intervencije. Takav skup kontrola je, najčešće, malen.

Ispravljanje podataka obavlja se, pretežnim dijelom, u odjelu OSO. To je, ponovno, moguće samo ako su prethodno precizno definirani i opisani postupci ispravljanja pogrešaka, za koje su ti radnici zaduženi. U takvoj podjeli poslova stručni radnici statističkih odjela rješavaju samo najsloženije pogreške i pomoću kontrolnih tabela analiziraju prikupljene podatke. Oni mogu otkriti pojavu sistematskih grešaka i eventualne nedostatke u skupu zadanih uvjeta kontrole. Nakon završetka ispravljanja cijelog ili dijela materijala radnici odjela OSO samostalno startaju batch kontrole i po potrebi nastavljaju s ispravljanjem. Nakon što je materijal potpuno čist, radnici odjela OSO startaju programe za izradu rezultata i predaju obrasce u arhivu ili u statistički odjel.

OSNOVNI TOK OBRAZACA



Slika 1.

Kad god je moguće točno definirati načine korekcije pojedinih pogrešaka, koji precizno mogu biti deterministički ili stohastički, onda je potrebno pripremiti programsku podršku koja takve korekcije i provodi. To je potrebno učiniti uvijek tada kada ne postoji mogućnost da se odredi istinita vrijednost za nedostajući ili netočan podatak, jer se tako osigurava korektno, jedinstveno i efikasno ispravljanje podataka. Takav pristup ima opravdanja i onda kada je traženje potpuno točnih individualnih podataka preskupo, kao i u slučajevima kada se pokaže da tako ispravljeni podaci kvalitetom ne zaostaju za onima ispravljenima na klasičan način.

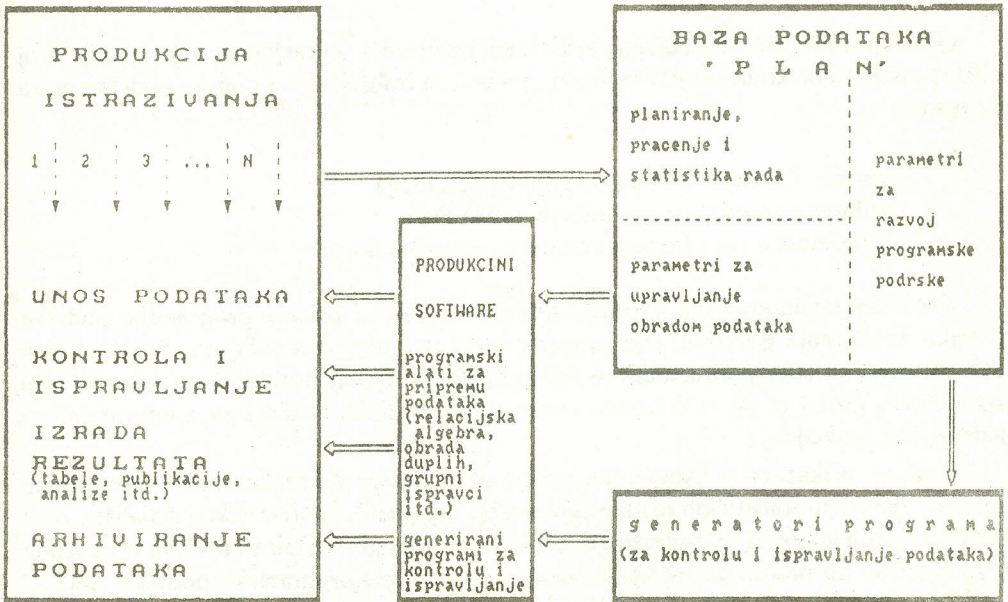
Ovakva organizacija rada na poslovima kontrole i ispravljanja podataka omogućava da se jednostavniji poslovi provođenja statističkih istraživanja prebace na jedan jedinstveni odjel za cijeli Zavod. Time su stvoreni preduvjeti za ublažavanje problema koji zbog periodičkog i neravnomjernog opterećenja nastaju u radu pojedinih statističkih odjela. Također se omogućava da se najstručniji radnici odvoje od jednostavnijih, operativnih poslova, te da se posvete analitičkim poslovima vezanim za prikupljene podatke.

