

CALITATEA DATELOR DE INTRARE ÎN APLICAȚII DISTRIBUITE

Ion Ivan

ionivan@ase.ro

Leonard Săcuu

leo.sacuiu@gmail.com

Daniel Milodin

daniel.milordin@ase.ro

Academia de Studii Economice, București

Rezumat: Se prezintă tipurile de date de intrare în aplicații distribuite, caracteristicile de calitate a datelor și interfețe folosite pentru procesele de creare a bazelor de date complete și corecte. Se stabilește o ierarhizare a caracteristicilor de calitate pentru datele de intrare și se identifică setul de cerințe pentru asigurarea, controlul și managementul calității datelor.

Pentru seturi de date foarte mari, se proiectează o structură software pentru implementarea componentelor de managementul calității datelor de intrare furnizate de utilizatori neomogeni.

Cuvinte cheie: date intrare, aplicații distribuite, caracteristici de calitate, certificare date.

Abstract: There are presented data entry types for distributed applications, quality characteristics for data, and interfaces used for processes of creating complete and concrete databases. It is established a hierarchy data entry characteristics quality and it is identified the set of request for assuring, control and management of data quality.

For very large data sets it is created a software structure for implementing the components of data entry management quality provided by the non-unitary users.

Key words: data entry, distributed applications, quality characteristics, data certification

1. Tipuri de date de intrare în aplicații informatice distribuite

Aplicațiile informatice distribuite reprezintă soluția viabilă pentru dezvoltarea informaticii moderne. Simpla folosire de aplicații în rețea de calculatoare nu creează premisele accesului nerestricționat al cetățenilor la resursele aplicațiilor, în vederea soluționării on-line de probleme.

Modul actual de a concepe interacțiunea om-calculator este încă tributat mentalității conform căreia experiența operatorilor asigură derularea corectă a fluxurilor de prelucrări. Într-un viitor foarte scurt, se va vedea că lipsa forței de muncă orientată spre calificarea de operatori va impune dezvoltarea de aplicații distribuite, care să se bazeze pe descentralizarea procesului de introducere a datelor, responsabilizând la maximum pe cei care derulează acest proces.

Sunt preferate aplicațiile distribuite în care diversitatea datelor de intrare este limitată prin:

- prezentarea nivelurilor pentru câmpuri prin enumerarea de valori din care utilizatorul să selecteze o singură valoare;
- utilizarea aceluiași mod de definire a intervalelor;
- validări extrem de severe a datelor introduse;
- introducerea de corelații care elimină neconcordanțele între valorile inițializate de utilizatori.

În cazul în care datele de intrare au un nivel de flexibilitate mai ridicat, fiind texte, se impune:

- construirea de vocabulare cu cuvinte acceptate a fi introduse în text;
- eliminarea cuvintelor din textele utilizate care nu aparțin vocabularului;
- construirea de liste ce conțin cuvinte de bază, astfel ca utilizatorul să aleagă noțiunea cea mai apropiată de necesitatea sa, asigurându-se folosirea de cuvinte agreate;
- efectuarea de corecții prin folosirea metricilor de similaritate, cu acceptarea sau respingerea de către utilizatori.

Din punct de vedere a datelor de intrare, aplicațiile informatice distribuite se clasifică în:

- aplicații cu selecție de legături și fără introducerea de valori pentru a inițializa variabile;
- aplicații cu introducere de date de autentificare și selecție de legături;
- aplicații cu autentificare și inițializare de variabile ce determină selecții de operații care nu au la bază alte inițializări, ci efectuează numai prezentări de texte și imagini;
- aplicații care inițializează variabile ce determină selecții urmate de prelucrări bazate pe inițializări de alte variabile;
- aplicații în care au loc inițializări și stocări de informații cu efectuări de prelucrări.

În tabelul 1, sunt prezentate tipurile de aplicații în raport cu operații de autentificare, inițializare de variabile, construirea de fișiere, utilizarea de baze de date, contribuții ale utilizatorilor în popularea bazelor de date destinate prelucrărilor.

Tabelul 1. Tipuri de aplicații informatice

Aplicație informatică	Autentificare	Selecție legături	Selecție după un câmp inițializat	Efectuarea de calcule cu variabile inițializate	Stocare de variabile inițializate	Adăugare baze de date
Aplicație publică de informare						
Aplicație de informare cu utilizatori selectați	*					
Aplicație de informare selectivă pentru anumiți utilizatori	*	*				
Aplicație de informare cu autentificare cu calcule folosind variabile inițializate				*		
Aplicație cu selecție și efectuare de calcule cu date din baze de date existente						
Aplicație cu autentificare și creare baze de date						
Aplicație cu autentificare și utilizare baze de date						
Aplicație cu autentificare, creare și utilizare baze de date și stocare de rezultate						
Aplicație cu autentificare, selecție, creare, utilizare și stocare de rezultate						

Diversitatea aplicațiilor informatice distribuite în raport cu datele de intrare este deosebit de mare. Sunt situații în care, pentru o aplicație, se regăsesc tipurile de date de intrare.

Datele de intrare într-o aplicație au rolul de a:

- intra direct în expresii și de a conduce la obținerea de rezultate;

- determina selecții de modele de calcul, de a selecta secvențe de text;
- defini probleme, fișiere, structuri de articole;
- preciza nume de câmpuri, tipuri de câmpuri;
- selecta comenzi de stocare în fișiere sau baze de date;
- defini texte de comunicare;
- inițializa articole într-o bază de date;
- stabili cine declanșează interacțiunea om-calculator;
- declanșa procese de autentificare;
- corecta câmpuri cu definiri multiple sau cuvinte introduse cu valori în afara intervalelor.