Naglašeno je da je, kod uvođenja suvremene informatičke opreme i tehnologije, jedan od bitnih zahtjeva korisnika, u ovom slučaju statističara, da im se omogući da sami vode cijeli tok obrade. To znači da im se mora omogućiti da samostalno startaju poslove na računalu u trenutku kada oni smatraju da je to opravdano. Da bi to realno bilo moguće, korisnici moraju znati što se dešava s njihovim podacima u svim fazama obrade. To posebno vrijedi za onaj dio koji se obavlja uz korištenje računala. Naravno, i informatička rješenja korištenja programske podrške moraju biti primjerena tom profilu neprofesionalnih korisnika velikog računala.

Da bi se riješili problemi praćenja obrade, analizirane su potrebe statističara za takvim podacima i napravljen je odgovarajući model podataka. Uz korištenje sistema za upravljanje relacijskim bazama podataka CA-UNIVERSE realizirana je i cijela aplikacija kojom se statističarima omogućava da planiraju i prate rad na provođenju statističkog istraživanja. Ta aplikacija je kasnije proširena s nizom podataka koji se koriste u toku nastajanja i održavanja računalskih aplikacija. Sada baza PLAN (slika 2) sadrži tri skupine podataka. To su podaci koji se koriste za:

- planiranje i praćenje obrada,
- upravljanje obradama na računalu i
- razvoj programske podrške.

Planski podaci i opis istraživanja za potrebe razvoja programske podrške upisuju sami korisnici u fazi postavljanja istraživanja, a većina podataka za praćenje preuzima se automatski s računala. Svaka obrada ostavlja u bazi poruku o vremenu kada je rađena, a po potrebi daje i osnovne podatke o rezultatu obrade.



Slika 2.

2. STANDARDNI DIZAJN KONTROLE I ISPRAVLJANJA

Uz opisanu organizaciju poslova na kontroli i ispravljanju podataka definirane su i osnovne karakteristike aplikacije kojom će se oni podržati na računalu. To su:

- kontrole se provode u batch modu (barem prva kontrola nakon unosa),
- ispravljanje, po mogućnosti s ugrađeno kontrolom, se provodi on-line i
- na unosu se provode samo osnovne kontrole.

Obim poslova na postavljanju aplikacija kontrole i ispravljanja, kao i velik broj podataka koje treba kontrolirati i ispraviti, povlači da izabrana programska podrška mora osigurati:

- visoku efikasnost u fazi postavljanja aplikacija,
- jednostavno održavanje gotovih aplikacija i
- efikasnost u svim fazama korištenja na računalu.

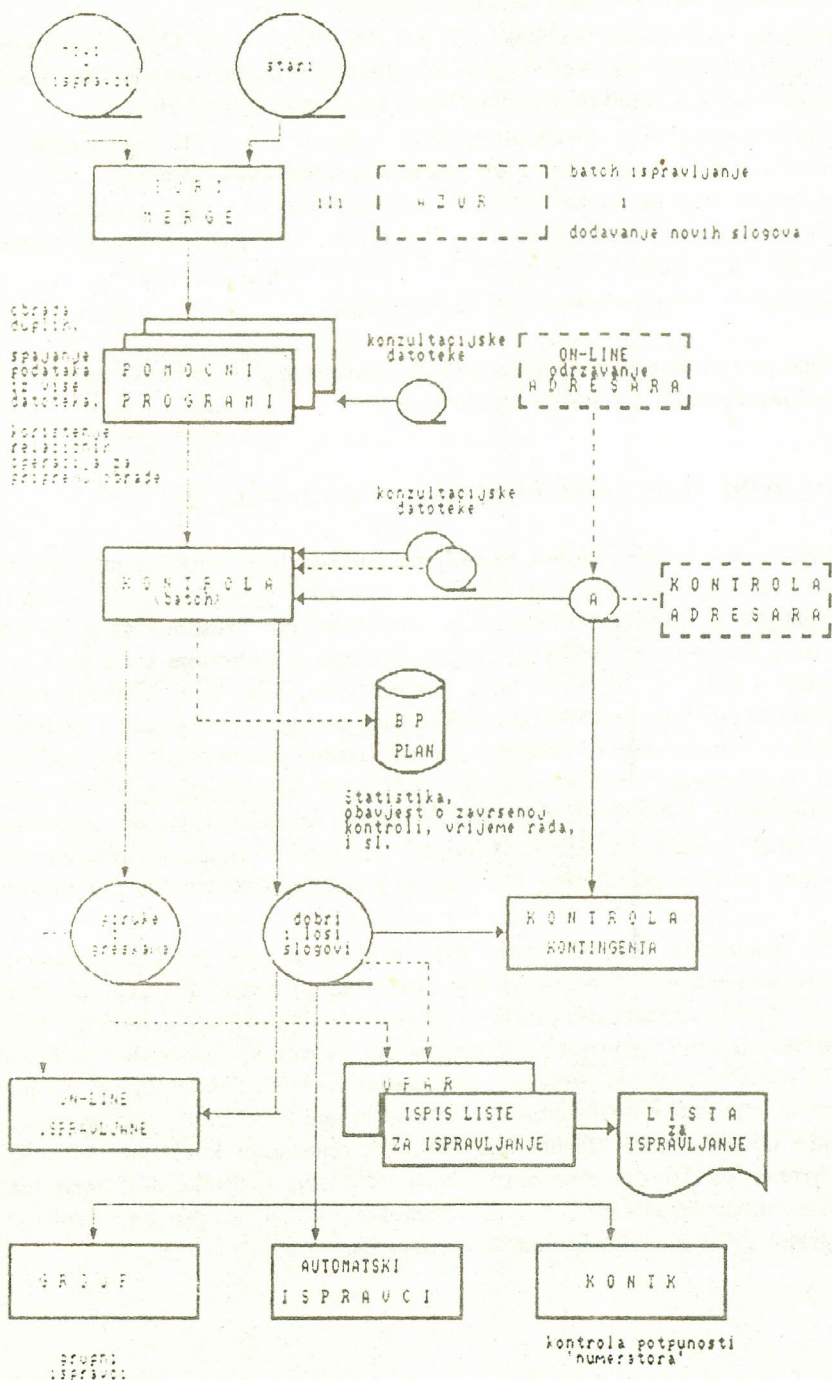
Klasičnim načinom rada, koji podrazumijeva dizajn i pisanje programske podrške za svako istraživanje posebno, nije moguće realizirati prva dva zahtjeva. Problem je u tome da takav pristup podrazumijeva dekompoziciju cijelog problema na nivo naredbi programskog jezika, te da se tada često rade nova i različita rješenja za, s informatičkog gledišta, iste funkcije.

Rješenje prikazano u ovom radu polazi od činjenice da se kontrola i ispravljanje podataka za većinu statističkih istraživanja sastoji od, opet s informatičkog stajališta, istih postupaka. Njih je potrebno prepoznati, precizno ih opisati i definirati njihove veze. Tada je, za svaki od tih postupaka, moguće napisati jedinstvenu programsku podršku, koja se zatim parametrima prilagođava potrebama pojedinih istraživanja.

Nakon toga potrebno je kod izrade programske podrške za konkretno istraživanje potrebno napraviti njegovu dekompoziciju na postupke koji su podržani općim programskim rješenjima, te od njih napraviti programsko rješenje cijele aplikacije. Takav pristup osigurava zadovoljenje zahtjeva za brzim postavljanjem i jednostavnim održavanjem aplikacija kojima je podržana kontrola i ispravljanje podataka.

Analiza procesa kontrole i ispravljanja, provedena u našem Zavodu, dovela je do jednog standardnog dizajna takve aplikacije (slika 3). Na njemu su uočeni postupci koji se, s malim varijacijama, koriste u mnogim istraživanjima. Za dio tih postupaka napisana je programska podrška opće namjene.