Se impune ca, în viitor, să se asigure crearea de baze de date care, prin analiza automată, să ofere informații importante în legătură cu:

- structura aplicației;
- absența unor funcții de prelucrare;
- existența informației redundante și a cauzelor care o generează;
- gradul de securitate a aplicației;
- determinarea cauzelor care conduc la creșterea timpului de acces la informații;
- cauzele care generează nivelul scăzut de satisfacție a utilizatorilor;
- punctele critice ale unui flux de prelucrare, traversat de utilizator;
- minimizarea fluxului de prelucrare;
- absența clarității unor informații și ambiguitatea unor comenzi.

În acest scop, aplicația distribuită este înzestrată cu câmpuri care măsoară accesările, în punctele cheie, dând posibilitatea ca, prin calcule de frecvențe de accesare și comparări, să se obțină aprecieri calitative privind construcția interfețelor și chiar asupra structurii aplicației.

2. Caracteristici de calitate a datelor de intrare

În [1] sunt prezentate caracteristici de calitate a datelor. Pentru fiecare aplicație informatică distribuită, este necesar să se facă un studiu și să se extragă unele caracteristici de calitate a datelor de intrare, care influențează și calitatea rezultatelor, în raport cu obiectivul definit. Aplicația informatică vizează cel puțin o colectivitate. În cazul în care membrii colectivității accesează aplicația pentru a se informa, caracteristicile urmărite sunt:

- completitudinea;
- corectitudinea;
- claritatea;
- consistența;
- controlul redundanței;
- asociativitatea;
- distributivitatea;
- simplitatea;
- accesibilitatea;
- comparabilitatea.

Site-ul destinat prezentării timpului probabil, din punct de vedere tehnic trebuie realizat astfel încât să se ajungă rapid la:

- intervalul de timp care îl interesează pe utilizator;
- zona geografică pentru care se dorește prognoza;
- tipul de informație furnizat;

- furnizarea de date istorice ale condițiilor meteorologice de-a lungul timpului și identificarea de secvențe comune la intervale de timp;
- precizia oferită asupra prognozei;
- asemănarea structurii informațiilor cu alte site-uri;
- asigurarea formei de prezentare, grade Celsius și grade Fahrenheit, la alegere sau împreună.

Dacă utilizatorul are limitări în a face selecții din volumul de informații furnizate, devine dificil de identificat acele informații de care are nevoie.

Dacă în aplicație sunt contorizate accesările, se restructurează criteriile de accesare astfel încât să se minimizeze calea până la informațiile selectate.

Corectitudinea datelor de intrare este caracteristica prin care se asigură ca, între nivelul măsurat, nivelul stocat în baza de date și nivelul real al caracteristicii de descriere a elementului din colectivitate, să nu existe diferențe.

Datele de intrare sunt absolut corecte, dacă diferențele între datele culese, cele introduse și stocate și nivelele reale sunt nule. Datele sunt parțial corecte, dacă există diferențe, sunt localizate aceste diferențe și în iterația următoare are loc efectuarea de corecții prin repetarea măsurărilor. Datele sunt incorecte dacă s-a identificat cel mult o diferență, și nu există modalități de efectuare a corecțiilor din cauza derulării dinamice sau absenței elementului pentru a relua măsurătorile.

Completitudinea este caracteristica prin care se asigură că, pentru toți membrii colectivității, au fost introduse date, iar datele de intrare au permis inițializarea tuturor câmpurilor care constituie caracteristicile de descriere a elementelor.

Este important să se cunoască:

- necesarul de elemente care furnizează colectivitatea;
- caracteristicile care descriu elementele colectivităților;
- domeniile de variație ale fiecărui câmp.

În construirea seturilor de date, se realizează tabele cu atâtea coloane câte caracteristici sunt utilizate la definirea elementelor și atâtea linii, câte combinații de niveluri măsurate acoperă diversitatea elementelor din colectivitate.

Se evită situațiile în care, pentru elemente și caracteristici, nu se măsoară nivelurile întrucât astfel de situații diminuează nivelul de completitudine. Completitudinea se gestionează prin improbabilitatea de a trece la următoarea preluare dacă datele despre un element sunt incomplete.

Reproductibilitatea procesului de măsurare conduce la reproductibilitatea obținerii de date neafectate de erori de măsurare. Este necesar să se definească proceduri, să se dea exemple, să se precizeze instrumente de măsură și, mai ales, să se arate care sunt erorile de culegere a datelor și care anume afectează această caracteristică.

Verificarea reproductibilității este asigurată prin:

- instruirea operatorilor care culeg date prin prezentări de proceduri;
- constituirea unui lot reprezentativ de elemente din colectivitate;
- efectuarea de către operatori de măsurători simultane pentru toate elementele din lot;
- efectuarea de măsurători de către cei care au definit procedurile sau au instruit pe operatori;
- calculul de indicatori ai analizei comparat.

Existența diferențelor semnificative este cauzată de erori generate de:

- calitatea procedurilor;
- calitatea instrumentelor;
- comportamentul subiecților;
- nivelul de instruire al subiecților cu privire la utilizarea aplicației;
- derularea procesului;
- calitatea operatorului.

Identificarea cauzelor impune modificări în proceduri, comportamentul subiecților, selecția

operatorilor și utilizarea instrumentelor de măsurare. În aplicațiile distribuite, un rol important revine operatorului care efectuează măsurătorile. În cazul în care fiecare utilizator este și operator, se impune ca:

- prezentarea procedurilor de măsurare să fie la dispoziția directă a utilizatorului;
- prezentarea dispozitivului de măsurare să se efectueze complet și pe înțeles;
- exemplele de măsurare corectă să fie sugestive și să acopere totalitatea situațiilor;
- modul eronat de culegere a datelor să fie prezentat, indicând efecte negative și mai ales modalități de îndreptare.

Aplicațiile cele mai importante, cu impactul cel mai puternic, sunt acelea în care utilizatorul este cel care furnizează datele de intrare. Instruirea utilizatorului pentru a fi un operator competent reprezintă în societatea informațională cheia succesului aplicațiilor distribuite majore.