STANDARNI DIZAJN KONTROLE I ISPRAVLJANJA



Trenutno su u upotrebi tri grupe programskih alata. To su:

- programi za pripremu podataka (UPAR, DUPLI, SORT). Oni pokrivaju dio funkcija relacijske algebre i mogu se koristiti u svim fazama obrade. Pozicijskim parametrima prilagodavaju se potrebama konkretnih istraživanja;
- programi za određene probleme (AZUR, GROUP, KONTK) koji su napisani tako da rješavaju neki konkretan problem u statističkim obradama. AZUR - batch ažuriranje podataka, GROUP - grupno ispravljanje (sistematske greške) i KONTK - kontrola kontingenata. Pojedinom istraživanju prilagodavaju se skupom parametara koji se koriste u trenutku izvođenja programa i
- generatori programa za kontrolu i ispravljanje.

Sve tri grupe programa namijenjene su informatičarima koji ih koriste u toku izrade programske podrške za neko statističko istraživanje.

3. GENERATORI PROGRAMA KONTROLE I ISPRAVLJANJA

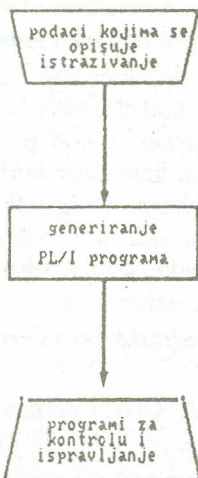
Programi za batch kontrole i on-line ispravljanje predstavljaju najsloženije, a ujedno i performansno najzahtjevnije dijelove cijele programske podrške za kontrolu i ispravljanje podataka prikupljenih statističkim istraživanjima. Izradom programskih generatora, a ne gotovih programskih paketa, omogućeno je dobivanje koda po mjeri pojedine aplikacije i time su osigurane bolje performanse rada. U slučajevima kada generirani program ne pokriva sve očekivane funkcije, moguće ga je nadopuniti potrebnim programskim kodom i time osigurati program po želji. Izrada takvih programa olakšana je činjenicom da svi generirani programi imaju istu strukturu, da su pisani uz pridržavanje istih standarda, te da je dio koda koji je generiran sigurno formalno ispravan. Činjenica je da u mnogim istraživanjima postoje specifičnosti koje je teško i skupo ugraditi u opća rješenja, pa ovakav pristup omogućava djelotvorno korištenje generatora i u takvim slučajevima.

Kod izrade generatora promatraju se dva različita, a ipak povezana procesa (slika 4). Ciljani proces je opisan funkcijama koje generirani programi obavljaju, ulaznim podacima koje oni koriste i sadržajem i formom izlaza koje daju. U inicijalnom procesu koriste se programi kojima je osnovna funkcija da generiraju programe traženih karakteristika. Ta osnovna funkcija programa generatora može se dekomponirati na niz funkcija nižeg nivoa koje zavise o funkcijama koje podržavaju generirani programi. One se mogu općenito opisati ovako: generiranje koda za realizaciju konkretne funkcije generiranog programa; na primjer, generiranje koda za čitanje datoteke ili generiranje koda za kontrolu numeričnosti itd... U tom kontekstu mogu se povezati funkcije generiranih programa i dio funkcija generatora programa.

GENERATORI PROGRAMA

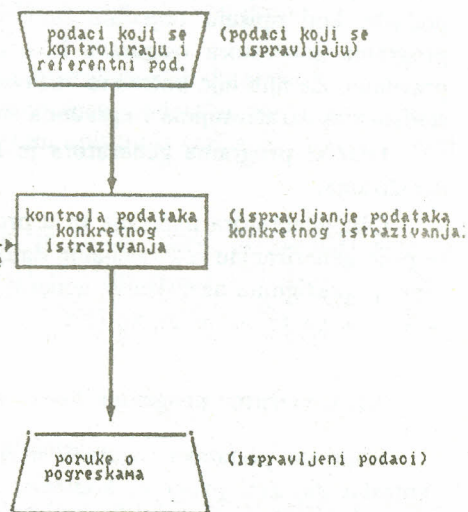
Generiranje programa

(inicijalni proces)



Koristenje generiranih programa

(ciljani proces)



Slika 4.

Zato su kod izrade generatora najprije definirane sve funkcije za koje se očekuje da će ih generirani programi morati podržavati. Naravno, svaki pojedini generirani program podržava samo dio tih funkcija. Zatim su određeni podaci, potrebni u fazi generiranja programa, koji omogućavaju da se sve predviđene funkcije kodiraju. Ti podaci opisuju karakteristike istraživanja za koje se generiraju konkretni programi kontrole i ispravljanja. Zbog performansi je traženo da generirani programi budu kodirani u programskom jeziku PL/I. Time je, u osnovi, zadan i rad generatora programa. I za generator kontrole i za generator ispravljanja vrijedi isti osnovni dizajn. Ulaz u programe generatore čine podaci koje opisuju karakteristike pojedinih istraživanja značajnih za kontrolu i ispravljanje. Upis i održavanje ulaznih podataka za oba generatora realiziran je uz korištenje sistema za upravljanje relacijskim bazama podataka CA-UNIVERSE. Podaci su smješteni u istim relacijama i dio podataka koriste oba generatora. Oni čine dio kompleksne baze podataka u kojoj se vodi niz podataka o statističkim istraživanjima za koja je zadužen naš Zavod. Model tog dijela baze prikazan je u nastavku rada (slika 5).

Osnovna funkcija programa generatora je ta da kodira program koji realizira sve tražene funkcije za istraživanje opisano ulaznim podacima. Ta funkcija može se podijeliti na dva osnovna podskupa. Jedan čine funkcije kojima se generira kod programa za realizaciju funkcija ciljanog programa. U tom dijelu program generator koristi ulazne podatke koji opisuju pojedinu funkciju konkretnog istraživanja. Drugi dio funkcija programa generatora osigurava kodiranje PL/I programa u skladu s odgovarajućim pravilima. Za njih nije potrebno zadavati nikakve ulazne podatke, jer ona nisu zavisna o statističkim istraživanjima i ugrađena su direktno u programe generatore.

Izlaz iz programa generatora je PL/I kod traženih programa po mjeri opisanog istraživanja.

Oba generatora napisana su u programskom jeziku PL/I. Ulazni podaci prepisuju se prije generiranja u sekvencijalne datoteke i programi generatori ih iz njih preuzimaju. Time je postignuta nezavisnost generatora od sistema za upravljanje relacijskim bazama podataka, pa se oni mogu koristiti i za slučajeve koji nisu opisani u bazi.

3.1. Generator programa kontrole

Kontrola podataka, opisana u ovom radu, realizirana je kao batch aplikacija. Centralni dio čini program kontrole. On učitava slogove osnovne datoteke dok god pripadaju istom obrascu i razmješta ih u odgovarajuće tabele. U programu je deklarirana struktura ulaznih podataka koja odgovara strukturi obrasca i na kraju su u nju učitani svi podaci jednog obrasca istovremeno. Nakon toga se kontrolira da li podaci zadovoljavaju sve zadane uvjete kontrole. Ako ne zadovoljavaju, najprije se, za pogreške za koje je to zadano, vrši automatsko ispravljanje, a zatim se ispisuju poruke za preostale pogreške.

Ulaz u program je datoteka s podacima koji se kontroliraju. Ako je potrebno, to mogu biti i datoteke koje sadrže šifarnike koji se koriste u kontroli.

Funkcije programa kontrole jesu:

- prepoznavanje hijerarhije obrazaca do četiri nivoa (obrazac, tabela, redak i polje) i rad s cijelim obrascima koji mogu biti prikazani nizom slogova različite strukture u ulaznoj datoteci,
- prepoznavanje nekonzistentnosti u podacima ulazne datoteke,
- provođenje zadanih determinističkih ispravaka,
- izvještavanje o otkrivenim pogreškama i
- statistika o radu programa.

Izlaz iz programa kontrole su dvije datoteke. Jedna sadrži kontrolirane i determinističkim ispravkama izmijenjene ulazne podatke. Druga sadrži poruke o otkrivenim pogreškama i statistiku o radu programa. Korespondentni slogovi prve i druge datoteke povezani su rednim brojevima dodanim u programu kontrole.