Caracterul ermetic pe care îl încorporează acum aplicațiile informatice generează erori în bazele de date create de operatori, compromițându-se credibilitatea aplicației prin avalanșa de corecții care se impun. Sunt numeroase aplicațiile distribuite în care transpunerea datelor despre cetățenii din evidențele organizațiilor s-a făcut fără certificarea datelor. Modul aberant în care organizațiile se substituie cetățeanului în a certifica rezultatele le aduce în foarte multe cazuri în situații umiltoare, iar viitorul apropiat va pune organizațiile în situația de a plăti disconfortul creat cetățeanului.

3. Interfețe pentru introducerea datelor

Datele de autentificare vizează:

- nume utilizator;
- parolă;
- date de descriere a utilizatorului;
- nivelul de acces și domeniu de vizibilitate al utilizatorului;
- în unele cazuri, utilizatorul trebuie să introducă un șir de caractere care să certifice faptul că, într-adevăr, o persoană introduce datele, și nu un produs software.

Aceste date de intrare se introduc o singură dată și se stabilește cheia de regăsirea numelui de utilizator fiind concatenat cu parola. Numele utilizatorului și parola se caracterizează prin:

- lungime minimă impusă;
- lungime maximă acceptată;
- alfabet;
- vocabular de cuvinte neacceptate;
- structură definită prin utilizare de simboluri speciale;
- unicitatea comparând cu șirurile deja existente;
- reguli de construire;
- nivelul de tărie al parolei, respectiv în ce măsură respectiva parolă face față unor cerințe de siguranță; prin nivelul de tărie, se urmărește informarea utilizatorilor pentru a crea parole care să conțină combinații de caractere cât mai nefirești și mai greu de generat;
- certificare parolă ca element de siguranță pentru ca definirea să fie cea dorită de utilizator având în vedere contextul ascuns al acestuia la introducere.

Există proceduri standard pentru a implementa elementele de autentificare. Întrucât accesul la aplicația distribuită se efectuează de la un calculator și orice calculator are tastatură, evident, definirea câmpurilor de autentificare și însăși autentificarea prin nume utilizator și parolă devin operaționale indiferent de context.

Aplicațiile cele mai des întâlnite evită ca datele de intrare să fie obținute ca șiruri de caractere introduse de la tastatură, în mod predominant. Dacă utilizatorul și-a introdus datele de identificare esențiale o dată, nu o va mai face și cu alte ocazii, întrucât o aplicație care solicită astfel de date în mod repetat este de fapt un rebut. Dacă utilizatorul trebuie să introducă șiruri de caractere, se precizează:

- numele câmpului;
- unitatea de măsurare;
- lungimea câmpului;

- exemple dacă este activat un buton alăturat;
- niveluri admise;
- tipul câmpului;
- separatorii;
- regulile de completare;
- alinierea stânga sau dreapta.

Cel care realizează interfața include secvențe care:

- inhibă tastarea altor caractere;
- efectuează conversii litere mici în litere mari acolo unde este cazul;
- poziționează caractere la stânga, la dreapta sau în raport cu un simbol predefinit.

Toate acestea au menirea de a-l dirija pe utilizator spre direcția furnizării de date corecte în raport cu criteriile impuse. Câmpurile opționale sunt precizate explicit. Câmpurile obligatorii sunt precizate, de asemenea, explicit. După completare se va activa un buton, de regulă SUBMIT sau o traducere a lui.

Crearea de aplicații care să identifice utilizatorul prin solicitarea de informații personalizate reprezintă o nouă etapă în încercarea de protejare a datelor furnizate în rețea. La tastarea numelui de utilizator, aplicația rulează o serie de programe care sunt parametrizate pentru un singur utilizator, astfel ca, un eventual intrus, chiar dacă trece o parte din etapele de identificare, prin restrângerea ariei de cunoaștere a datelor despre cont, să fie pus în imposibilitatea de a mai continua operațiunea de logare.

O nouă provocare este reprezentată de crearea de produse software, care generează aleator utilizatori și parole. Sunt astfel asediate baze de date în încercarea de a sparge conturi. O soluție pentru prevenirea acestor situații o reprezintă implementarea de elemente de securitate suplimentare, care să se adreseze în mod direct unei persoane și care să solicite realizarea de acțiuni pe care doar un om le poate face. Se pune accentul pe generarea de secvențe text aleatorii, care sunt prezentate utilizatorului aplicației distribuite și pe compararea răspunsului furnizat de acesta cu privire la secvența text citită.

În acest sens, o altă nouă soluție este reprezentată de implementarea unor teste logice, la care utilizatorul trebuie să răspundă. Astfel, se realizează atât interacțiunea om-mașină, care certifică faptul că o persoană este cea care a inițiat secvența de acces la date, cât și realizarea de procese logice, specifice doar omului, știut fiind faptul că, pentru a da soluție unei probleme, o mașină trebuie să fie programată în prealabil cu datele și modul de rezolvare a problemei.

Secvența de analiză a datelor de intrare, furnizate liber prin tastare de utilizator, are două mesaje:

- inițializare câmpuri, dacă datele au fost corecte;
- revenirea la faza de introducere de date și semnalizarea de erori, precum: stabilirea câmpurilor obligatorii neinițializate, introducerea de șiruri ce nu respectă reguli ca lungime, ca alfabet utilizat sau ca structură de sub-șiruri.

Se reia procesul și trecerea la următoarea etapă se efectuează dacă și numai dacă inițializarea câmpurilor din articol este validată și articolul este înscris într-o bază de date. Astfel de abordări sunt necesare atunci când un utilizator interacționează cu aplicații pentru rezervări la hotel și pentru achiziționări de bilete de avion și este necesar să se înscrie un card debit/credit și elemente de identificare a utilizatorului.

În interesul aplicației distribuite, prelucrările au la bază date de intrare, selectate de utilizatori. Selecțiile din listele de valori reprezintă modul flexibil de definire a problemei utilizatorului. Astfel, se definesc:

- anul de începere și de terminare a unui interval;
- luna de începere și de terminare;
- ziua din lună pentru începere și pentru terminare;
- șirul de localități ce definesc un traseu;
- locația de unde se solicită alocarea de resurse;
- tipul de resursă dorit;
- cantitatea din resursa selectată;
- modul de efectuare a plății;
- modul de comunicare mesaje.

Selecția unei variante dintr-un set sau a unei combinații de variante presupune:

- stabilirea tuturor variantelor;
- inexistențe de conflicte între variante;
- existența formulei de calcul, care inițializează cu 1 coeficienții selectați și care generează pentru r variante 2^r combinații de comportament pentru aplicație.

La definirea datelor de intrare, este necesar să existe validări puternice care privesc:

- apartenența la domenii specifice;
- folosirea de termeni corect introduși;
- utilizarea de cuvinte cheie cu o frecvență a cărei valoare trebuie să fie situată peste un anumit nivel;
- apartenența la un interval de valori, pentru câmpurile numerice, și la un vocabular, pentru câmpurile alfa-numerice;
- identificarea momentelor când au fost interschimbate șiruri ale câmpurilor din vecinătate;
- completitudinea și, mai ales, capacitatea de prelucrare a informațiilor din șirurile de date introduse în timp;
- respectarea încadrării nivelurilor în distribuțiile de frecvențe ce caracterizează elementele colectivității;
- respectarea unor inegalități definite la prima accesare a aplicației de către utilizator; utilizatorul fiind definit prin resursele pe care le posedă, iar consumul de resurse nu va depăși nivelul disponibil existent.

Se consideră caracteristicile $c_1, c_2, \dots, c_m$ pentru care utilizatorul introduce datele. Fiecare caracteristică are un nume descris prin șirul de cuvinte $s_1, s_2, \dots, s_m$. Fiecare caracteristică are un domeniu de definiție dat ca listă de valori din care este selectat un element, ca interval în raport cu care se selectează sau se generează un element. Fiecărei caracteristici i se asociază o unitate de măsură. Se urmărește ca, pentru același tip de element din sistemul internațional de măsură, să se folosească aceeași unitate. Se precizează pentru fiecare caracteristică intervalul aderat de măsurare.

Se definește procedura care asigură măsurarea corectă. Se prezintă modalități eronate, care determină involuntar măsurători incorecte. Se atrage atenția asupra cauzelor generatoare de erori și cum se preîntâmpină.

Interfețele se proiectează astfel încât, la mesajele formate din cuvintele $c_1, c_2, c_3, \dots, c_m$, utilizatorul să furnizeze date de intrare $s_1, s_2, s_3, \dots, s_m$, astfel încât perechile (c_i, s_i) , $i=1,2,\dots,m$ să permită selectarea procedurilor pentru a efectua strict prelucrările pe care le dorește respectivul utilizator. Mai mult, la elaborarea specificațiilor pentru interfețe:

- se definesc acțiunile de prelucrare $a_1, a_2, a_3, \dots, a_k$ cu care este înzestrată aplicația, k fiind numărul maxim de soluții pe care utilizatorul le obține de la problemele sale;
- se construiește vocabularul VK de cuvinte cheie, cu care se elaborează interfețele specifice traversării fluxurilor necesare declanșării setului de acțiuni;
- se impune în corespondență cuvintele din vocabularul VK cu acțiunile construind o regrupare de forma dată în tabelul 2.

Tabelul 2. Gruparea acțiuni/cuvinte cheie

Cuvinte cheie Acțiuni	c_1	c_2	...	c_i	...	c_m
a_1						
a_2						
...						
a_j				m_{ij}		
...						
a_k						

La intersecția liniei i și coloanei j , se regăsește cuvântul m_{ij} pe care îl furnizează utilizatorul astfel încât să compună un mesaj complet cu toate datele de intrare de pe linie, dacă dorește să obțină soluția corespunzătoare acțiunii a_j .

La proiectarea interfeței, se prezintă:

- mecanismele de construire a cuvintelor m_{ij} ;
- testele de cuvinte dacă ceea ce se utilizează pentru un cuvânt cheie c_i aparține unei mulțimi de șiruri impuse;
- ordinea în care se dispun cuvintele cheie este dată de frecvența de completare, de stabilirea conținutului șirurilor generate de utilizator și de efortul de construire a șirului; primele se dispun șirurile cu constante și ultimulele cele cu variabilitate maximă; numele, codul numeric personal, adresele sunt constante pentru un utilizator în timp ce produsele achiziționate, cantitatea de produs diferă de la o accesare la alta a aplicațiilor distribuite.

Interfețele se bazează pe faptul că datele de intrare privind un utilizator sau o entitate provin din mai multe surse, se corelează și se creează premisele pentru solicitarea de modificări și activarea procedurilor de corecție pentru a asigura corectitudinea datelor. Interfețele se construiesc astfel încât să fie eliminate în mod categoric erorile grosolane în procesul de introducere a datelor. Erorile grosolane vizează:

- lungimea de cuvinte;
- lungimea de secvențe de caractere identice;
- frecvențe de apropiere a unor simboluri care nu se încadrează în distribuțiile de frecvențe construite într-un interval de timp suficient de lung, distribuții care au devenit stabile.

În cazul în care validările presupun prelucrarea de date din mai multe baze de date, se impune asigurarea aceluiași nivel de calitate pentru bazele de date. În cazul neactualizării unei baze de date, calitatea acesteia se diminuează, iar non-calitatea se propagă la nivelul interfețelor. La schimbarea denumirii de stradă, compararea adresei din contract cu adresa din actul de identitate generează diferențe care, în cazul operatorilor cu imposibilitatea de a opera actualizări în baze de date a contractelor, determină neefectuarea de acțiuni pe care le solicită utilizatorul. De aceea, la proiectarea de aplicații, în cazul elementelor volatile, trebuie să apară definiții multiple pentru un câmp marcând în acest fel istoricul conținutului (figura 1).