Generator programa za kontrolu generira program kontrole opisanih karakterisitka. Za to koristi ulazne podatke koji se mogu podijeliti u dvije osnovne grupe. To su:

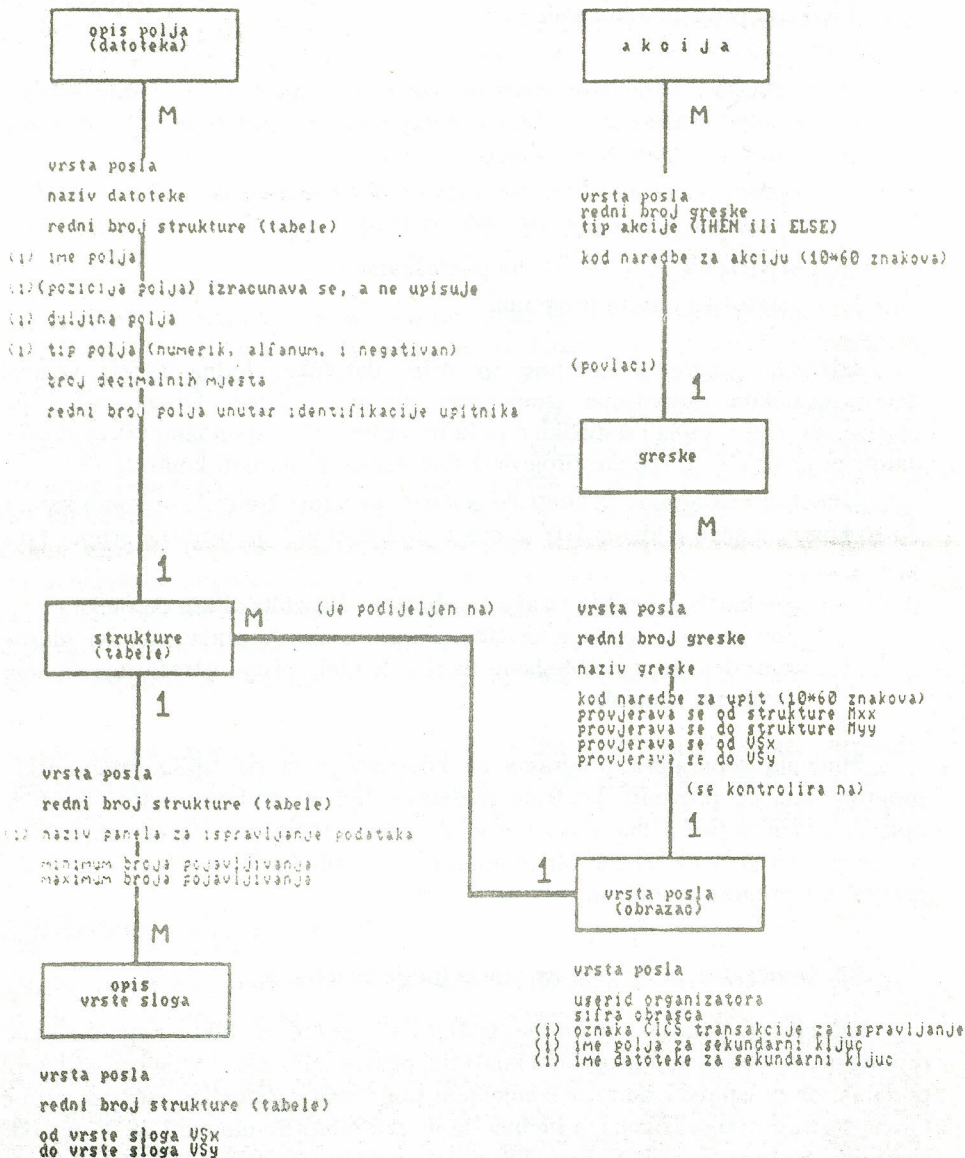
- podaci koji opisuju strukturu obrasca i konzultacijskih datoteka i
- podaci koji opisuju uvjete nekonzistentnosti podataka. Svi ulazni podaci navedeni su u ERM shemi na slici 5. Oni opisuju istraživanje za koje se radi program kontrole.