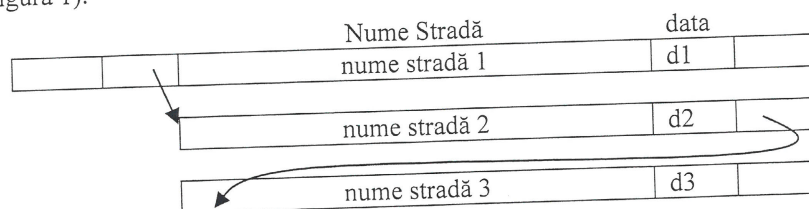


Figura 1. Evoluția dinamică a câmpului nume stradă

Construirea de baze de date care să preia istoricul modificării unor câmpuri constituie o soluție pentru reducerea datelor redundante. Bazele de date cu caracter istoric trebuie întreținute și actualizate de fiecare dată când apar modificări în câmpurile de identificare a utilizatorilor.

Schimbarea numelui sau adresei generează un set de noi înregistrări și, în același timp, dealocă vechile înregistrări care devin astfel inutilizabile. Ștergerea acestor înregistrări nu reprezintă o soluție întrucât este foarte posibil ca respectivul utilizator să aibă articole generate pe numele lui. Stocarea vechilor informații într-o bază de tip istoric și crearea unei chei de legătură între vechiul nume și noul nume rezolvă această problemă, permițând în orice moment vizualizarea înregistrărilor generate pentru un utilizator, chiar dacă anumite câmpuri au suferit modificări.

Între interfețele aplicațiilor și prelucrările asociate acestora trebuie să existe o concordanță perfectă. Selecțiile utilizatorului trebuie să-i ofere acestuia toate elementele pentru a furniza datele de intrare strict necesare obținerii rezultatului dorit. Dacă unele date există deja în bazele de date ale aplicației, nu vor apărea cuvinte cheie care să solicite introducerea de către utilizator, acestea extrăgându-se din baze de date. Dacă unele șiruri au fost introduse într-o etapă anterioară, dar în aceeași sesiune, se vor prelua din punctul acela fără a se solicita reintroducerea lor.

4. Creșterea calității datelor de intrare

Creșterea calității datelor de intrare urmărește:

- la nivelul fiecărui utilizator, tastarea cuvintelor m_{ij} din vocabularul său, ori de câte ori se va efectua, să aibă un număr de reluări din ce în ce mai redus;
- la nivelul aplicației, modul în care se proiectează interfețele și se efectuează analiza șirurilor furnizate;
- prin modul în care se introduc șirurile, să lase utilizatorului cât mai puține câmpuri de introdus prin tastare.

Creșterea interesului dezvoltătorilor este o cale sigură de îmbunătățire a calității datelor de intrare prin faptul că aceștia:

- vor cunoaște tehnici și metode de măsurare a cunoștințelor utilizatorilor;
- vor orienta aplicația către utilizatorii ale căror caracteristici le știu și pentru care au efectuat măsurători;
- vor construi vocabularul VK folosind cuvinte pe înțelesul utilizatorilor, ținând seama de alte aplicații accesate de aceștia;
- vor acoperi toate tipologiile de interfețe și vor alege strict interfețele ce se potrivesc perfect utilizatorilor;
- vor cunoaște tehnici și instrumente de dezvoltare de aplicații distribuite cu niveluri de complexitate foarte ridicate, pentru a reuși preluarea datelor de la utilizatori în vederea personalizării interfețelor pentru aceștia.

Instruirea utilizatorilor vizează:

- stabilirea datelor de intrare pentru aplicație (acces și prelucrare);
- cunoașterea tipurilor de probleme pe care aplicația le rezolvă;
- cunoașterea modului de selecție a pașilor de prelucrare;
- cunoașterea modului în care se efectuează prelucrarea rezultatelor;
- cunoașterea modului în care se corectează datele introduse;
- cunoașterea parcursului datelor și a influențelor pe care acestea le au la obținerea rezultatelor;
- cunoașterea de prelucrare și a timpilor de răspuns ai aplicației;
- cunoașterea tipurilor de validări pe care le realizează aplicația;
- interpretarea mesajelor.

Instruirea sau autoinstruirea trebuie să se efectueze astfel încât rezultatele să fie vizibile rapid prin scăderea duratei dată de momentul când utilizatorul accesează prima dată aplicația și momentul în care obține prima soluționare de problemă care îi satisface interesul față de aplicație. Calitatea datelor de intrare este influențată de claritatea cu care sunt definite câmpurile. Sunt aplicații care cer prenumele și apoi numele. Sunt aplicații în care data calendaristică este descrisă prin zi, lună, an. Alte aplicații impun o abordare diferită. Pentru o dezvoltare corectă a fluxurilor, trebuie să se acorde o flexibilitate legată de colectivitățile ce accesează aplicația: în varianta în limba română: nume/prenume, zi, lună, an; în varianta în limba engleză: prenume / nume, lună, zi, an.

Atunci când datele de intrare sunt întregi pe un anumit interval, se preferă selecția. Când șirurile de cuvinte sunt denumiri de orașe, de străzi, mărci de produse, se preferă enumerarea într-o listă și relația. În cazul în care utilizatorul introduce șiruri de caractere, este rezonabil să se numereze aceste șiruri și să se creeze liste proprii din care acesta selectează.

Utilizarea de indicatori și efectuarea de măsurători pentru caracteristici de calitate au menirea de a orienta eforturile spre creșterea calității.

Atunci când datele de intrare conțin șiruri identice, se produc regrupări de câmpuri astfel încât introducerea repetată să fie exclusă pentru un același câmp. Pentru fluxurile de prelucrare date în figura 2, câmpul codul numeric personal (CNP) figurează pentru toate fluxurile de prelucrare.

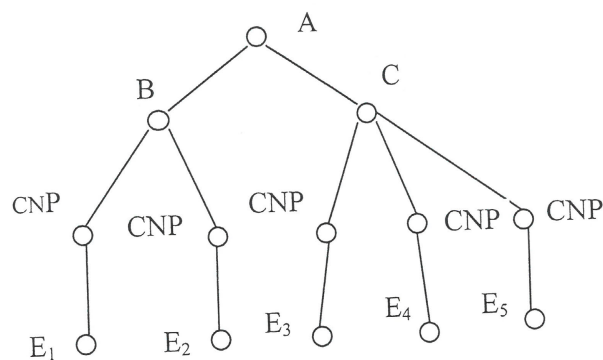


Figura 2. Fluxuri de prelucrare cu date de intrare repetitivă

Instrumentele de măsurare a repetabilității au rolul de a stabili frecvențele de apariție și de a găsi cel mai potrivit loc pentru câmpurile utilizate frecvent. Câmpul CNP își schimbă poziția rezultând structura dată în figura 3.

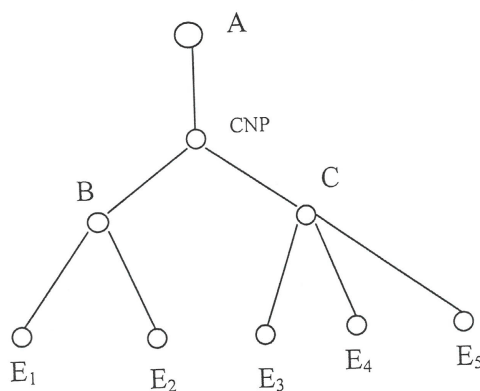


Figura 3. Fluxuri de prelucrare fără repetabilitatea datelor de intrare

Calitatea datelor de intrare este determinată de factorii externi aplicației, dar și de factorii interni ai aplicației, de bazele de date incluse în aplicație, de procedurile de prelucrare și de validare a datelor. Definirea specificațiilor joacă un rol esențial în asigurarea datelor de intrare. Prin specificații, se definesc reguli de construire a cuvintelor din vocabularul utilizatorului. Astfel, numele și prenumele conțin numai litere și cel mult liniuțe de unire. Aceasta se află numai în interiorul cuvântului. Numele nu conține subșiruri de litere identice cu lungime mai mare ca 2. Numele nu conține subșiruri asociate unui vocabular dat. Lungimea șirului asociat numelui nu este mai mare decât o valoare impusă. Numele nu conține subșiruri multiplicare mai mult de două ori.

Specificațiile includ exemplificări sugestive, care denotă o cunoaștere foarte bună a problemelor de rezolvat. Calitatea datelor de intrare este influențată de definirea regulilor care permit construirea de șiruri ce se pun în corespondență cu elementele din realitate, asigurând o reflectare cât mai corectă a realității.

Aplicațiile distribuite combină în dezvoltarea de aplicații textele formate din cerințe cu imagini. Datele de intrare se construiesc inclusiv prin preluarea de șiruri construite cu ajutorul imaginilor.

Pentru o suprafață dreptunghiulară, datele de intrare presupun existența unei interfețe de tipul dat în figura 4.

Suprafața:

Lățime

▽

Lungime

▽

Figura 4 Interfața cu introducerea de date folosind și o reprezentare grafică

Utilizatorul selectează valori pentru lăţime şi lungime, activează butonul de calcul suprafeţe şi va apărea valoarea suprafeţei.

Datele de intrare trebuie certificate de cel care le introduce. Faptul că accesul la resursele unei aplicaţii distribuite este condiţionat de o parolă şi semnarea unui document prin care responsabilitatea utilizatorului este total dependentă de parola de acces, orice date introduse se consideră certificate de utilizator. Nu se va lua în considerare că o altă persoană a introdus datele din moment ce parola marchează faptul că accesul la resurse a fost garantat de posesorul parolei.

Există şi situaţia în care transmiterea datelor este condiţionată de semnătura electronică. Aplicaţiile informatice pretenţioase, care privesc fluxuri băneşti, informaţii de fundamentare a deciziilor, date confidenţiale includ obligativitatea butonului de trimitere spre soluţionare să fie urmată de semnătura electronică.

Creşterea calităţii datelor de intrare este şi responsabilitatea utilizatorului, dar şi a dezvoltătorilor, şi trebuie să-i preocupe pe toţi în egală măsură.

Orice deficienţă trebuie semnalată şi trebuie să-şi găsească soluţie la nivelul interfeţei prin:

- adăugarea de câmpuri care permit asigurarea completitudinii datelor pentru a obţine o soluţie;
- construirea unui *help* foarte dezvoltat, care să explice utilizatorilor aplicaţia;
- construirea de liste de selecţie cu termeni cheie pe care utilizatorul să-i selecteze;
- validarea câmpurilor în timp real şi înştiinţarea utilizatorului cu privire la incorectitudinea datelor;
- mutarea unui câmp pe un alt nivel pentru a asigura nivelul de repetabilitate optim;
- eliminarea de câmpuri care nu reflectă cerinţe în raport cu soluţia dorită.

În planul aplicaţiei, apar situaţii ce privesc:

- creşterea numărului de acţiuni, dacă la definirea specificaţiilor nu au fost luate în considerare toate situaţiile sau dacă au apărut noi probleme de rezolvat;
- regrouparea interfeţelor atunci când, pentru unele soluţii, se identifică clase de soluţie şi pentru o clasă este necesară definirea unei singure interfeţe;
- rafinarea de soluţii, reuşind să se obţină un nivel de ortogonalitate cât mai ridicat pentru soluţii, chiar dacă se includ într-o clasă.

Pentru creşterea calităţii datelor de intrare trebuie să existe o permanentă preocupare, trebuie definit un management eficient şi trebuie făcută o educaţie permanentă spre a construi datele de intrare corect, complete, consistente şi comparabile, în mod natural, obişnuit.

5. Certificarea datelor de intrare

În contextul utilizării intense de aplicaţii distribuite, datele de intrare trebuie certificate. A certifica datele de intrare înseamnă:

- stabilirea dacă, din punct de vedere cantitativ, datele de intrare asigură bazele declanşării fluxurilor de prelucrare pentru obţinerea de soluţii;
- analizarea dacă, din punct de vedere calitativ, datele de intrare permit obţinerea de rezultate corecte şi complete, care să satisfacă exigenţele utilizatorilor.