Funkcija generatora programa za kontrolu je ta da ispiše programski kod za program koji će provoditi kontrolu podataka čija je struktura, kao i uvjeti kontrole, opisana ulaznim podacima u generator. Pri tome dobiveni program mora biti pisan u programskom jeziku PL/I i ne smije imati nikakav suvišan kod, a time je zadan i izlaz iz generatora programa kontrole.

3.2. Generator programa za ispravljanje podataka

Ispravljanje podataka je on-line aplikacija koja radi na nivou sloga datoteke. Preko rednog broja, dodanog u programu kontrole, poziva se na ekran slog koji sadrži pogrešan podatak, on se ispravi i zatim se izmijenjeni slog vraća u datoteku. Očito se radi o malim i jednostavnim transakcijama, a budući da ih ima relativno mnogo i da ih provode brojni korisnici, izabran je CICS kao TP za programe za ispravljanje. Oni su pisani u programskom jeziku PL/I, reentrant su i podržavaju pseudokonverzijski način rada.

E R M - shema za generatore kontrole i ispravljanja



Slika 5.

Time su uz najmanje opterećenje računala osigurane najpovoljnije performanse rada korisnika.

Generirani program za ispravljanje radi s datotekama materijala i poruka o pogreškama koje se dobiju kao izlaz iz programa kontrole. Slogovi s podacima i oni s porukama o pogreškama povezani su rednim brojevima dodanim u kontroli.

Ulazni podaci u generirani program za ispravljanje:

- datoteka s podacima koji se ispravljaju,
- datoteka s porukama o greškama. Program, u on-line načinu rada, podržava slijedeće funkcije:
- ispisuje poruke kojima olakšava rad korisnika,
- pretraživanje po ključu,
- pretraživanje po sekundarnom ključu,
- izmjenu,
- brisanje,
- dodavanje,
- pozivanje slijedećeg sloga,
- pozivanje prethodnog sloga,
- korištenje različitih ekrana za različite strukture slogova u datoteci.

Koristeći datoteku s porukama o pogreškama generirani program za ispravljanje podržava i slijedeće funkcije:

- pozivanje slijedećeg pogrešnog sloga i
- pregled pogrešaka o pojedinom slogu.

Izlaz iz programa za ispravljanje zavisi o aktivnosti korisnika. On može izmijeniti sadržaj datoteke osnovnih podataka, a sadržaj datoteke s podacima o pogreškama ostaje nepromijenjen do idućeg kruga kontrole. Pored toga izlazni podaci su i poruke koje program ispisuje na ekrane u toku rada korisnika.

Dobiveni programi ne sadrže module za kontrolu podataka, osim onih koji su za tu svrhu razvijeni u programskom paketu CA-UFORM pomoću kojeg su rađeni ekrani za ispravljanje. No, uvijek je moguće proširiti generirani program s programskim kodom za provođenje željenih kontrola i tako spriječiti nastajanje novih pogrešaka u fazi ispravljanja podataka.

Ulaz u generator programa za ispravljanje su podaci kojima se opisuje struktura datoteke za koju se radi ispravljanje. Pored toga to su i nazivi transakcije pod CICS-om i ekrana koji će se koristiti u ispravljanju. Svi ulazni podaci navedeni su u ERM shemi na slici 5. s oznakom I na početku.

Funkcija generatora programa za ispravljanje je ta da ispiše programski kod za program koji će omogućiti izvođenje svih ranije navedenih funkcija ispravljanja, i to za zadanu strukturu ulaznih podataka. Pored toga, generator mora ispisati reentrant PL/I program koji podržava pseudokonverzacijski način rada, a time je opisan i izlaz iz ovog generatora programa.

4. ZAKLJUČAK

Dosadašnja iskustva pokazuju da opisana organizacija rada omogućava veću djelotvornost u obavljanju poslova kontrole i ispravljanja podataka prikupljenih statističkim istraživanjima. To je prvenstveno rezultat činjenice da je ona napravljena na način koji u većoj mjeri koristi mogućnosti računala.

Kod dizajna programske podrške pokušala su se jedinstveno sagledati sva istraživanja i izdvojiti različite informatičke funkcije koje se javljaju kod većine njih. Za dio tih funkcija napisani su programski moduli. Oni se parametrima zadanim kod izvođenja prilagođavaju potrebama konkretnog istraživanja. Za programe logičko-računske kontrole i ispravljanja napisani su generatori programa. Time je većina poslova kontrole i ispravljanja pokrivena standardnom programskom podrškom opće namjene, što je i najbolji način da se osigura primjena usvojenih standarda.