Se preiau din specificaţii indicatorii prin care se defineşte colectivitatea ca număr şi caracteristici de descriere. Se analizează dacă interfeţele asigură:

- introducerea datelor pentru iniţializarea câmpurilor asociate caracteristicilor;
- controlul asupra numărului de câmpuri iniţializate;
- verificarea numărului de elemente pentru care s-a efectuat iniţializarea.

În cazul în care se specifică elementele, se furnizează lista pentru care nu s-au efectuat iniţializările. În cazul iniţializărilor parţiale, se editează rapoarte privind articolele complete şi elementele din colectivitate pentru care trebuie efectuate completări.

Certificarea vizează capacitatea aplicaţiei de a gestiona latura cantitativă a datelor de intrare. Procesul de certificare avansează la nivelul de detaliere mai ridicat:

- enumerând tipurile de rezultate pe care le aşteaptă utilizatorii de la aplicaţii;

- stabilind structura pe conținut a rezultatelor, ce date de identificare conțin, ce indicatori calculați și ce date inițiale sunt incluse în raport;
- stabilind procedurile prin care se obțin rapoartele asociate rezultatelor așteptate de utilizator cu construirea modelelor de calcul de indicatori;
- analizând dacă bazele de date conțin câmpurile care se preiau ca atare în raport și câmpurile care intră în formulele de calcul;
- găsind modul de creare a rapoartelor cu posibilitatea de căutare și regăsire, ori de câte ori e nevoie, fără a fi necesară reluarea procesului de realizare pentru a-l feri pe utilizator să reconstruiască procese de selecție, chiar dacă bazele de date asigură prin conținut reproductibilitatea procesului de realizare a unui raport dintr-o perioadă anterioară;
- identificând procedurile de încărcare în bazele de date a câmpurilor cu validări astfel încât prelucrările să demareze pentru un element din colectivitate dacă și numai dacă datele de intrare respectă cerințele impuse prin procesul de certificare.

Certificarea datelor de intrare evidențiază că aplicația informatică distribuită are toate componentele care să-i permită să fie operațională, să satisfacă exigențele utilizatorilor. Nu înseamnă că, o dată cu terminarea cu succes a procesului de certificare, aplicația va funcționa astfel încât să-i satisfacă pe utilizatori.

Dacă în colectivitate sunt n elemente și sunt introduse date de intrare de către m utilizatori, $m < n$, rezultă că pentru $n-m$ utilizatori aplicația este inaccesibilă.

Dacă datele de intrare furnizate de utilizatori sunt incorecte, prin corelații cu datele deja existente se stabilesc erori grosolane precum:

- șiruri de cuvinte prea lungi;
- numere cu multe ordine de măsură în plus față de situațiile frecvente;
- frecvențe prea mari și intervale de timp prea mici legate de accesarea de resurse.

Certificarea include etape prin care se constată existența procedurilor care au menirea de a găsi erori grosolane.

La crearea bazelor de date, apar și o serie de erori privind:

- introducerea repetată a unor câmpuri;
- asocierea mai multor chei pentru un element din colectivitate;
- interschimb între câmpuri;
- decalarea câmpurilor privind mai multe elemente din colectivitate;
- omiterea datelor privind un element sau un grup de elemente din colectivitate.

Certificarea are menirea de a vedea care sunt secvențele de program, care analizează conținutul bazei de date pentru a vedea repetările și pentru a stabili dinamic domeniile variabilelor din colectivitate.

Este posibil ca, în final, să se prezinte un raport asupra ceea ce s-a întâmplat în procesul de creare a bazei de date.

În cazul interogărilor, certificarea, prin efectuarea de teste, are menirea de a stabili:

- proporția interogărilor de succes;
- gruparea interogărilor eșuate;
- lungimea fluxurilor de prelucrare;
- calitatea mesajelor comunicate utilizatorilor fie că este vorba de interogări de succes și, cu atât mai mult, dacă este vorba de interogări eronate;
- modalitățile și punctele de unde se reiau interogările în cazul insuccesului; se vor preciza situațiile care preiau din punctul inițial și cele care preiau cu efectuarea numai a corecțiilor.

Rezultatul certificării este transferul de credibilitate că aplicația informatică distribuită conține toate elementele de:

- interfață,
- prelucrare,
- securitate,
- acces,

care îi permit să ofere utilizatorului exact resursele de care acesta are nevoie.

Certificarea datelor de intrare este o componentă a certificării aplicației informatice și sunt vizate, de asemenea, interfețele, prelucrările și elementele de securitate, strict din punct de vedere al informațiilor furnizate de utilizatori.

Sunt analizate cerințele afișate de aplicație inițial pentru utilizatori și mesajele furnizate de utilizatori față de șirurile de cuvinte introduse de la tastatură sau selectate de utilizator sau preluate din fișierele utilizatorului. Sunt analizate corespondențele existente în aplicație între șirurile de cuvinte selectate sau furnizate de utilizator și gama prelucrărilor pentru a vedea dacă:

- textul furnizat activează procedurile de prelucrare în mod corect;
- există proceduri neactivate;
- procedurile care se activează incorect au la bază expresii condiționale corecte sau sunt plasate incorect în secvențe de apel.

Cei care certifică aplicații distribuite, prin experiența pe care o au, dispun de structuri de seturi de date pentru a evidenția tipologii de erori incorporate în aplicații.

Certificarea datelor de intrare revine la a structura șiruri de caractere, inclusiv de lungime vidă, la a vedea comportamentul aplicației în zonele de comunicare, la prelucrarea, securizarea și stabilirea măsurii în care aplicația răspunde unor exigențe de calitate a datelor de intrare.