U takvom pristupu svodi se dizajn kontrole i ispravljanja pojedinog istraživanja na dekomponiranje cijelog procesa na definirane standardne funkcije. Napravljen je i jedan opći dizajn tog procesa koji se može direktno primijeniti na mnoga istraživanja. Implementacija tih funkcija na računalo podrazumijeva određivanje i upisivanje potrebnih parametara. Samo za dio poslova koji se tako ne mogu riješiti piše se programska podrška po mjeri.

Korištenjem standardiziranog dizajna podržanog programskom podrškom opće namjene povećana je djelotvornost postavljanja takvih aplikacija i, posebno, njihovog održavanja. Budući da generatori programa daju programe po mjeri pojedine aplikacije, te da su pisani u jeziku treće generacije, osigurano je i efikasno izvođenje aplikacije na računalu.

U daljnjem razvoju želimo na efikasan način riješiti problem on-line aplikacija koje u sebi, pored mogućnosti ispravljanja, sadrže i istovremenu kontrolu izmijenjenih podataka.

Drugi pravac razvoja mora se usmjeriti na mjerenje efikasnosti provođenja svih potupaka kontrole i ispravljanja podataka, kao i utjecaja tih postupaka na kvalitet podataka koji se dobiju na kraju tog procesa. Tek na temelju tako dobivenih pokazatelja moguće je unaprijediti organizaciju rada na tim poslovima, a zatim i razviti potrebnu programsku podršku.

LITERATURA:

1. Dumičić, S. (1987). Kontrolni modul, Statističar, broj 13.
2. Dumičić, S. (1988). Grupne operacije s podacima u datotekama, X Međunarodni simpozij Kompjuter na sveučilištu
3. Lyberg, L. and Sundgren, B. (1987). The impact of the Development of EDP on Statistical Methodology and Survey Techniques, Statistics Sweden
4. Milonja, Z., Gredelj, M. i Dumičić, S. (1985). Koncept obrade podataka statističkih istraživanja, Bilten Statističkog društva SR Hrvatske, broj 11.
5. Milonja, Z. (1987). Primjena informatičkih tehnologija na neke funkcije u obradi statističkih istraživanja, Statističar, br. 13.
6. Rogan, Š.N. i Dumičić, S. (1987). Modul za ispravljanje podataka, Statističar, broj 13.
7. Shanks, J.M. (1989). Information Technology and Survey Research: Where Do We Go From Here?, Journal of Official Statistics, 5(1), pp.3-21.
8. Sundgren, B. (1985). Outline of an Algebra of Base Operators for Production of Statistics, Statistics Sweden
9. ... (1988). Izvještaj o rezultatima probnog uvođenja nove tehnologije obrade statističkih istraživanja, interna dokumentacija, Republički zavod za statistiku SR Hrvatske

Primljeno: 1990-07-06

Dumičić S. The Control and Correction of Data Gathering for Statistical Research

SUMMARY

The paper shows one solution for productivity improvement in designing and maintaining applications for statistical survey data checking and correction. About 200 surveys were analyzed and this led to a general design, suitable for generalized software. Checking and correction methods applied in those statistical surveys were also analyzed, which led to a general application design that includes all relevant methods. Along with the generalized software a data model was built to hold descriptions which are characteristic of every survey in particular. This part of software, including data entry and checking, was made by means of CA-UNIVERSE relational data base management system. A number of utility programs has been developed to deal with more simple operations (join, duplicates, contingencies etc.). These utilities are easily tailored for inclusion in every survey application by means of some control parameters. Program generators were developed for more complicated operations, such as consistency checking and on-line data correction. Program generator input is made of parameters describing the survey in particular, and the result is PL/I code for checking and correction of survey data. The program generators themselves are also written in PL/I. The program generators and utility programs have proved suitable for checking and correction of 80% of statistical surveys, according to one-year experience of our Statistical Office.