Datele de intrare determină calitatea rezultatelor. Preluând din specificații informații privind:

- tipurile de rezultat,
- structurile de rezultate,
- nivelurile caracteristicilor de calitate,

se testează aplicația cu seturi de date de intrare existente, dar mai ales cu seturi construite de echipa care efectuează certificarea și se măsoară riguros corelația între calitatea rezultatelor și calitatea datelor de intrare. În urma certificării datelor de intrare, va rezulta cu precizie contribuția datelor de intrare în a certifica aplicația pentru un uz corect sau pentru a nu acorda certificat de calitate, ceea ce presupune că revenirea asupra:

- structurilor de șiruri de cuvinte;
- construirii vocabularului de cuvinte cheie;
- mecanismelor de generare cuvinte utilizator;
- punerii în corespondență a cerințelor cu acțiuni de prelucrare;
- secvențelor interne de validare;
- procedurilor de asigurare a ortogonalității bazelor de date

este absolut necesară.

Echipa care efectuează certificarea elaborează un raport în care în mod transparent prezintă:

- cum s-a efectuat certificarea;
- care sunt testele efectuate;
- cum s-a comportat aplicația pentru fiecare baterie de teste;
- care sunt rezultatele testării pentru care nu se acordă certificat;
- ce condiții trebuie să îndeplinească aplicația pentru a primi certificat în viitor.

Numai în acest caz, certificarea își atinge obiectivul, iar procesul de certificare devine credibil în fața celor care au elaborat procesul necertificat.

6. Concluzii

Aplicațiile distribuite operează cu volume de date foarte mari. Față de alte situații, deja există aplicații care derulează sute de mii de tranzacții pe zi. Achiziționarea on-line de bilete, accesarea de blog-uri, plățile electronice, instruirea asistată, vizualizarea unor informații sunt numai câteva din aplicațiile distribuite aflate în uz curent, care și-au dovedit viabilitatea. Numeroase aplicații distribuite se caracterizează prin introducerea de date de către utilizator prin selecție dintr-un vocabular preluat și prezentat sau generat și prezentat. Sunt puține elementele ce privesc datele care se introduc de la tastatură,

cel mult nume și prenume, cod card de credit, un nume utilizator și o parolă.

În perioadele viitoare, datele de intrare vor fi mult mai multe, mai diverse și utilizatorii vor contribui **direct** la crearea bazelor de date, furnizând date de intrare sub formă de:

- șiruri de cuvinte din vocabularul unei limbi cu restricții de neutilizare a cuvintelor dintr-un subvocabular;
- fișiere proprii unde sunt texte structurate;
- componente ale bazelor de date ce provin din alte aplicații distribuite.

În definirea de specificații, trebuie:

- să se prezinte cu claritate importanța calității datelor;
- modalitățile care impun restricțiile de responsabilizare și ce se întâmplă dacă nu sunt introduse corect datele;
- modalități și termene de efectuare a corecțiilor.

În noul context în care utilizatorul aplicației informatice distribuite preia rolul operatorului, responsabilitatea lui este directă, întrucât tot ce decurge din datele introduse și mai ales corecțiile de orice natură, au la bază date corecte prin certificare directă de către utilizator. Certificarea se realizează prin două elemente: accesarea la resurse prin nume utilizator și parolă și prin transmiterea datelor activând un buton *Trimite* sau *Submit*. Aplicația certificată are toate elementele care elimină transmiterea accidentală și includerea în baza de date a șirurilor de cuvinte incorecte, incomplete. Activarea butonului *Trimite* sau *Submit* este rezultatul deciziei că utilizatorul consideră datele sale de intrare ca reflectând realitatea.

Există în aplicația informatică distribuită următoarele situații:

- utilizatorul furnizează date de intrare, care îndeplinesc toate cerințele de calitate și reflectă calitatea, obținând tot timpul rezultate corecte;
- utilizatorul furnizează date de intrare, care îndeplinesc cerințele de calitate, dar nu reflectă calitatea, obținând uneori rezultate corecte alteleori incorecte;
- utilizatorul furnizează date ce conțin elemente de noncalitate, care sunt preluate, și se declanșează prelucrări, ceea ce arată că aplicația are carențe esențiale și, în scurt timp, va trebui să fie retrasă sau va genera pierderi suportate de proprietari;
- utilizatorul nu reușește să participe la crearea bazei de date datorită faptului că datele sale nu corespund calitativ și utilizatorul nu trece de etapa de corectare a șirurilor de cuvinte pentru a avea succes la închiderea acțiunii generate de activarea unui buton *Trimite* sau *Submit*.

Rezultă că succesul aplicației informatice distribuite are ca fundament managementul calității datelor de intrare. Contribuția dezvoltătorilor de aplicații și a utilizatorilor în a asigura calitatea datelor de intrare este singura cale de a face aplicațiile să fie viabile pe termen lung.

Bibliografie

1. **IVAN, I., Gh. NOSCA, M. POPA:** Managementul calității aplicațiilor informatice. Editura ASE, București, 2006.
2. **IVAN, I., C. BOJA:** Practica optimizării aplicațiilor informatice. Editura ASE, București, 2007, ISBN 978-973-594-932-7, 483 p.
3. **IVAN, I., M. POPA:** Uniformitatea, caracteristica de calitate a entităților text, Informatică economică, vol. 9, nr. 1, 2005.
4. **IVAN, I., P. POCATILU, D. CAZAN:** Certificarea bazelor de date utilizate în aplicațiile Internet, Informatică Economică, vol. 5, nr. 2, 2001.
5. **IVAN, I., P. POCATILU, S. CAPIZISU:** Certificarea în informatica aplicată, Informatică Economică, vol. 4, nr. 2, 2000.
6. **IVAN, I., A. POCOVNICU:** QSPM – Software pentru managementul calității proiectelor, Informatică Economică, vol. 6, nr. 4, 2002.
7. **IVAN, I., C. TOMA:** Aplicații informatice orientate către utilizatorii finali, Informatică Economică, vol. 7, nr. 4, 2003.
8. <http://www.theserverside.com>: Comunitatea online a dezvoltătorilor ce folosesc limbajul de programare Java.